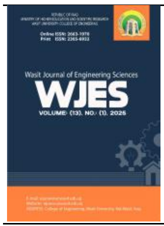


Wasit Journal of Engineering Sciences
Journal homepage: https://ejuow.uowasit.edu.iq

ISSN 2305-6932
Original Research

Vol. 14 No. 2 June 2026, pp.260-276
https://doi.org/10.31185/ejuow.v14i2.862

Statistical and Experimental Assessment of Solar Irradiance Variability in Kuala Lumpur: Comparing Driest and Wettest Months in 2023

Sajad Mohammed Aboud

Republic Islamic of Iran, Tabriz, Tabriz University, Mechanical Engineering College, Mechanical Department

Corresponding Author Email: Sajad.m.aboud@tabrizu.ac.ir

Received Dec.22, 2025 Revised Jan.17, 2026 Accepted Jan.31, 2026 Online Jun.1, 2026

ABSTRACT

The experimental data is obtained in regional Kuala Lumpur region via three stages. First, develop a statistical model using real-world meteorological data, such as sunlight hours, precipitation/rainfall, and relative humidity, to identify the highest and lowest relative monthly sunshine durations. The solar panel was used to directly measure sun irradiation in June and November 2023, from 9:30 am to 4:30 pm. Checking the statistical model's accuracy is the final step. Research indicated that June had the highest relative sunshine duration (51.7%) and mean daily irradiance (≈619 W/m²), while November had the lowest (46.4%) and mean irradiance (≈560 W/m²). Temperature and day duration vary slightly across annual months, although moisture and cloud cover vary more significantly. The model-estimated decline from June to November (≈10.3%) matches the observed irradiance-based reduction (mean ≈9.4%, range 8.9-9.85%), indicating good agreement within experimental uncertainties. Smoothing the irradiance time series (5-10 min intervals) reduced absolute variability but did not significantly affect the relative decline between months (≈8.7-9.8%), making the comparison robust. Validation metrics and tests (KS and cross-validated regression diagnostics in Methods) confirm the statistical model's reliability for monthly assessments in this tropical urban setting. KL's monthly and seasonal rainfall fluctuations result in a 9-11% variation in solar energy between the driest and wettest months. This variability is significant in system design, financial planning, yield forecasts, reserve capacity planning, and battery size for rooftop and distributed PV. The provided regression model and real-time irradiance measurements agree within experimental error, suggesting that the model may be a helpful tool for monthly-scale PV yield estimate and planning when combined with additional local meteorological inputs and multi-site validation.

Keywords: Solar Energy Irradiation, Meteorological Parameters, Statistical and Regression Model, Kuala Lumpur Region, Annual Highest/Lowest Irradiation

1. Introduction

Solar energy is the most abundant renewable energy source. However, obtaining this solar energy is difficult due to climatic and geological conditions, which are primarily responsible for these solar modules' low performance. Despite a lack of data and studies on how solar irradiance intensity and daylight length affect solar power, recent attention has shifted to the kind of solar cells and their positioning systems. To assess solar technologies and optimize the deployment of solar energy systems, climatic data from a typical or standard period and a thorough knowledge of the distribution and quantity of solar resources are required. A detailed understanding of solar radiation and meteorology is essential for selecting technologies, evaluating, and modelling solar technology performance, planning installation sizes, and obtaining finance. This research is being conducted in Kuala Lumpur, Malaysia's capital, at latitude and longitude coordinates 3.140853 and 101.693207.

A few studies have looked at typical general meteorological data for Malaysia and Kuala Lumpur. Muhammad et al. [1] conducted the most complete study for Subang (near Kuala Lumpur), calculating the mean daily data for a model year by selecting typical months from 21 years of data. They discovered an annual dry bulb temperature of 27.6°C, a relative humidity of 83%, a 7-okta cloud cover, an average wind speed of 1.2 m/s, and worldwide daily solar radiation of 16.4 MJ/m^2 (4.56 kWh/m^2).

Early solar radiation studies employed four monitors, fluctuations inferred from radiation at the top of the atmosphere, and regression analysis to calculate sunlight hours. Chuah and Lee [2] estimated monthly average daily global solar radiation levels in Malaysia by evaluating sunlight duration data from 12 cities between 1961 and 1975. They found daily values ranging from 3.5 to 6 kWh/m^2 in northern and east coast towns, and 4 to 5 kWh/m^2 in southern cities. Rosato et al. [3] utilized linear regression to link worldwide solar radiation and sunlight duration, revealing comparable seasonal daily values ranging from 3.56 kWh/m^2 in December to 7.53 kWh/m^2 in April. They also discovered that variations in global solar radiation are primarily due to rainfall, which is affected by prevailing winds, local land/sea breezes, terrain, and convection. Sopian and Othman [4] used a larger worldwide network of radiation detectors at eight locations. The daily average values for the eight cities ranged from 3.3 kWh/m^2 in Kuching and 5.5 kWh/m^2 in Bayan Lepas.

Azhari et al. [5] combined satellite and ground station cloud cover data to estimate the lowest daily solar radiation at 4.9 kWh/m^2 and the highest at 5.2 kWh/m^2. Compared to prior research, they discovered that the average daily worldwide solar radiation rose from 4.8 kWh/m^2 in 1982 to 4.965 kWh/m^2 in 2006. While the studies above produced comparable yearly average findings, substantial regional and seasonal variance exists. It may significantly influence the operations and economics of a solar power plant.

Pollution may also diminish sun radiation levels, as smoke and haze have been shown to decrease UV and global solar energy by 16-23% (Engel-Cox et al.) [6]. Liew et al. [7] utilized principal component analysis to find four unique regimes of particulate matter less than 10 microns in diameter (PM10) in Malaysia. For example, the Kuala Lumpur regime has a distinct seasonal pattern in PM10 concentrations, with peak levels occurring during the summer monsoon.

Kuala Lumpur has a tropical rainforest climate, with annual precipitation ranging between 2,400 and 2,700 mm and 180 to 200 wet days yearly. Meteorological data showed a slight seasonal fluctuation in temperature and day duration. The appearance of clouds often peaks in the afternoon. Precipitation is a proxy for deep convective cloud cover, which lowers surface shortwave irradiance by increasing optical thickness and improving multiple scattering. This study investigates the variations in PV energy output between the "monthly driest month" and the "monthly wettest month" circumstances in KL. The emphasis is on daily totals to ensure consistency in energy yield and operational planning.

The tropical atmosphere is ideal for solar energy production with its constant sunlight. However, unpredictability in precipitation influences solar radiation levels, which are critical for the efficient operation of a PV system. Monthly precipitation changes substantially impact solar radiation, with lower levels seen during the wet season (Engel-Cox et al.) [6]. As a result, Tsung et al. [8] demonstrated that the Angstrom-Prescott model accurately correlates meteorological phenomena like rainfall and cloud cover with power radiation. Despite the government's efforts to encourage renewable energy, the high initial cost of PV systems remains a hurdle. Tan et al. [9] propose that sophisticated forecasting models may improve solar radiation prediction, allowing for more effective solar technology design and deployment.

