Estimation of monthly average daily of the global solar radiation using the linear regression algorithm

Hussein H. Hadi1*
1* Ministry of Education, Iraq
hunzgz@yahoo.com

Abstract. Solar radiation is the backbone for the existence of life on the earth. Its measurement is too expensive. Therefore, it has been great importance to propose an efficient method to use as a solar predictor based on other more readily available meteorological data. In this study, Linear regression method has been applied in two locations that have two different climates in Egypt, location A that has a coastal climate on the Mediterranean coast and location B that has a dry desert climate, using three weather vectors (minimum, maximum and average) of temperature values. The performance of the linear regression models provides better predictions for global solar radiation (GSRa) at different locations. The Root Mean Square Error (RMSE) = (2.7716 and 2.9392) MJ m\(^{-2}\) with Relative Root Mean Square Error (rRMSE) = (14.5962 and 14.0861) % in location A and B respectively. The accurate prediction results of the GSRa using this approach can be employed in the various purposes of the solar applications.

Keywords: Egypt, linear regression, prediction model, solar radiation, temperature value.

1. Introduction
In recent years, there is a growing interest in renewable energy sources. Solar energy is one of renewable energy sources and most viable options to replace fossil fuel power plant. It is one of the best solutions to handle reduced production problem of the power in many countries. Egypt has many amounts of solar radiation, which is available around the seasons that can be invested to obtain energy [1] [2]. The optimal location selection for solar radiation intensity is essential to develop solar energy-based projects and long-term evaluation of the solar energy conversion systems performances [3]. The some difficulties in solar radiation measurement as environment and high cost, many researchers began using machine learning models to calculate the amounts of solar radiation. The models are depending on various climate variables such as sunshine duration, temperature, relative humidity and solar declination angle to predict the amounts of the solar radiation. Machine learning methods are a very strong regularization strategy and guaranteed to build best models that are useful in predicting [4] [5].

A lot studies used machine learning techniques to estimate GSRa and have a perfect outcome prediction. One of the studies used one of the most important machine learning techniques named Deep Neural Networks (DNNs) to solve two issues, wind energy and daily solar radiation prediction [6]. Proposed study applied machine learning algorithms to estimate hourly of the solar irradiance based on two types of the input data that have been used to develop forecasting models. The first model uses pressure, local time, wind speed, temperature and relative humidity as input variables; the second model is the time-series prediction of solar irradiance in Abu Musa Island [7]. Comparative study used air temperature as inputs for prediction of daily GSRa in China. Four machine learning
models, original artificial neural network, hybrid mind evolutionary algorithm, random forests and wavelet neural network, beside four empirical temperature-based models (Fan model, Hargreaves-Samani model, Jahani model and Bristow-Campbell model) have been applied [8]. Assessment study to estimate monthly average daily of the GSRa in four locations in Turkey (Ankara, Karaman, Kilis, and Şırnak) by building models based on machine learning algorithms (feed-forward neural network), empirical models (3 Angstrom-type models), time series (Holt-Winters) and mathematical model. As well as, the dataset including relative humidity, pressure, ambient temperature, wind speed and sunshine duration [9]. A novel intelligence model named hybridization of Adaptive Neuro-Fuzzy Inference System (ANFIS) with two metaheuristic optimization algorithms, Salp Swarm Algorithm (SSA) and Grasshopper Optimization Algorithm (GOA) (ANFIS-muSG) have been used to estimate GSRa in North Dakota, USA. The maximum, mean and minimum of the air temperature data for nine years from 2010 to 2018 have been used to apply the models. In addition, the intelligence models using the term root mean square error to calculate the performance of each model [10].

In our work, a linear regression method has been used to predict monthly average daily of the GSRa. This method is able to produce an accurate prediction depending on the values of temperature variables (maximum, minimum and average) as input. As well as, the estimation technique used some statistics indicator such as RMSE and rRMSE to assess the models that are valid for evaluating the linear method. This study has been applied in two locations in Egypt, location A that has coastal climate and location B with desert weather as a case of study.

2. Data collection and location description
This study, has been applied on two locations that have different climates, location A (Latitude 30° 55’ N and longitude 29° 41’ E) and location B (Latitude 27° 12’ N and longitude 31° 10’ E) which chosen from the north and south of the Egypt according to own weather characteristics for each location as a case study figure 1 [1][11][12].

