

This is a peer-reviewed prepublished version of the paper to be published in **Vol. 26, No. 1, March 2027**.

Predicting solar radiation in Iraq using principal component analysis and multiple regression model

Mervat A. Al-Obaidi, Jasim M. Rajab and Ali M. Al-Salihi

RECEIVED
8 May 2026

CITE THIS ARTICLE

Al-Obaidi, M.A., Rajab, J.M. and Al-Salihi, A.M., 2027. Predicting solar radiation in Iraq using principal component analysis and multiple regression model. *Nature Environment and Pollution Technology*, 26(1), D1949.

<https://doi.org/10.46488/NEPT.2027.v26i01.D1949>

EARLY-ACCESS STATEMENT

This manuscript, after peer review, is being released online in its original, unedited form to provide early access to its findings. Prior to final publication, it will undergo additional editing and formatting. Readers may note that, as an unedited version, the manuscript may have linguistic and other minor errors. The DOI will be active after the final publication of the paper.

PUBLICATION STATUS

PEER REVIEW
Completed

ONLINE VERSION
Prepublished

FINAL ISSUE
March 2027

© The Author(s) 2027. This article is Open Access under the Creative Commons Attribution 4.0 International License, allowing use, sharing, adaptation and reproduction with proper credit. Images or third-party material are covered by this licence unless otherwise noted. For uses beyond the licence, permission must be obtained from the copyright holder. Licence details: creativecommons.org/licenses/by/4.0/.

Predicting Solar Radiation in Iraq Using Principal Component Analysis And Multiple Regression Model

Mervat A. Al-Obaidi^{1,2†}, Jasim M. Rajab¹ and Ali M. Al-Salihi¹

¹Research Commission, Space Research and Technology Center, Geophysics Research Department, Baghdad, Iraq

²Mustansiriyah University, College of Science, Department of Atmospheric Sciences, Baghdad, Iraq

†Corresponding author: Mervat Ayad Abdul Khaleq. Email: mervatali2005@gmail.com

ORCID IDs of Authors <https://orcid.org/0000-0003-3458-3606> , <https://orcid.org/0000-0002-0901-472X> , <https://orcid.org/0000-0002-6099-2246>

Abstract: Daily Total Solar Radiation (TSLR) is a key input in environmental modeling. Due to the shortage of mensuration facilities, high-quality, long-term data are bounded. It is possible to get reasonably accurate radiation estimates of solar radiation from climatological parameters, which may complement or extend presenting data. This study aimed to develop a combined model (PC-MR) for predicting TSLR levels using Principal Component Analysis (PCA) and multiple regressions (MR) models utilizing fourteen meteorological parameters (RHmax, RHmin, RHavg, ET, ATmin, ATmax, ATavg, WSavg, WSmin, 2-meter dewpoint temperature, SP, CS, UV downward radiation, and CC) data during (2014-2023) from five Stations (Abu Ghraib, Al Qurnah, Balad, Tikrit, and Khanaqin) in Iraq. The models demonstrated a strong relationship with TSLR, yielding a high correlation coefficient R between (0.872 - 0.893) and an adjusted R² value ranging from (0.761 to 0.798) displaying a strong fit model. The performance metrics during the training phase showed an MAE (2.461-3.01) MJ.m⁻², MSE (4.89- 9.13) MJ.m⁻², RMSE (3.02-3.53) MJ.m⁻², and MAPE (16.46% - 21.23%). Validation conducted independent last year data (2024), yielded a strong R (0.871-0.889). In addition, the asset values of R² were (0.711-0.780), MAE ranged (1.75-2.459) MJ.m⁻², MSE (3.92-7.18) MJ.m⁻², RMSE (2.24-2.95) MJ.m⁻², and MAPE ranging from 13.55% to 18.47%. These evaluation metrics indicate moderate to strong fit predictive models, showing a closely approximate between the predicted and observed TSLR values across stations. These findings are recommended to use ground-based and observation datasets within hybrid statistical frameworks to overcome data scarcity in regional solar resource mapping. Furthermore, these models are recommended as reliable tools for operational solar energy planning and grid-integration studies in Iraq.