As a result, a thorough model is necessary to investigate how the sun's movement influences the operation of the solar module system. The simulation model found that daylight length is more essential than available solar irradiance. Furthermore, the greater the sun irradiation and daylight length, the more electricity the solar module produces. The length of daylight is also impacted by latitude, with higher latitudes having more extended daylight [10].

2. Methodology

2.1. Experimental Work

The experimental procedure was conducted in three chronological parts. The first part involves developing a statistical model to determine the highest and lowest relative sunshine durations using empirical meteorological data, including solar hours, precipitation/rainfall, and relative humidity. The second part was performed experimentally by directly measuring solar irradiation using a solar photovoltaic panel, as shown in Figure 1, during June and November 2023, between 9:30 am and 4:30 pm. The third part involves testing the accuracy of the statistical model.

Figure 1: Photovoltaic/thermal collector

2.2. Location and Solar Geometry

The interaction of precipitation, rainfall, and daily sunlight hours in Kuala Lumpur (Figure 2) is heavily impacted by urbanization and its accompanying climate consequences. Cloudiness and precipitation are more common in urban environments, reducing sunlight availability. Furthermore, greater local temperatures promote convective processes, increasing precipitation by more than 30% in urban areas [11]. The increase in rainfall is associated with an increase in cloud cover, notably during the southwest and northeast monsoons, when urban areas have 1.22 to 1.24 times more cloud cover than rural areas (Syed Mahbar & Kusaka) [12]. In addition, the amount of sunlight hours decreases because clouds block solar energy. Statistical models show a substantial link between sunlight length and cloud cover, with precipitation also having an important impact (Zateroglu)[13].

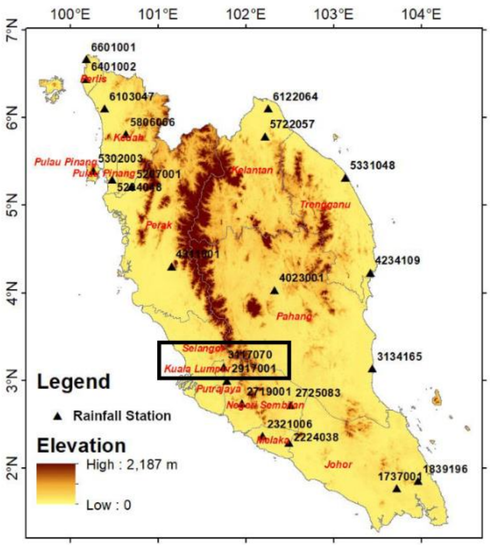


Figure 2. Map of Kuala Lumpur (Mohamed Salem Nashwan et.al)[14]

2.3. Meteorological Data Collection

In this study, six meteorological parameters that affect the amount of solar power were chosen to get the highest and lowest relative percentage monthly sun hours. These parameters include the minimum temperature, maximum temperature, precipitation/rainfall, relative humidity, and average sun hours, as shown in Table 1.

Table 1. Meteorological parameters that affect the gathered data of the solar power

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
Avg. Temperature °C	25.1	25.7	26	26.1	26.3	26.2	26.1	26	25.8	25.7	25.2	25.1
Min. Temperature °C	21.9	21.9	22.7	23.1	23.4	23.1	22.8	22.8	22.8	22.8	22.7	22.3
Max. Temperature °C	29.1	30.2	30.3	30.1	30	30	29.9	29.8	29.7	29.5	29	28.8
Precipitation / Rainfall mm	209	174	268	300	246	174	183	219	243	308	373	284
Humidity(%)	85%	82%	85%	87%	87%	85%	84%	84%	85%	87%	90%	88%
Rainy days (d)	15	13	18	20	19	17	18	18	19	19	20	18
avg. Sun hours (hours)	7.5	8.0	7.8	7.4	7.7	8.2	8.1	7.9	7.7	7.6	6.5	7.0

2.3.1. Minimum and maximum temperature

In Kuala Lumpur's regional tropical climate, minimum and maximum temperatures are related to solar hours, which have a non-direct impact on humidity. Kuala Lumpur has consistently high temperatures, ranging from 28.8 °C (December) to 30.3 °C (March) (Nucifera et al.)[15]. Meanwhile, Ooi et al. [16] studied the impact of urbanization on temperatures in Kuala Lumpur and found that the average daily temperature rose by 0.9 °C. High temperatures and humidity increase cloud cover, reducing relative sunshine hours (Aktas et al.)[17]. Conversely, although high temperatures may reduce relative sunshine hours due to increased cloudiness, they may also increase solar radiation on clear days, revealing a complex dynamic that warrants more investigation.

2.3.2. The Length of daylight

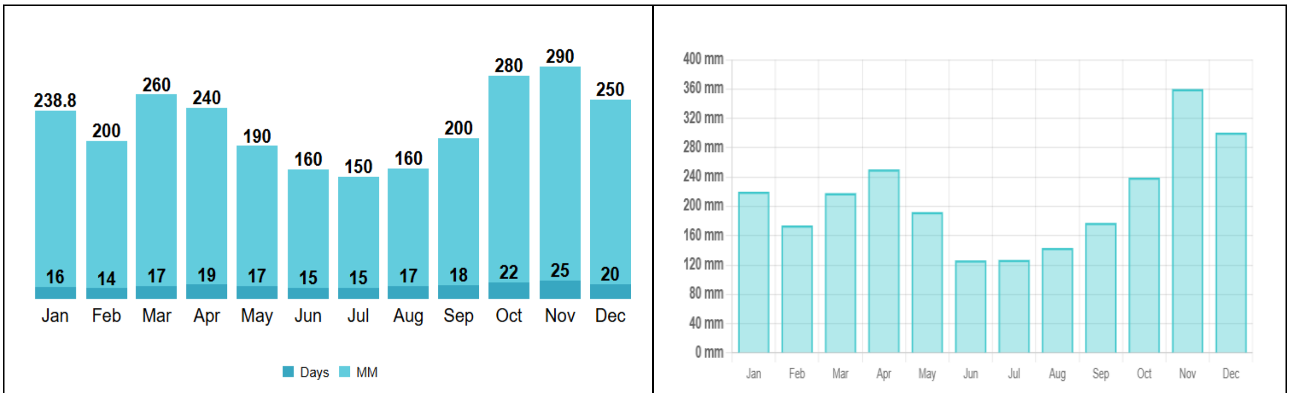
The length of daylight is different from the monthly average sun hours due to the nature of Malaysian weather, which is influenced by its location in a sub-tropical region. However, for the two months chosen (June and November) in the previous paragraph, the length of the daylight is 12:20 h and 12:01 h, respectively, as shown in Table 2. Hence, the insignificant difference has no effect; in reality, this difference has other impacts on the availability of sunlight during the day (Rajendran & Smith)[10].

Table 2. Monthly sunrise and sunset for Kuala, Lumpur, Malaysia (Rajendran & Smith)[10].