The dataset that using in building linear regression models consist of various variables such as monthly average daily values of GSRa, minimum, maximum and average of temperature values, during 2003-2012. The dataset has been provided from Scientific Research and Technological Applications in Alexandria, Egypt.

![Figure 1: Geographical locations for locations A and B.](image-url)
3. Linear regression theory
The dataset has been distributed into the two groups the first group is training dataset and the second is testing dataset. The training dataset consists of inputs, denoted as \( x \), and the corresponding data is the output as class label, denoted as \( y \). The main goal is to predict the testing data to the corresponding class label [13]. In the regression method the output \( y \) has forms of one or more real numbers. Therefore, predict for input \( x \) an output \( f(x) \) that is close to the true \( y \). The learn a continuous function as the following formula:

\[
f(x) = b_0 + b_i x_i \quad i=1,2,...,n \tag{1}
\]

Where \( x_i \) is independent variables, \( b_i \) is a coefficient of the independent variables, \( b_0 \) is a constant and \( n \) is the number of the features in the feature space [14][15]. If \( x^{(i)} \) is used to denote the input variables and \( y^{(i)} \) to denote the output or target variables, a pair \( (x^{(i)}, y^{(i)}) \) represents a training example. The main aim of the linear regression method is to create a regressor that can be used to generalize a new instances form training examples. The regressor is a hypothesis function \( h_\theta(x) \) that can be used as a decision boundary to predict instances according to predefine of the class label elements. The regressor hypothesis function has the following form:

\[
h_\theta(x) = \sum_{i=1}^{n} \theta_i x^{(i)} \tag{2}
\]

Where \( \theta = \{\theta_i\}_{i=1}^{n} \) is a set of parameters model, and \( x = \{x^{(i)}\}_{i=1}^{n} \) is the set of training examples [16] [17]. To understand how the regressor initiated in linear regression technique, the previous series of analysis can illustrate in figure 2.

![Figure 2: Regressor initiated of the linear regression technique.](image-url)
4. Statistical testing approaches

In this paper, the evaluation of the models is established using the common indicators errors, RMSE, rRMSE. The smallest value of these two statistics is accepted value. The statistics of the errors are calculated using the following equations:

\[
RMSE = \left[ \frac{1}{n} \sum_{i=1}^{n} (C_i - M_i)^2 \right]^{1/2}
\]

\[
rRMSE = \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{C_i - M_i}{C_i} \right)^2 \right]^{1/2} \times 100
\]

Where \(C_i\) Is the \(i\) of the predicted values, \(M_i\) is the \(i\) of the measured values and \(n\) is the number of observations. The RMSE value explains information about the short-term performance of the prediction method and permanently positive value. A smaller value indicates best estimation performance, and zero refers to the ideal case. The rRMSE is the percentage of RMSE and the small value refer to the best estimation outcomes [13] [18].

5. Results and discussion

In this paper, a linear regression method in two locations in Egypt using temperature values elements (minimum, maximum and average) as input variables has been applied to predict monthly average daily of the GSRa. The results in Table 1 explain that the both linear regression models perform well in site A and B. A small values of the RMSE that are (2.7716 and 2.9392) MJ m\(^{-2}\) with rRMSE (14.5962 and 14.0861) % in location A and B respectively, indicate that the results are very close to measure values, also the excellent values of mean accurate explained that the models have high performance in prediction.

| Table 1: Statistical errors of the linear regression models in location A and B. |
|---------------------------------------------------------------|
| Location | RMSE (MJ m\(^{-2}\)) | rRMSE (%) | Mean Accurate |
|-----------|------------------------|-----------|---------------|
| Location A | 2.7716                  | 14.5962   | 0.9907        |
| Location B | 2.9392                  | 14.0861   | 0.9907        |