Keywords: Iraq, principal component analysis, Multiple linear regression, ECMWF, Solar Radiation

1. INTRODUCTION

Solar energy (SE), among renewable energies, considered of great importance worldwide due to the rising oil cost, global warming, and increasing the energy consumption (El Chaar & Lamont, 2010), its systems turn on solely on SR, without any noise, risk gas emission, and its maintenance and repair expense relatively low. Thus, SE is well recognized as new energy and a focus of recent research (Bezir et al., 2010). However, SR data remains limited in many countries due to the high expense involved observation and monitoring equipment. Additionally, operational faults in measurement devices often result from incomplete or inaccurate TSLR data, even in locations with established monitoring stations (Abdelhafidi et al., 2021; Rahimikhoob, 2010). Since

TSLR is affected by local climatic conditions, it is substantial to study it under the specific climatic situations of the location of interest (El-Sebaei et al., 2010). Thus, it is essential to develop approaches or techniques to estimate SR utilizing meteorological variables. The primary challenge in predicting daily TSLR under Iraq climatic condition is from the highly dynamic and multidimensional interactions of local variables. Iraq experiences extreme seasonal variations such high temperature, rapid fluctuation in relative humidity, shifting surface pressure, and cloud cover acts as co-predictors that govern solar radiation transmittance. In the other hand, the using traditional approaches with limited parameters introduce significant physical oversimplification and limits predictive local adaptability. Conversely, incorporating a comprehensive suite of parameter into standard regression framework causes sever multicollinearity which not stabilize models.

Multiple Regression Analysis (MRA) is one of the most widely employed in the atmospheric sciences to model the relationship between a dependence of a response variable on several independent variables by fitting a linear equation to observed data (Hwee et al., 2008; Jasim et al., 2014; Shams et al., 2021; Achite et al., 2023). Several articles have adopted and resolved this model into utilized in atmospheric science (Noi et al., 2017; Mohammadi, B., & Moazenzadeh, R. 2021; Lim et al., 2022; Abed et al., 2025; Elbeltagi et al., 2025). Although its obvious success in many applications, data collection may involve several variables from single population. The correlation of these factors may be too difficult to comprehend properly (Rajab et al., 2012; Minh et al., 2025). In these statuses, there are likely too many pairs of correlations among the factor must be considered. Additionally, graphical displays are often unsuitable or inefficient for extremely large datasets. In fitting these data with a multiple linear regression model, interpreting the full structure of inter-variable correlations can be challenging. Principal Component Analysis (PCA) has been widely applied to disentangle multicollinearity by transforming correlated variables into a set of statistically uncorrelated (orthogonal) principal components (Maenhaut et al., 1989; Murariu et al., 2025; Wang et al., 2021). Additionally, analyzing the interrelationships among the original independent variables can be used to reduce the dataset, allowing only those variables that significantly contribute to explaining the dependent variable (Marukat, 2023; Tamilselvi, 2024).

Many researchers have used Angstrom model as primary tool for solar energy patterns, the study was a conducted in developing the basic linear model for forecasting solar radiation utilize sunshine duration data (Angstrom, 1924). Several methods have been presented utilizes empirical relationships to predict TSLR, mostly employed meteorological factor such as sunshine duration (Li et al., 2011), humidity, rainfall, and temperature (Bilgili & Ozgoren, 2011).

Recently, multivariate statistical techniques have been widely applied to resolve the atmospheric problems. For instance, Rajab et al. (2013) introduced a hybrid statistical approach by combining the MLR and PCA on peninsular Malaysia, demonstrating that using PCA for dimensionality reduction significantly improve ozone forecasts precision, eliminating data redundancy and multicollinearity, proving the strong framework for regional atmospheric monitoring. In addition to the accuracy and efficiency of the (PC-MR) statistic used. Fariz and Basha, 2024 utilizes one year of TSLR data collected from the Vellore district in India by calculating the

average daily TSLR and subsequently combines neural networks, metaheuristic optimization algorithms, and PCA to identify prevailing models. The proposed technique achieved optimal error metrics and correlation coefficient values across all seasons, highlighting its consistency. For instance, a recent study Nguyen Luu Minh et al. (2025) applied MLR to predict adsorption capacity specifically for the purification of plastic waste pyrolysis oil, demonstrating the model's efficiency in optimizing purification parameters and handling multi-variable datasets. Al-Obaidi et al. (2025) was applied Stepwise Multiple Regression (SMR) to establish and evaluate models with new correlation coefficients for estimating TSLR data in several site, across Abu Ghraib station, Iraq. However, these previous studies successfully demonstrate significant of individual statistical approaches, many limitations are reliance on raw atmospheric predictors without taking into account the inherent multicollinearity among highly correlated variables. The using approaches like stepwise regression under a large number meteorological inputs can lead to unstable regression coefficients and missed physical realism. Consequently, it remains critical analyzing gap in using advanced dimensionality reduction to stabilize multi variable linear estimation frameworks under Iraq climatic conditions. Thus, our study employed PCA with MLR to estimate daily TSLR utilizing readily available meteorological data for five governorates in Iraq. SLR was considered as the dependent variable in the statistical (regressions) model to assess the relationship between TSLR and other meteorological elements. The objectives study, utilizing ten years of data, to develop and evaluate regression model with new correlation coefficients based on different meteorological elements for five stations, Iraq.