Month	Sunrise (am)	Sunset (pm)	Day Length (h)
January	07:22	07:22	12:00
February	07:25	7:29	12:03
March	7:18	7:26	12:08
April	7:06	07:19	12:13
May	07:00	07:18	12:18
June	07:03	07:23	12:20
July	07:09	07:28	12:19
August	07:10	07:25	12:15
September	07:03	07:13	12:10
October	06:56	07.01	12:05
November	06:57	06:58	12:01
December	07:08	07:07	11:59

2.3.3. Monthly Average Rainfall and Precipitation

The impact of rainfall on solar energy output in Kuala Lumpur varies dramatically between the peak and the lowest precipitation months. Various climatic parameters, including rainfall severity and precipitation, impact this connection, which affects solar panel efficiency and total energy production. Engel-Cox et al. [6] discovered that increased sun hours and precipitation diminish solar energy availability, resulting in an annual average of 3.73 to 5.11 kWh/(m²d). The tilt and direction of solar panels also impact their best performance, which may reduce the effects of sunshine hours. Kuala Lumpur's metropolitan layout increases localized rainfall, possibly impacting solar energy output (Ghielmini et al.)[18]. The fluctuation of rainfall patterns, especially during monsoon seasons, substantially influences solar energy output, with monthly rainfall showing larger coefficients of variation (Mohamed & bin,)[19]. Rainfall typically lowers solar radiation but may also improve solar panel performance by cleaning surfaces and increasing efficiency. This dichotomy emphasizes the intricate relationship between climatic conditions and solar energy generation in metropolitan areas such as Kuala Lumpur. Figure 3 shows a schematic representation of rainfall (a) and precipitation (b).



(a) Rainfall in Kuala Lumpur (b) Precipitation in Kuala Lumpur

Figure 3. (a) Rainfall in Kuala Lumpur (Mohamed Salem et.al)[14] and (b) Precipitation in Kuala Lumpur (Ghielmini et al.)[18]

The data shows that the rainfall and precipitation in Kuala Lumpur for 2023 for June are 160 mm and 135 mm, respectively, and those for November are 290 mm and 360 mm, respectively.

2.3.4. Relative Humidity

The influence of relative humidity on solar energy in Kuala Lumpur, especially during the peak and lowest precipitation months, provides essential information on solar radiation and photovoltaic (PV) performance. According to the data, relative humidity and precipitation levels directly impact solar energy output, with varied effects depending on the month. Wazwaz and Khan [20] found that the average relative humidity is greatest in November (90%) and lowest in February (82%). Direct scattering and loud cover reduce sunlight (Engel-Cox et al.)[7]. Higher humidity may cause more dust deposition, which reduces energy production (Sengupta et al.)[21].

2.3.5. The Relative Sunshine Duration

The relative sunlight duration (Y) was employed in this investigation since it is recommended in scientific literature over measured values (Takilalte et al.[22] ; Gouda et al.,[23]). The values of Y (between 0 and 1) define the ratio of recorded sunlight duration to day length, which is measured in hours per day.

2.4. Real-Time Data Collection

The data collection was accurately instrumented throughout the daylight between 9:30:00 and 16:30:00, utilizing a pyranometer close to the photocell PV (Lai)[7]. The data was collected between June and November 2023. The two months were chosen using the statistical model for the relative. The aim for choosing June and November was to produce a representative sample while capturing the seasonal extremes in sunshine duration and weather fluctuations between the driest and wettest months (Solanki et al.)[24]. The data collection considers the atmospheric conditions typically seen in the area. Meanwhile, climate change and sun radiation were the most important components in the data gathering procedure. To counteract these variations, an accurate smoothing method was used, measuring the irradiance at 5- and 10-minute intervals to ensure future studies' correctness (Coddington et al.)[25].

3. Irradiance Modeling

3.1. The Kolmogorov–Smirnov (KS) one-sample test

The One-Sample Kolmogorov-Smirnov (KS) Test is a nonparametric statistical technique for determining if a sample belongs to a specific continuous distribution. The KS test calculates the most significant difference between the sample's empirical distribution function and the cumulative distribution function of the hypothesised distribution. This test is very beneficial for determining goodness-of-fit without requiring a specific distribution form (Zheng et al.)[26]. The test compares the agreement of two sets of values based on a randomly selected sample and an empirical frequency distribution of the observed and empirical samples. The test employs the most significant divergence to calculate a two-tailed probability estimate p, determining whether the samples are statistically similar or dissimilar.

3.2. Linear Regression Model

A linear regression model for relative sunlight duration in Kuala Lumpur to compare irradiation between June and November 2023 is developed in terms of sun hours, precipitation, rainfall, relative humidity, maximum temperature, and lowest temperature. The regression analysis approach has been widely used in forecasting climatological variables. The relationships between the dependent variable and other variables may be represented mathematically using linear regression (Tyagi et al.)[27]. The determination coefficient R2 is a scale that indicates how well the estimated model fits the observed data. The measure R2 represents the calculated value as a percentage and ranges from 0 (indicating that the dependent variable is unexplained by the independent variables) to 1 (indicating that the dependent variable is explained by the independent variables at the maximum degree). Equation (1) is used to compute the multiple determination coefficient for the statistical model.

$$R^2 = 1 - \frac{\sum(Y_e - \bar{Y})^2}{\sum(Y_o - \bar{Y})^2} \quad (1)$$

Where the term $\bar{Y}$ denotes the average value of the estimated values Y_e, the phrase Y_o expresses the observed value. The value 1 for obtained R2 determines that the model is consistent with the dataset in high level, but the value 0 denotes the model is inconsistent with the data. The second statistical term is the standard error estimation (SEE), which interprets the accuracy of the constructed models by determining the difference between predicted and observed values. This expression indicates that the amount of the deviation the measured values from the estimated values on average and computed by the formula as shown in Equation (2),

$$SEE = \sqrt{\frac{\sum(Y_o - Y_e)^2}{n-2}} \quad (2)$$

where the term Y_o defines the observed value, the phrase Y_e denotes the estimated value and n depicts the number of observations.

3.3. Modeling Weather and Solar Energy

The hourly daylight availability of the Malaysian sky was computed using daylight modelling techniques based on empirical and observed solar irradiance and rain days (Zain-Ahmed et al.)[28]. Global illuminance is normally high, reaching 80,000 lux at noon during peak solar irradiation. Even in months when the ground receives less solar radiation, the peak illuminance may exceed 60,000 lux. Such data has applications and uses in daylighting design, including visual and thermal comfort, task illumination, and energy-efficient building design. The Perez sky diffuse model has been validated multiple times; nevertheless, research has indicated that the estimate's accuracy may be improved by localizing model coefficients and calibrating embedding coefficients using EnergyPlus' source code (Meshkinkiya et al.)[29]. The diffuse percentage on the wettest days reaches 0.9-1.0 utilizing Orgill-Hollands type relationships, whereas on the driest days it varies between 0.2-0.4 at midday (Jahangir et al.)[30].

The experimental setup and data gathering were carried out in the Kuala Lumpur region because of its diverse locations and careful selection to provide the best circumstances for solar systems in tropical climates. The data collection was meticulously instrumented using a pyranometer to estimate the time and location based on the highest available solar radiation. The pyranometer was put as close to the photocell PV as possible per previous suggestions by Lai [7]. The data was collected in the afternoon, when the time period was carefully selected to account for the boundaries of when solar radiation was most relevant (Zhang & Chen)[31]. The monthly average of each meteorological indicator was derived to compare the wettest and driest months. The data collection demonstrates real-world variability for three weather scenarios: clear sky, partial cloud cover, and overcast circumstances. The three scenarios reflect the behavior of air conditions often seen in the region.