The Table 2 explains the values of measured and predicted data. The convergence of results between measured and predicted data is indicated that linear regression is useful in predict of the GSRa. In addition, the data of testing set that is represent a quarter of the total data set, which used in creating linear regression models. The Figure 3 is a comparison between measured and predicted values of monthly average daily of the GSRa for both cases of the study. Also, it showed that is remarkable agreement between the measured and predicted values. Figure 4 show the trend of measured and predicted values of the monthly average daily of the GSRa is very much similar. The adopted estimation of GSRa using linear regression formula on the basis ten years gives the best estimated values of monthly average daily of the GSRa using temperature values. Also, the convergence of results between the measurement and prediction data explained that temperature values are one of the best climate elements to estimate the amounts of the solar energy. The linear regression technique that has linear behavior proved high performance in estimation GSRa. Further, the predicted results of the linear approach were best from outcomes another machine learning techniques that have different estimation manners. The purpose of this study in calculation the prediction results were significant and reliable. Also, as further work in the future the performance of linear method can be improved through increasing the quality of the variables to support best feature space.
Table 2: A comparison between measured data and the data that has been predicted.

| Location A | Location B |
|------------|------------|
| Measured   | Predicted  | Measured | Predicted |
| 9.7600     | 7.2266     | 12.4500  | 10.7814   |
| 11.1500    | 11.1639    | 13.2800  | 16.5226   |
| 9.4100     | 9.6914     | 10.8700  | 12.5758   |
| 20.3100    | 27.4996    | 23.4100  | 27.7079   |
| 14.1500    | 16.5076    | 16.4000  | 16.6126   |
| 11.1600    | 13.1882    | 14.6700  | 17.5048   |
| 25.7700    | 28.0773    | 26.7200  | 26.9025   |
| 11.0500    | 15.0568    | 13.0900  | 18.1812   |
| 15.7100    | 18.9989    | 18.7300  | 21.7490   |
| 29.1300    | 26.5261    | 29.7200  | 28.7100   |
| 16.1200    | 21.9191    | 19.7500  | 25.4696   |
| 18.4700    | 14.6661    | 24.2000  | 17.3267   |
| 17.3000    | 17.8787    | 20.5900  | 18.6698   |
| 10.6800    | 10.8911    | 15.3600  | 15.0643   |
| 9.4100     | 9.6914     | 11.8800  | 12.1505   |
| 18.3300    | 17.9019    | 23.6900  | 21.4363   |
| 28.5300    | 31.6607    | 29.1700  | 31.7195   |
| 23.9200    | 27.4879    | 25.1000  | 28.4839   |
| 22.7500    | 25.9788    | 25.4400  | 27.9650   |
| 28.6500    | 25.1371    | 28.5400  | 26.1770   |
| 23.3100    | 24.1209    | 23.7400  | 23.5733   |
| 9.2200     | 9.3178     | 11.1200  | 13.7551   |
| 27.8600    | 29.6974    | 28.3800  | 30.1291   |
| 10.0300    | 9.5687     | 14.0200  | 11.9338   |
| 19.3600    | 21.3450    | 22.8300  | 21.6379   |
| 13.0000    | 12.4892    | 17.4900  | 12.7604   |
| 28.6400    | 28.1706    | 28.1900  | 27.0080   |
| 15.1000    | 19.4982    | 19.4900  | 22.1704   |
| 17.5700    | 16.3249    | 21.4200  | 17.6043   |
| 22.0300    | 21.9660    | 24.1500  | 23.6943   |
6. Conclusion
In this paper, a linear regression method using climate variables (minimum, maximum and average) of the temperature values to estimate monthly average daily of the GSRa from June 2003 to December 2012, in two locations in Egypt with two different climates have been applied. The RMSE for both cases were (2.7716 and 2.9392) MJ m$^{-2}$ in locations A and B respectively. The low values of RMSE parameters showed that the linear regression method with temperature values can be very efficient technique and best proposed solution to reduce the costs of calculating the amounts of solar energy.

Acknowledgment
The author wishes to thank the Ministry of Education, Iraq and Faculty of Science, Zagazig University, Dr. Zahraa El Slayed Mohammed for the great support for this paper. Also, thanks for the Informatics Research Institute, New Borg El Arab City, Alexandria, Egypt for providing the weather and solar data.
Reference

[1] El-Metwally, M. (2005). Sunshine and global solar radiation estimation at different sites in Egypt. *Journal of Atmospheric and Solar-Terrestrial Physics*, 67(14), 1331-1342. https://doi.org/10.1016/j.jastp.2005.04.004