2. MATERIALS AND METHODS

2.1. Methodology

This research employed a variety of specialized software tools, including Origin (2021 and 2026), Lucidchart, Jamovi (2.7.5), SPSS, GIS ArcMap (10.4 and 10.8), and Sigma Plot, for rigorous data processing and visualization. A 10-year dataset (2014–2023) encompassing fourteen daily meteorological parameters was analyzed. PCA was integrated with MLR to optimize the predictive framework for daily TSLR, a Principal Component Regression (PCR) was developed to maintain model integrity, PCA was used as a dimensionality reduction technique focused exclusively on independent variables to bypass multicollinearity issues. The final modeling framework follows the PCR approach, where the computed PC Scores from components with eigenvalues > 1 refined by varimax rotation to analyze significant factor loadings (> 0.70) were extracted and used as the direct independent predictors in the regression analysis. Pre-processing involved Z-score normalization (Eq. 2) to standardize variables, ensuring that variations in units or magnitudes did not bias the PCA results (Jain et al., 2005; Demircioğlu, 2024).

$$Z = \frac{x - \mu}{\sigma} \quad \dots (2)$$

Where Z is the Z-score normalized one values, μ refer to mean of the variable, x original value of the variable, and σ is the standard deviation of the variable (Henderi et al., 2021; Jain et al., 2005).

Then PCA was applied to the meteorological dataset to reduce dimensionality for each station. The retained PCs, with eigenvalues greater than 1, account the maximum variance of dataset and interpreted through their correlations with the original meteorological parameters to interpret the suitability of dataset in a more meaningful form. For the Al-Qurnah , Balad, and Khanqin stations, initial diagnostic tests revealed a state of perfect multicollinearity, where the determinant of the correlation matrix was found to be zero (0.000). This mathematical singularity was attributed to the high redundancy in humidity metrics. Consequently, the Minimum Relative Humidity (RHmin) was excluded from the input vector causes very low factor loadings (< 0.40), indicating statistical insignificance, and to rectify the matrix instability. For all stations a high KMO value and a significant Bartlett's Test of Sphericity ($p < 0.001$) confirmed that the inter-variable correlations were sufficient for meaningful component extraction (Shrestha, 2021) examines the null hypothesis that the correlation matrix is an identity matrix, content variables are not significantly correlated. Combining both tests provides strong statistical evidence of meaningful inter-variable correlations (Guttman, 1954). A high KMO value together with a significant Bartlett's test result indicates the presence of correlation patterns in dataset and effectively reduced into fewer PCs without substantial loss of information. The identified PCs were incorporated into Principal Component Regression (PCR) models, developed for each of the five study sites.

MLR analysis was implemented by employing variables with high loadings on varimax-rotated PCs into regression equations, providing only those statistical significance level at the $p < 0.001$ (Abed et al., 2025). Finally, five predictive models were developed and validated by comparing the predicted TSLR values with the observed measurements. The execution of these models was evaluated using several statistical indicators, including Determination coefficient (R^2), the correlation coefficient (R), adjusted R^2 , Mean Absolute Error (MAE), Root Mean Square Error (RMSE), the F-statistic test, and Mean Absolute Percentage Error (MAPE); Fig.2 illustrates the main steps of the research methodology. The most accurate PCR models enclose the highest R , the lowest RSME, and the greater value of the F statistic test (Duzen & Aydin, 2012; Mecibah et al., 2014a). The models were validated using an independent dataset from the year 2024 to ensure predictive accuracy.