3.4. Building Model

Specifying the variables is the first stage in developing a linear regression model to predict relative sunlight duration (output) using sun hours (h), precipitation and rainfall (mm), relative humidity (%), and maximum and lowest temperatures (°C). The second phase is data gathering, which may be accomplished by regularly using NASA sources from Malaysian authorities. The data should be cleaned to deal with the missing values. Model-based imputation is utilized to handle minor gaps to complete the cleaning phase, while specialized imputation algorithms are employed for larger gaps. To verify consistency, set rainfall/precipitation to values ranging from 0 (no rain) to the actual rainfall (mm). Pearson statistics are used to determine correlation, with a negative correlation with cloud cover and humidity and a positive correlation with maximum temperature. Multicollinearity is used to calculate the Variance Inflation Factor (VIF), which evaluates the severity of multicollinearity in regression analysis on a scale of 5 to 10. Rainfall data is usually biased. The cloud cover in Oktas varies from 0 to 1. The data is divided into time series using K-fold cross-validation (CV) based on the original sample, which is randomly partitioned into k equal-sized subsamples, sometimes known as "folds".

3.5. Model Choice

The Ordinary Least Squares is an excellent place to start when choosing the appropriate model since its interpretation is straightforward. However, machine learning techniques may be used for multicollinearity or overfitting: Ridge to stabilize coefficients, Lasso for feature selection, and Elastic Net (combination). The root mean square error (RMSE) statistic evaluates a model's short-term performance by comparing the actual difference between estimated and measured values over time. The lower the value, the better the model works. The mean absolute error (MAE) is then used to determine the differences between paired observations that reflect the same event, such as two variables y and x to compare expected vs observed, subsequent time compared to the starting time, and one measurement method compared to another. The third metric is the coefficient of determination (R2), a statistical measure in a regression model that reflects how much variation in the dependent variable can be explained by the independent variable. In other words, the R-squared value indicates how well the data fits the regression model. The last statistic is the mean absolute percentage error (MAPE), which determines the average magnitude of error produced by a model, or how far off predictions are on average. The MAPE score of 20% means that the average absolute percentage difference between predictions and the actual value.

3.6. Define the Regression Model

Regression is the method of connecting two or more variables of interest to provide an instrument for prediction or forecasting. The approach used to do regression analysis aids in determining which characteristics are essential, which parts are overlooked, and how a researcher promotes all of them. Regression relies on supervised machine learning algorithms in predicting data patterns with usually one dependent variable (DV) and one or more independent variables (IV). A collection of machine learning algorithms that serve as a framework for illustrating interpretations of correlations between consistent variables. These processes are applicable in almost all investigations, with the common practical estimate being the most routine of all data analysis procedures. In machine learning, predictions are made based on three parameters: independent variable count, regression form line, and dependent variable type (Khan et al.)[32].

The data sources are available from Malaysia Meteorological Department (MetMalaysia) or global datasets like NOAA, NASA POWER, ERA5). The dependent variable (prediction) is the relative sunshine duration which characterizes the efficiency of the solar power. The independent variables are sunshine hours (h), precipitation (mm), rainfall (mm), relative humidity (%), maximum temperature (°C), and minimum temperature (°C). Hence, the regression model of the prediction of relative sunshine (Y measured as hours/day) depends the linear function of the predictors including the intercept (β_0), sunshine hours ($\beta_{Sunshine\ Hours}$), precipitation ($\beta_{Precipitation}$), rainfall ($\beta_{Rainfall}$), humidity ($\beta_{humidity}$), maximum temperature (β_{Tmax}), minimum temperature (β_{Tmin}), and error term (ε), as depicted inequation 3:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \qquad \ldots\ldots\ldots.(3)$$

Where:

- Y = Relative sunshine duration (%), X_1 = Sunshine hours (h). X_2 = Precipitation (mm), X_3 = Rainfall (mm), X_4 = Relative humidity (%), X_5 = Maximum temperature (°C), X_6 = Minimum temperature (°C), β_0 = Intercept, β_1 through β_6 = Regression coefficients, ε = Error term

The building of the regression model requires cleaning missing values, standardization of (normalization) of the predictors, and checking the correlation between the input variables.

3.7. Implementing the Code

The model can be diagnosed according to the following criteria: the significance (p-value) which is the probability that the observed effect within the study that is considered statistically significant between $p<0.05$ or $p<0.01$, multicollinearity test by the variance inflation factor (VIF)which measures the severity of multicollinearity in regression analysis, indicating an increase in the variance of a regression coefficient as a result of collinearity, and residual plots. The expected insights of the independent variables show that the higher sunshine hours, precipitation, rainfall, and lower sunshine duration, the higher Tmax is likely related to more sunshine (inverse with cloudiness), and the higher Tmin correlates with humidity/cloudiness at night. The regression model was implemented using Python's scikit-learn library, employing Ridge regression with cross-validation for coefficient optimization. Time-series split validation was used to assess model performance. Detailed implementation code, including data preprocessing, feature engineering, and model validation procedures, is provided in Appendix A.

4. Results and Discussion

4.1. Monthly Sunshine Hours

The daily sunshine hours depend on the day length, which is almost the same in Kuala Lumpur because of the geographical location. The interaction of precipitation, rainfall, and daily sunlight hours in Kuala Lumpur is heavily impacted by urbanization and its accompanying climate consequences, as shown in Table 3. Cloudiness and precipitation are more common in urban environments, reducing sunlight availability. Furthermore, greater local temperatures promote convective processes, increasing precipitation by more than 30% in urban areas (Argüeso et al.)[11]. The increase in rainfall is associated with an increase in cloud cover, notably during the southwest and northeast monsoons, when urban areas have 1.22 to 1.24 times more cloud cover than rural areas (Syed Mahbar & Kusaka)[12]. In addition, the amount of sunlight hours decreases because clouds block solar energy. Statistical models show a substantial link between sunlight length and cloud cover, with precipitation also having an important impact (Zateroglu)[13].

Table 3. Meteorological and Solar Parameters for Kuala Lumpur, Malaysia (2023)

Month	Day Length (h:mm)	Solar Noon	Precipitation (mm)	Rainfall (mm)	Sunshine Hours (h)	Relative Sunshine Duration (%)
January	12:00	01:22 PM	220	238	7.5	51.8
February	12:03	01:27 PM	170	200	8.0	49.7
March	12:08	01:22 PM	222	260	7.8	47.0
April	12:13	01:13 PM	250	240	7.4	48.9
May	12:18	01:09 PM	190	190	7.7	50.3
June	12:20	01:19 PM	125	160	8.2	51.7
July	12:19	01:17 PM	128	150	8.1	47.8
August	12:15	01:18 PM	140	160	7.9	51.5
September	12:10	01:17 PM	185	160	7.7	48.3
October	12:05	01:08 PM	240	280	7.6	46.9

November	12:01	12:59 PM	360	290	6.5	46.4
December	11:59	01:08 PM	298	250	7.0	47.6

4.2. Kolmogorov-Smirnov Test Results

The statistical models depend on the parameters mentioned in Equation (3). The data are taken from the sources highlighted in the previous sections. These parameters include sunshine, precipitation, rainfall, relative humidity, and minimum and maximum temperatures as depicted in Table 4. According to Zateroglu [13], the rainfall and precipitation substantially impact sunlight duration, reducing solar energy. The average sunshine hour in Kuala Lumpur ranges between 6.5 h in November and 8.2 h in June. The sunshine hours are calculated previously as depicted in Table 3. The regression statistical model shows that the dependent variable, the relative sunshine duration, is maximum at 51.7% in June and minimum at 46.4% in November. The corresponding minimum and maximum temperatures in June and November are 23.1 (30.0 °C) and 22.7 (29.0 °C), respectively.