[2] Tao, H., Ewees, A. A., Al-Sulttani, A. O., Beyaztas, U., Hameed, M. M., Salih, S. Q., ... and Yaseen, Z. M. (2021). Global solar radiation prediction over North Dakota using air temperature: development of novel hybrid intelligence model. *Energy Reports, 7*, 136-157. https://doi.org/10.1016/j.egyr.2021.110996

[3] Mohammádi, B., and Moazenzadeh, R. (2021). Performance Analysis of Daily Global Solar Radiation Models in Peru by Regression Analysis. *Atmosphere, 12*(3), 389. https://doi.org/10.3390/atmos12030389

[4] Ağbulut, Ü., Gürel, A. E., and Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews, 135*, 110114. https://doi.org/10.1016/j.rser.2020.110114

[5] Liu, J. N., and Hu, Y. X. (2013). Support Vector Regression with Kernel Mahalanobis Measure for Financial Forecast. In *Time Series Analysis, Modeling and Applications* (pp. 215-227). Springer, Berlin, Heidelberg. Díaz–Vico, D., Torres–Barrán, A., Omari, A., and Dorronsoro, J. R. (2017). Deep neural networks for wind and solar energy prediction. *Neural Processing Letters, 46*(3), 829-844. https://doi.org/10.1007/s11063-017-9613-7

[6] Khosravi, A., Koury, R. N. N., Machado, L., and Pabon, J. J. G. (2018). Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production, 176*, 63-75. https://doi.org/10.1016/j.jclepro.2017.12.065

[7] Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L., and Cui, N. (2019). Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Conversion and Management, 198*, 111780. https://doi.org/10.1016/j.enconman.2019.111780

[8] Gürel, A. E., Ağbulut, Ü., and Biçen, Y. (2020). Assessment of machine learning, time series, response surface methodology and empirical models in prediction of global solar radiation. *Journal of Cleaner Production, 277*, 122353. https://doi.org/10.1016/j.jclepro.2020.122353

[9] Tao, H., Ewees, A. A., Al-Sulttani, A. O., Beyaztas, U., Hameed, M. M., Salih, S. Q., ... and Yaseen, Z. M. (2021). Global solar radiation prediction over North Dakota using air temperature: development of novel hybrid intelligence model. *Energy Reports, 7*, 136-157. https://doi.org/10.1016/j.egyr.2020.11.033

[10] El-Metwally, M. (2004). Simple new methods to estimate global solar radiation based on meteorological data in Egypt. *Atmospheric Research, 69*(3-4), 217-239. https://doi.org/10.1016/j.atmosres.2003.09.002

[11] El Mghouchi, Y., El Bouardi, A., Sadouk, A., Fellak, I., and Ajzoul, T. (2016). Comparison of three solar radiation models and their validation under all sky conditions–case study: Tetuan city in northern of Morocco. *Renewable and Sustainable Energy Reviews, 58*, 1432-1444. https://doi.org/10.1016/j.rser.2015.12.354

[12] Chen, J. L., Li, G. S., and Wu, S. J. (2013). Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. Energy conversion and management, 75, 311-318. https://doi.org/10.1016/j.enconman.2013.06.034

[13] Czȋbula, G., Czȋbula, I. G., and Gaceanu, R. D. (2014). A support vector machine model for intelligent selection of data representations. *Applied Soft Computing, 18*, 70-81. https://doi.org/10.1016/j.asoc.2014.01.026
[14] David, M., Diagne, M., and Lauret, P. (2012, May). Outputs and error indicators for solar forecasting models. *In Proceedings of the World Renewable Energy Forum (WREF)* (pp. 13-17).

[15] McCloskey, A. (2020). Asymptotically uniform tests after consistent model selection in the linear regression model. *Journal of Business and Economic Statistics, 38*(4), 810-825. https://doi.org/10.1080/07350015.2019.1592754

[16] Ajayi, O. O., Ohijeagbon, O. D., Nwadialo, C. E., and Olasope, O. (2014). New model to estimate daily global solar radiation over Nigeria. *Sustainable Energy Technologies and Assessments, 5*, 28-36. https://doi.org/10.1016/j.seta.2013.11.001

[18] Rusen, S. E., and Konuralp, A. (2020). Quality control of diffuse solar radiation component with satellite-based estimation methods. *Renewable Energy, 145*, 1772-1779. https://doi.org/10.1016/j.renene.2019.07.0