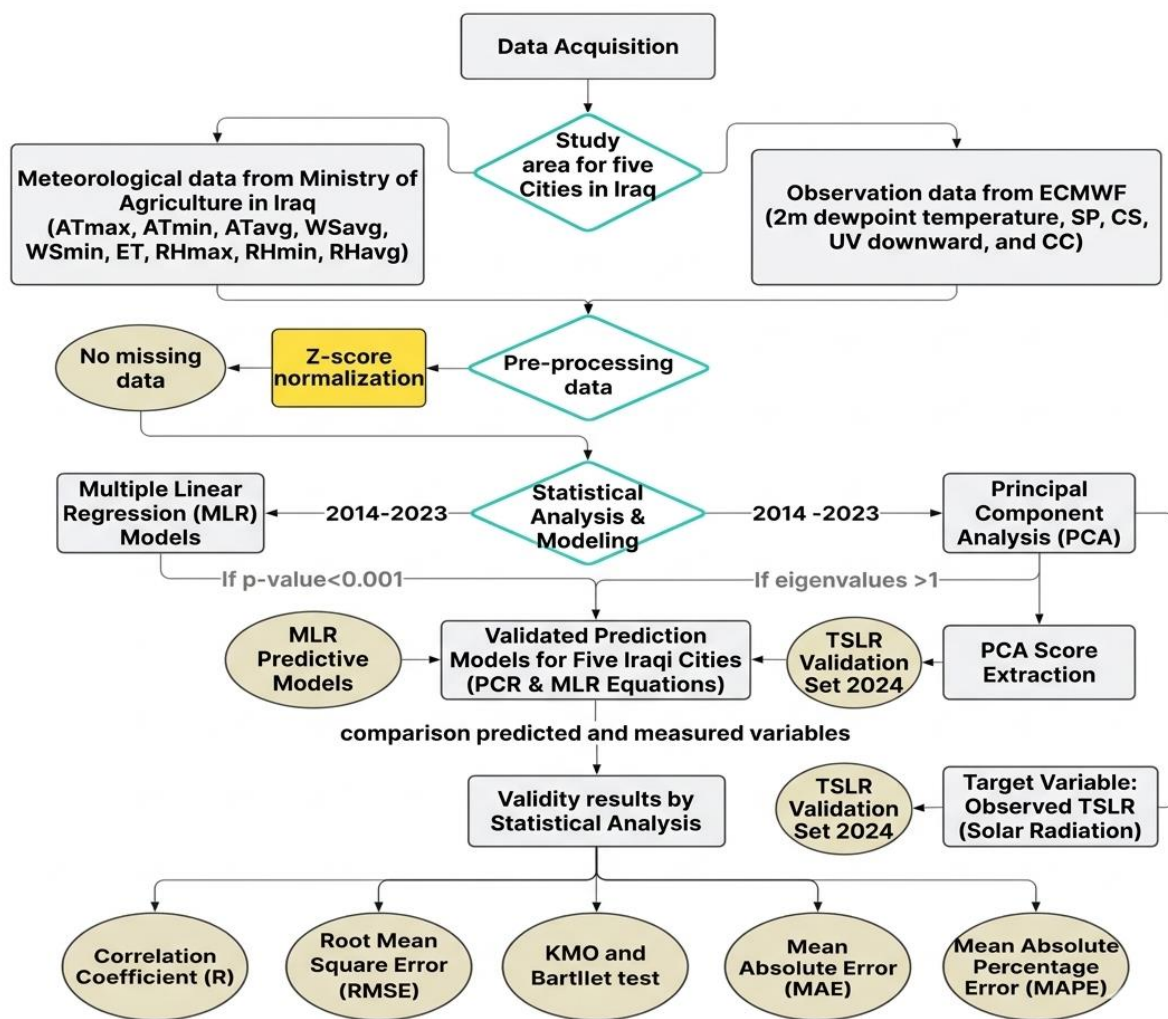


Fig. 1: Method flowchart of main steps.

2.2. Dataset Used and Study Area

The study covers an eleven-year period, from January 1, 2014, to December 31, 2023, used for model predictions (PC-MR), and the entire last year (2024) was strictly reserved as an independent blind dataset for validation. The average of nine daily (Accumulated) meteorological variables RHmax (%), RHmin (%), RHavg (%), ET (mm), ATmin (°C), ATmax (°C), ATavg (°C), WSavg (m/s), and WSmin (m/s) was obtained from the Ministry of Agriculture in Iraq's Agricultural Meteorology Center for five meteorological local Sites (Abu Ghraib, Tikrit, Balad, Al Qurnah, and Khanaqin) well distributed in four Iraqi governorates. The geographic coordinates of these stations are illustrated in Fig 1. Additionally, five other daily averaged variables—2-meter dewpoint temperature (k), SP (pa), CS (J/m^2), UV downward radiation (J/m^2), and CC—were acquired from the ERA5 Reanalysis Platform of the European Centre for Medium-Range Weather Forecasts (ECMWF). These fourteen parameters served as independent variables in a PCA model to estimate daily TSLR (J/m^2), which was measured by the Ministry of Agriculture and was kept separate as the dependent variable. The missing value