Table 4. Parameters required for the model

Date (2023)	Independent Variables						Dependent Variable
	Sun shine hours (h)	Precipitation (mm)	Rainfall (mm)	Relative Humidity (%)	Tmin (°C)	Tmax (°C)	Relative Sunshine Duration (%)
January	7.5	220	238	85	21.9	29.1	51.8
February	8.0	170	200	82	21.9	30.2	49.7
March	7.8	222	260	85	22.7	30.3	47.0
April	7.4	250	240	87	23.1	30.1	48.9
May	7.7	190	190	87	23.4	30.0	50.3
June	8.2	125	160	85	23.1	30.0	51.7
July	8.1	128	150	84	22.8	29,9	47.8
August	7.9	140	160	84	22.8	29.8	51.5
September	7.7	185	160	85	22.8	29.7	48.3
October	7.6	240	280	87	22.8	29.5	46.9
November	6.5	360	290	90	22.7	29.0	46.4
December	7.0	298	250	88	22.3	28.8	47.6

4.3. Real-Time Data Collection

The primary focus is on solar irradiance, a critical parameter in solar energy research. Researchers have consistently noted that one of the most challenging aspects of solar energy utilisation is accounting for weather-related changes, including rain, dust, wind speed, azimuthal angles, and geographical location. Addressing these challenges necessitates the estimation of average solar irradiance over reasonable time intervals. The data collection was instrumented precisely during the daytime, 9:30:00 and 16:30:00, at an interval of 1.8457 seconds during June and November 2023. The two months were selected on the statistical model for the highest and lowest relative sunshine duration. For this work, the months of June 2023 and November 2023. Figure 4 shows the average of the irradiance during June 2023, while the subsequent smoothing at 5-minute and 10-minute intervals is shown in Figures 5 and 6, respectively.

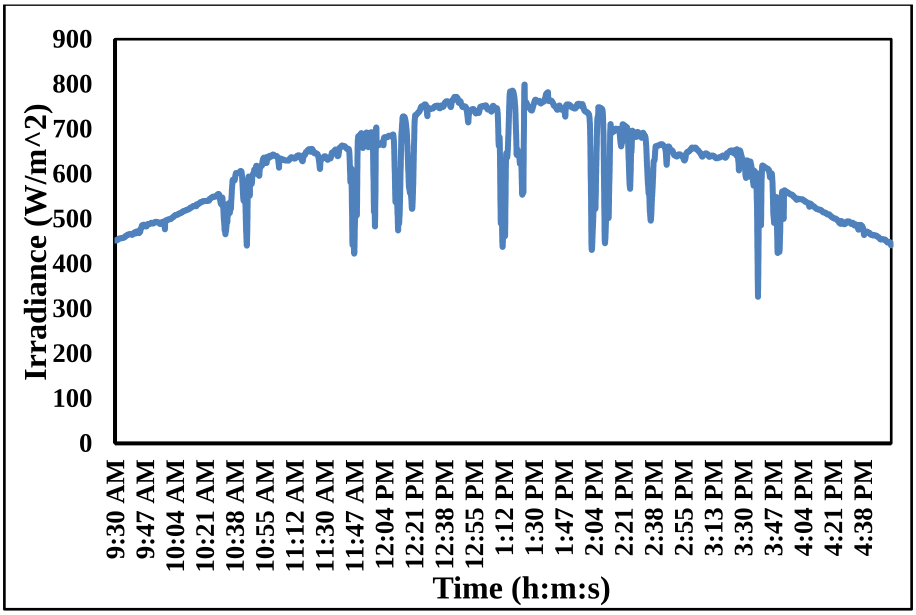


Figure 4. The average irradiance during June 2023

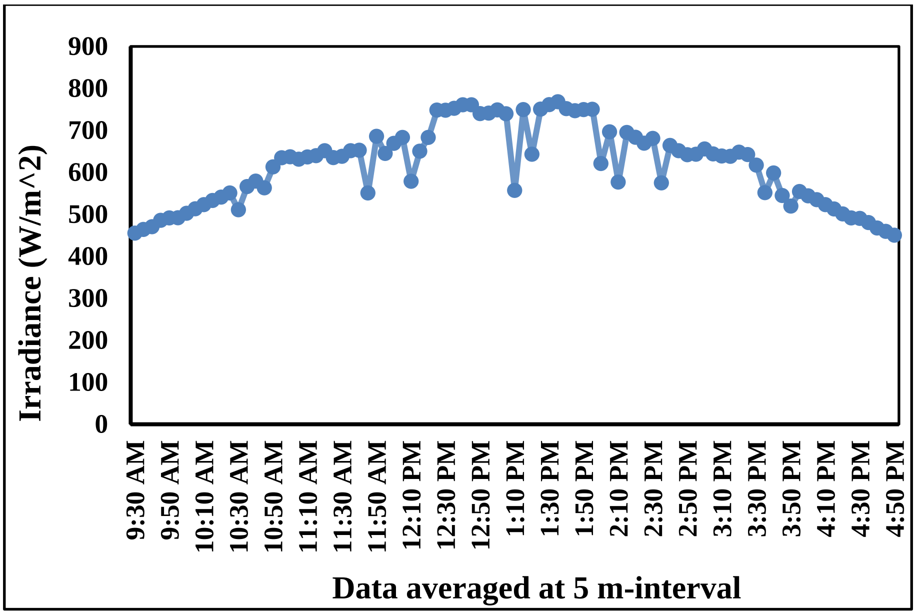


Figure 5. Smoothing of the irradiance at interval of 5 minutes

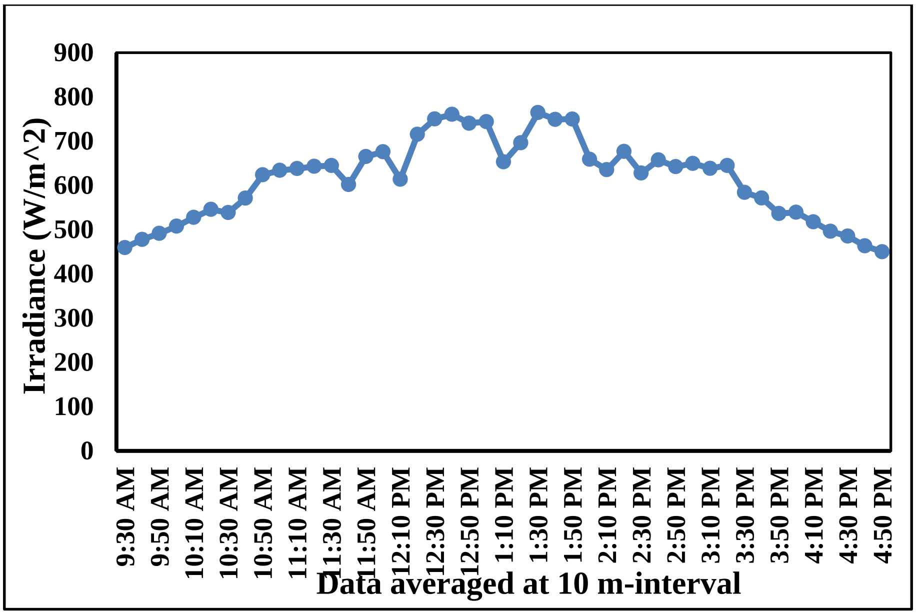


Figure 6. Smoothing of the irradiance at interval of 10 minutes

The same procedure was performed for November 2023 and the original result is shown in Figure 7. The subsequent analyses for smoothing the average irradiance at 5-minute and 10-minute intervals is shown in Figure 8 and 9, respectively.

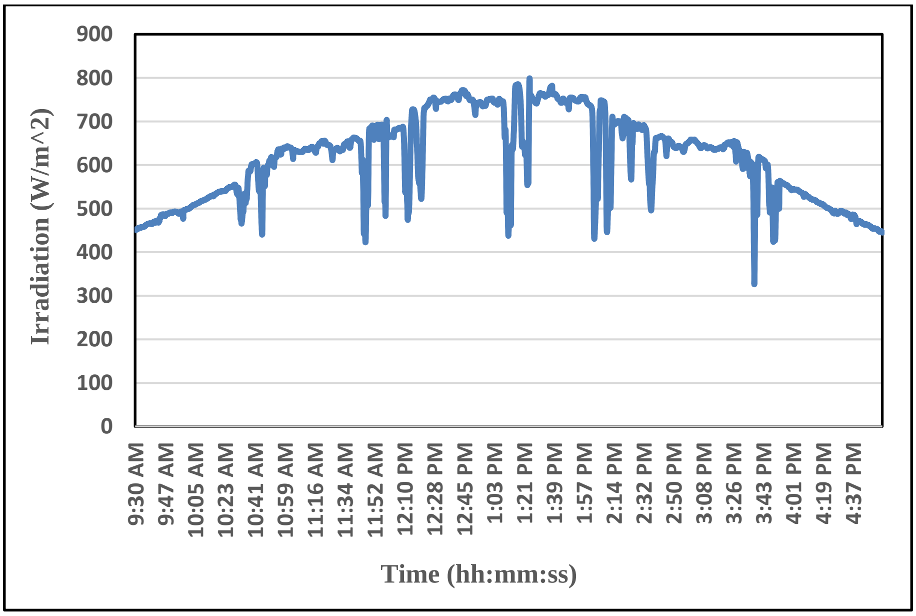


Figure 7. The average irradiance during June 2023

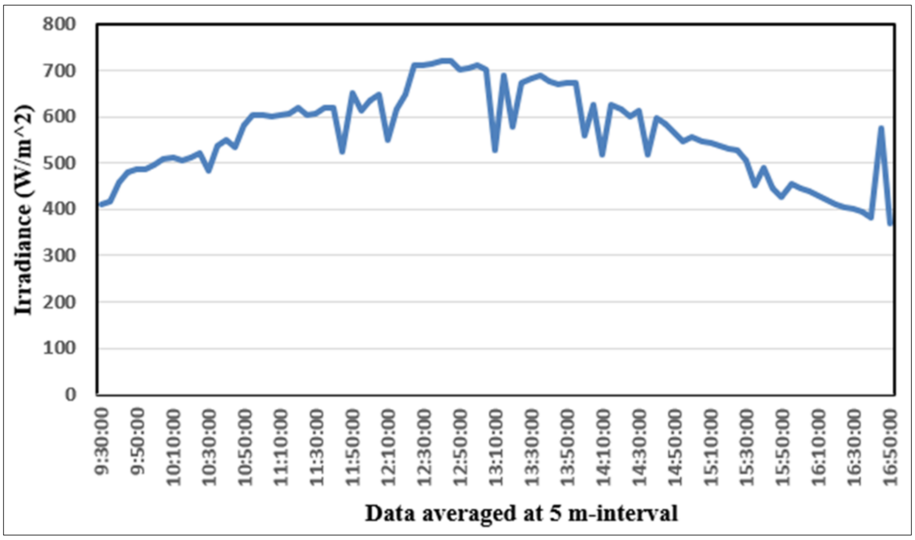


Figure 8. Smoothing of the irradiance at interval of 5 minutes

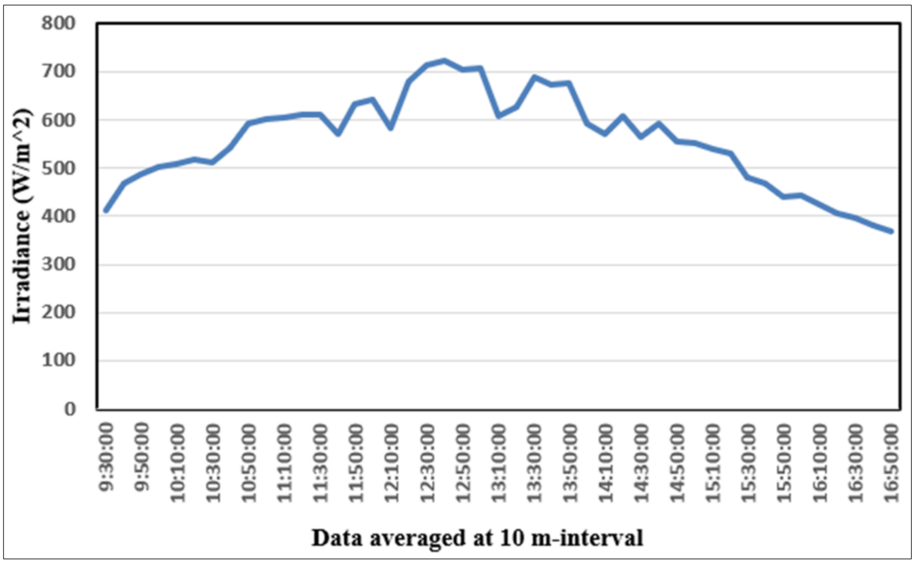


Figure 9. Smoothing of the irradiance at interval of 10 minutes

4.4. Validation of the Statistical Model

Verifying a statistical model using empirical and real-time data necessitates numerous systematic procedures that evaluate the model's correctness and dependability. The procedure usually includes comparing model predictions to actual observations, using statistical tools, and assuring data consistency. The deviations (prediction minus observation) are shown against real observations to measure model bias and accuracy. Real and statistical outputs may be compared using statistical tests such as the two-sample Student t statistic (Abdelsattar et al.)[33]. A criterion for determining how effectively a model produces data that is comparable to observed data is critical. This method applies to various data types and may improve the validation process (Lindholm et al.)[34].