was replaced by the long-term daily mean calculated from the available historical records for that specific day. Table 1; shows the analysis involved a ten-year period from January 1, 2014, to December 31, 2023, as training for all stations.

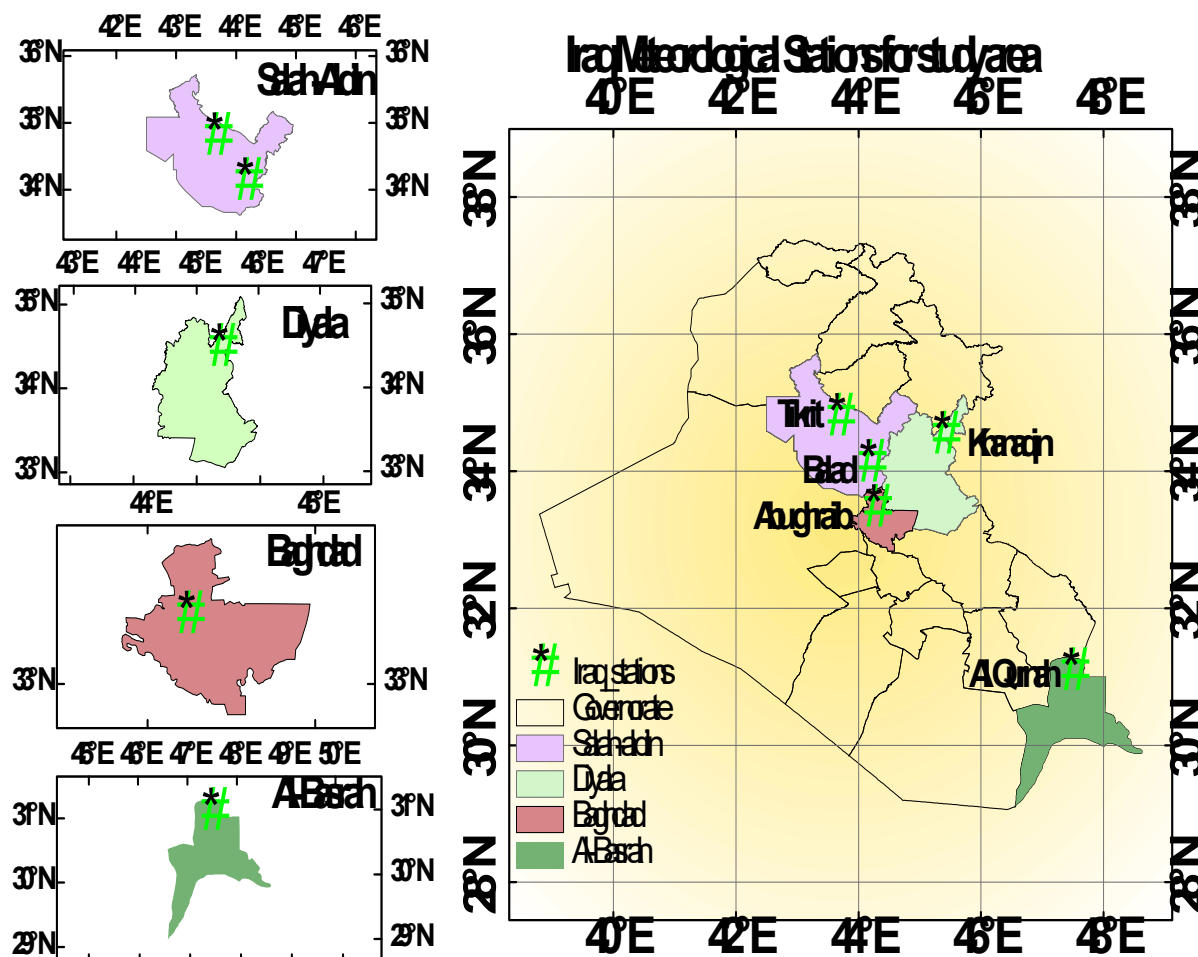


Fig. 2: Geographical study area.