Table 5 compares the relative reduction using irradiance real-time data collection in June 2023 (the highest) and November (the lowest). The relative reduction (Table 5A) for the non-smoothed data and the smoothed data at 5 and 10 minutes is 9.5%, 8.7%, and 9.2%, respectively. When the standard deviation of the data gathered is included, the average relative reduction for the non-smoothed data and the smoothed data at 5 and 10 minutes is 9.8%, 8.9%, and 9.5%, respectively. The result of the actual relative reduction in solar irradiance ranges between 8.9% and 9.85%, averaging at 9.4%. Meanwhile, Table 5B shows the relative reduction using the most important parameter (the relative sunshine hours) between June 2023 (the highest) and November (the lowest), compromising the statistical regression model. The relative reduction is 10.3%, which is within experimental error. Hence, the statistical model is reliable and can be employed in various conditions, although there are some restrictions, especially in the regional geographical location. The accuracy of the model can be calculated at 8.7%.

Table 5. Comparison between the real-time and statistical regression model

A: Irradiance Real-time Reduction (W/m2)						
Statistical Metrics	June 2023			November 2023		
	Smoothing Minutes			Smoothing Minutes		
	0	5	10	0	5	10
Average	619.419	616.451	614.594	560.303	562.558	558.201
Std	97.798	92.359	89.991	100.940	94.546	95.465
Actual Reduction				9.5%	8.7%	9.2%
Mean+Std	717.217	708.810	704.585	661.243	657.104	653.666
Mean-Std	521.621	524.092	524.603	459.363	468.012	462.736
Min reduction				7.80%	7.3%	7.2%
Max Reduction				11.9%	10.7%	11.8%
Average Relative Reduction				9.8%	8.9%	9.5%
B: Statistical Model Reduction (Relative Sunshine Duration %)						
	June 2023			November 2023		
	51.7%			46.4%		
Relative Reduction	10.3%					

5. Limitations

The research focuses on two typical months (June and November 2023) and a single metropolitan location. Broader geographical sampling (many stations) and lengthier multi-year records would enhance generality while decreasing sampling uncertainty. Including aerosol/haze data, cloud fraction/optical depth, and diffuse-direct partitioning would improve irradiance and PV performance estimations, particularly during haze episodes or partial-cloud regimes. Using regularized regression (Ridge/Lasso) or non-linear models with standard validation measures (RMSE, MAE, R^2, MAPE) on holdout data improves predictive performance evaluation for operational forecasting and system design.

While this study provides valuable insights into monthly solar irradiance variability in Kuala Lumpur, the two-month sampling period (June and November 2023) represents a temporal limitation. To account for inter-annual climate variability and improve the generalizability of our findings, we explicitly recommend:

1. **Multi-year data collection:** Future studies should collect data spanning at least 3-5 years to capture inter-annual variations associated with El Niño-Southern Oscillation (ENSO) events, which significantly influence Malaysian rainfall patterns.

2. **Expanded temporal coverage:** Monthly measurements throughout an entire year would enable more comprehensive seasonal characterization and improve model robustness.

3. **Long-term validation:** The statistical model should be validated against multi-year historical data to assess its predictive accuracy across different climatic conditions.

These extensions would enhance the model's reliability for long-term PV system planning and financial forecasting in tropical urban environments

6. Conclusion

This research used meteorological data, real-time pyranometer readings, and a multivariable regression framework to compare solar irradiance and relative sunlight duration in Kuala Lumpur. The experimental data is obtained in regional Kuala Lumpur region by developing a statistical model using real-world meteorological data, such as sunlight hours, precipitation/rainfall, and relative humidity, to identify the highest and lowest relative monthly sunshine durations. The third stage was performed by utilizing solar panel to directly measure sun irradiation in June and November 2023, from 9:30 am to 4:30 pm. Checking the statistical model's accuracy is the final step. The highest relative sunlight duration (51.7%) and the biggest mean daytime irradiance (≈619 W/m²), while November, the wettest, had the lowest relative sunshine duration (46.4%) and mean irradiance (≈560 W/m²). Rainfall/precipitation and relative humidity are the principal negative determinants of relative sunlight duration and irradiance; increasing cloudiness and precipitation during the monsoon season significantly diminish surface shortwave fluxes. Temperature differences were minimal between months and had a secondary impact compared to moisture/cloud-related factors.

The regression approach using sunshine hours, precipitation/rainfall, relative humidity, and min/max temperatures produced consistent predictions of relative sunshine duration. The model-estimated reduction from June to November (≈10.3%) is close to the observed irradiance-based reduction (mean ≈9.4%, range 8.9-9.85%), indicating good agreement within experimental uncertainty.

Smoothing the irradiance time series (5 - 10 min intervals) reduced absolute variability but did not significantly affect the relative decrease across months (≈8.7-9.8% depending on smoothing), indicating robustness of the comparison. Validation metrics and tests (KS and cross-validated regression diagnostics as described in Methods) show that the statistical model is reliable for monthly-scale assessments in this tropical urban setting. However, accuracy is limited by spatial representativeness and meteorological measurement uncertainty.

Monthly and seasonal rainfall variability in Kuala Lumpur leads to a 9-11% variation in available solar energy between the driest and wettest months analyzed. This level of variability is not insignificant in system size and financial planning, and it should be included in yield estimations, reserve capacity planning, and battery sizing for rooftop and distributed PV.

Frequent precipitation and high humidity also indicate higher soiling dynamics and possible post-rain cleaning advantages; system maintenance schedules and inverter/MPPT setups should account for such local impacts.

Because day length at Kuala Lumpur's latitude varies little, meteorological factors (cloudiness, rainfall, humidity, and urban effects) dominate seasonal PV yield differences; thus, local high-frequency irradiance measurements and humidity/precipitation records are critical inputs for accurate performance modelling.

7. Declaration of Competing Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

8. Funding Information

No funding was received from any financial organization to conduct this research

9. Author Contributions

Author proposed the research problem. In addition, collected recent articles and organized them in simple shapes, Verified the recommendation in the proposed work, Designed and proposed work, Discussed the proposed design. The author discussed the results and the final version of this paper.

10. Acknowledgments

The author expresses his gratitude to Tabriz University/ College of Engineering/ Mechanical Engineering department in Islamic Republic of Iran for supporting this study. In addition, many thanks to Dr. Razavi for their advice for the academic writing of this paper.

11. References

[1] N. S. Muhammad, J. Abdullah, and P. Y. Julien, "Characteristics of rainfall in peninsular Malaysia," in J. Physics: Conf. Series, vol. 1529, no. 5, p. 052014, May 2020.

[2] D. G. Chuah and S. L. Lee, "Solar radiation estimates in Malaysia," Solar Energy, vol. 26, no. 1, pp. 33-40, 1981.

[3] A. Rosato, A. Ciervo, F. Guarino, G. Ciampi, M. Scorpio, and S. Sibilio, "Dynamic simulation of a solar heating and cooling system including a seasonal storage serving a small Italian residential district," Thermal Science, vol. 24, no. 6 Part A, pp. 3555-3568, 2020.

[4] K. Sopian and M. Y. H. Othman, "Estimates of monthly average daily global solar radiation in Malaysia," Renewable Energy, vol. 2, no. 3, pp. 319-325, 1992.

[5] A. W. Azhari, K. Sopian, A. Zaharim, and M. Alghoul, "Solar radiation map from satellite data for a tropical environment-case study of Malaysia," in Proc. 3rd IASME/WSEAS Int. Conf. Energy & Environment, Feb. 2008, pp. 528-533.