Table 1: Geographic and data records period of the studied stations.

Station Name	Longitude E	Latitude N	Data series for establishing models	Data series for comparison and validation models
Abu Ghraib	44.23°	33.32°	2014-2023	2024
Al Qurnah	47.45°	30.94°	2014-2023	2024
Balad	44.14°	33.98°	2014-2023	2024
Tikrit	43.63°	34.65°	2014-2023	2024
Khanaqin	45.35°	34.39°	2014-2023	2024

Iraq is located in the southwest of Asia, situated between 29° 5' to 37° 22' N and 38° 45' to 48° 45' E, at region 35 North, above the equator (Al-Obaidi & Al-Timimi, 2022). Geographically, Iraq situated within the global solar belt and benefits from high annual solar exposure with minimal cloud coverage, making it highly

favorable for SE utilization. However, the high solar radiation in western region comparison with the other parts of Iraq (Salman et al., 2025), the average Global Horizontal Irradiance in Iraq ranges between approximately 5.0 to 5.6 kWh/m²/day, with higher values observed in southern regions compared to the north. These substantial irradiance levels position Iraq as a promising candidate for deploying SE projects in electricity generation, water desalination, and advancing sustainable development (Al-Obaidi et al., 2025).

2.3. Principal Component Analysis (PCA) and Multiple linear Regression (MLR)

PCA remains one of a most widely used methods in academic research due to its effectiveness in reducing data dimensionality while preserving most of the original information. In a simple term, PCA optimize the correlation between the variables that were originally used to create a new collection of uncorrelated variables and mutually orthogonal knowns as principal components (PCs). These PCs capture the maximum variance in the data and ensuring linear combinations of the original variables (Héberger, 1999; Lengyel et al., 2004; 46. Maćkiewicz and Ratajczak, 1993). The PCs are arranged systematically so the first PCs elucidates the majority of the variety in the data, and each successive component accounts the largest proportion of variability that not been accounted by its predecessors (Bakken et al., 1997).

The first few PCA, were used to represent the original observations typically account for a majority of variation in data set even though the number of PCAs is equal to independent original variables. For choosing others subsets of variables for regression equation the PCA also used. One common application is getting a varimax rotation of the PCs, and to keep a subset of the original variables connected with each of the initial rotated components, which are then utilized as predictors in the regression—is one typical use (Mecibah et al., 2014b; Rösler & Manzey, 1981). Varimax rotation encloses that any variable is highly linked with only one PC and a near zero associated with the other components (Abdelhafidi et al., 2021; Abed et al., 2026).

The MLR method is a commonly applied technique to improve a linear input-output model fit for a given data set. When the practical relationship is non-linear, then one or more other inputs can be modified first into a non-linear form (Abdul-Wahab et al., 1996). An Enter-type MLR model was used in this study to explain the impact of influential variables, as independent data, on the daily TSLR treated as dependent variable. To get subset of predictor variables for TSLR regression model, a variable selection technique with high loadings from varimax rotated PCs was utilized. Generally, the PCR model can be written as in Eq.:

$$\text{Pre-dependent variable} = B_0 + B_1PC_1 + B_2PC_2 + B_3PC_3 \quad \dots (1)$$

PC₁, PC₂ and PC₃ are factor scores as independent component. B₀, denote to the constant of the regression coefficients, whilst B₁, B₂ and B₃ indicate the independent coefficient component PC₁, PC₂ and PC₃ respectively (Akan & Keskin, 2019).

3. RESULTS AND DISCUSSION

As shown in Table 2, the KMO Measure of Sampling Adequacy for the five stations—Abu Ghraib, Al-Qurnah, Balad, Tikrit, and Khanaqin—was 0.824, 0.826, 0.792, 0.814, and 0.822, respectively. According to Kaiser's (1974) criteria, these values fall within the "meritorious" to "middling" range, indicating adequate sampling and strong intercorrelations among the variables (Kaiser, 1974). Also, Bartlett's Test of Sphericity produced highly significant results, with approximate chi-square values of 111726; 190964; 85802; 129112; and 86572 respectively, and degrees of freedom ranging from 78 to 92 (p-value less 0.001). These results confirm the null hypothesis that the correlation matrix is an identity matrix, sufficient inter-variable correlations to justify factor analysis, and the dataset is highly suitable for PCA.