[6] J. A. Engel-Cox, N. L. Nair, and J. L. Ford, "Evaluation of solar and meteorological data relevant to solar energy technology performance in Malaysia," J. Sustainable Energy & Environment, vol. 3, no. 3, pp. 115-124, 2012.

[7] T. Lai, "Analyses of efficiency degradation and electroluminescence imaging in crystalline-Si and CdTe thin-film PV modules via accelerated lifecycle testing," Ph.D. dissertation, The University of Arizona, 2023.

[8] K. Y. Tsung, R. Tan, and G. Y. Ii, "Estimating the global solar radiation in Putrajaya using the Angstrom-Prescott model," in IOP Conf. Series: Earth and Environmental Science, vol. 268, no. 1, p. 012056, Jun. 2019.

[9] L. K. Tan, S. M. Ong, K. Q. Lee, Y. S. Gan, W. Y. Tey, and K. N. Su, "Principal component analysis on meteorological data in UTM KL," Progress in Energy and Environment, pp. 40-46, 2017.

[10] P. Rajendran and H. Smith, "Modelling of solar irradiance and daylight duration for solar-powered UAV sizing," Energy Exploration & Exploitation, vol. 34, no. 2, pp. 235-243, 2016.

[11] D. Argüeso, A. Di Luca, and J. P. Evans, "Precipitation over urban areas in the western Maritime Continent using a convection-permitting model," Climate Dynamics, vol. 47, no. 3, pp. 1143-1159, 2016.

[12] S. F. Syed Mahbar and H. Kusaka, "Urbanisation increases cloudiness over Greater Kuala Lumpur during southwest and northeast monsoons," Weather, vol. 80, no. 5, pp. 144-151, 2025.

[13] M. T. Zateroglu, "Statistical models for sunshine duration related to precipitation and relative humidity," Avrupa Bilim ve Teknoloji Dergisi, no. 29, pp. 208-213, 2021.

[14] Mohamed Salem Nashwan, Tarmizi Ismail, and Kamal Ahmed. " Non-Stationary Analysis of Extreme Rainfall in Peninsular Malaysia" Journal of Sustainability Science and Management Volume 14 Number 3, June 2019: 17-34.

[15] F. Nucifera, W. Riasasi, and K. Ichii, "Spatio-temporal distribution of heat index and land cover change in tropical cities of Southeast Asia," in Proc. 2022 5th Int. Conf. Information and Communications Technology (ICOIACT), Aug. 2022, pp. 226-231.

[16] M. C. G. Ooi, A. Chan, M. J. Ashfold, K. I. Morris, M. Y. Oozeer, and S. A. Salleh, "Numerical study on effect of urban heating on local climate during calm inter-monsoon period in greater Kuala Lumpur, Malaysia," Urban Climate, vol. 20, pp. 228-250, 2017.

[17] Y. D. Aktas, K. Wang, Y. Zhou, M. Othman, J. Stocker, M. Jackson, and J. Hunt, "Outdoor thermal comfort and building energy use potential in different land-use areas in tropical cities: case of Kuala Lumpur," Atmosphere, vol. 11, no. 6, p. 652, 2020.

[18] C. Ghielmini, F. S. Pausata, D. Argüeso, M. Demuzere, and R. Vhuiyan, "Quantifying the role of city heterogeneity in modelling the urban precipitation effect of Kuala Lumpur," Available at SSRN 4537311.

[19] D. Mohamed, "Characterisation of urban rainfall in Kuala Lumpur, Malaysia," Ph.D. dissertation, Lund University, 1997.

[20] A. Wazwaz and M. S. Khan, "Variation in direct solar irradiation with relative humidity and atmospheric temperature," J. Ecological Engineering, vol. 20, no. 9, 2019.

[21] S. Sengupta, S. Sengupta, C. K. Chanda, and H. Saha, "Modeling the effect of relative humidity and precipitation on photovoltaic dust accumulation processes," IEEE J. Photovoltaics, vol. 11, no. 4, pp. 1069-1077, 2021.

[22] A. Takilalte, A. Dali, M. Laissaoui, and A. Bouhallassa, "Prediction of direct normal irradiation using a new empirical sunshine duration-based model," J. Renewable Energies, vol. 26, no. 1, pp. 91-103, 2023.

[23] S. G. Gouda, Z. Hussein, S. Luo, and Q. Yuan, "Review of empirical solar radiation models for estimating global solar radiation of various climate zones of China," Progress in Physical Geography: Earth and Environment, vol. 44, no. 2, pp. 168-188, 2020.

[24] S. K. Solanki, N. A. Krivova, and J. D. Haigh, "Solar irradiance variability and climate," Annual Review of Astronomy and Astrophysics, vol. 51, no. 1, pp. 311-351, 2013.

[25] O. Coddington, J. Lean, P. Pilewskie, M. Snow, E. Richard, G. Kopp, and T. Baranyi, "Solar irradiance variability: comparisons of models and measurements," Earth and Space Science, vol. 6, no. 12, pp. 2525-2555, 2019.

[26] W. Zheng, D. Lai, and K. L. Gould, "A simulation study of a class of nonparametric test statistics: a close look of empirical distribution function-based tests," Communications in Statistics-Simulation and Computation, vol. 52, no. 3, pp. 1132-1148, 2023.

[27] K. Tyagi, C. Rane, and M. Manry, "Regression analysis," in Artificial Intelligence and Machine Learning for EDGE Computing, Academic Press, 2022, pp. 53-63.

[28] A. Zain-Ahmed, K. Sopian, Z. Z. Abidin, and M. Y. H. Othman, "The availability of daylight from tropical skies—a case study of Malaysia," Renewable Energy, vol. 25, no. 1, pp. 21-30, 2002.

[29] M. Meshkinkiya, R. C. G. M. Loonen, R. Paolini, and J. L. M. Hensen, "Calibrating Perez model coefficients using subset simulation," in IOP Conf. Series: Materials Science and Engineering, vol. 556, no. 1, p. 012017, Jul. 2019.

[30] J. B. Jahangir, M. Al-Mahmud, M. S. S. Shakir, A. Haque, M. A. Alam, and M. R. Khan, "A critical analysis of bifacial solar farm configurations: theory and experiments," IEEE Access, vol. 10, pp. 47726-47740, 2022.

[31] Y. Zhang and L. Chen, "Estimation of daily average shortwave solar radiation under clear-sky conditions by the spatial downscaling and temporal extrapolation of satellite products in mountainous areas," Remote Sensing, vol. 14, no. 11, p. 2710, 2022.

[32] M. A. Khan, R. Khan, F. Algarni, I. Kumar, A. Choudhary, and A. Srivastava, "Performance evaluation of regression models for COVID-19: a statistical and predictive perspective," Ain Shams Engineering J., vol. 13, no. 2, p. 101574, 2022.

[33] M. Abdelsattar, M. A. Ismeil, M. M. A. Zayed, A. Abdelmoety, and A. Emad-Eldeen, "Assessing machine learning approaches for pH," 2024.

[34] A. Lindholm, D. Zachariah, P. Stoica, and T. B. Schön, "Data consistency approach to model validation," IEEE Access, vol. 7, pp. 59788-59796, 2019.