Table 2: Kaiser-Meyer-Olkin, Bartlett's test of sphericity, and statistical metric analysis over period 2014-2024.

No. of Model	Station name	KMO Test	Bartlett's Test of Sphericity			Enter regression analysis			
			Approx. Chi-Square	df	Sig.	R	R ²	Adjusted R ²	F
1	Abu Ghraib	0.824	111726	92	0.000	0.886	0.785	0.785	4439.336
2	Al Qurnah	0.826	90964	78	0.000	0.893	0.798	0.798	4806.653
3	Balad	0.792	85802	78	0.000	0.872	0.761	0.760	3862.389
4	Tikrit	0.814	129112	91	0.000	0.880	0.775	0.774	4178.391
5	Khanaqin	0.822	86572	78	0.000	0.873	0.763	0.763	3912.308

3.1. Site-dependent models

3.1.1. Baghdad Governorate /Abu Ghraib Station

As shown in Table 3, based on the Kaiser criterion (Eigenvalue > 1), only the first three PCs, were retained for further analysis, with eigenvalues of 8.915, 1.862, and 1.146, respectively. These PCs, for 63.677%, 13.297%, and 8.312% of the total variance, resulting in a cumulative explained variance of 85.287%. This high cumulative percentage indicates that the retained components effectively capture the majority of the variability in the dataset, reflecting strong data structure and high efficiency of dimensionality reduction. Following varimax rotation (see Table 4), the retained PCs were grouped into three distinct categories depend on their variable loadings. The first component, termed the Thermal and Evapotranspiration Component, illustrates a climatic dimension characterized by high positive loadings on temperature variables, UV, CS and evapotranspiration.

Table 3. Total Variance Explained across all station over period 2014-2023.

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Eigenvalues	% of Variance	Cumulative %	Eigenvalues	% of Variance	Cumulative %
Station 1: Abu Ghraib						
PC ₁	8.915	63.677	63.677	8.639	61.704	61.704

PC ₂	1.862	13.297	76.974	2.122	15.155	76.858
PC ₃	1.164	8.312	85.287	1.180	8.428	85.287
Station 2: Al Qurnah						
PC ₁	8.066	62.046	62.046	7.327	56.365	56.365
PC ₂	1.500	11.536	73.583	2.224	17.111	73.476
PC ₃	1.280	9.847	83.429	1.294	9.953	83.429
Station 3: Balad						
PC ₁	7.320	56.305	56.305	7.204	55.413	55.413
PC ₂	1.879	14.456	70.761	1.886	14.507	69.920
PC ₃	1.301	10.009	80.770	1.411	10.851	80.770
Station 4: Tikrit						
PC ₁	9.011	64.363	64.363	8.356	59.688	59.688
PC ₂	1.507	10.763	75.126	2.134	15.244	74.932
PC ₃	1.174	8.383	83.509	1.201	8.577	83.509
Station 5: Khanaqin						
PC ₁	7.854	60.412	60.412	7.704	59.262	59.262
PC ₂	1.674	12.874	73.286	1.756	13.510	72.772
PC ₃	1.240	9.539	82.825	1.307	10.053	82.825

This component shows a great inverse relationship with relative humidity and surface pressure, showing that higher SE and thermal conditions are typically related with lower humidity and pressure levels. This reflects the strong inverse relationship between heating processes and moisture obtainable in this semi-arid environment. Second class namely, specified the wind speed element, exclusively represents the wind impact, showed by high loading both average and maximum wind speed. Third component, Atmospheric Moisture and Cloud Cover component, reflects saturated atmospheric moisture conditions and the presence of cloud cover, specificity the potential condensation and atmospheric stability changes. Overall, after rotation three PCs explained 61.704%, 15.155%, and 8.428% of the total variance, resulting in a cumulative variance of 85.287% shows a classical arid-zone pattern. Although the first component still accounts for the largest proportion of the variance, rotation has redistributed the explained variance more evenly across the PCs. This improves the interpretability of the factor solution, allowing each component to be more clearly associated with specific groups of variables. The remaining variables were represented by the rest of the PCs accounting successively.

3.1.2. Basra Governorate /Al Qurnah Station

The first three PCs were retained for further analysis, with initial eigenvalues of 8.066;1.500;1.280 which explain 62.046%;11.536%; and 9.847% respectively yielding a cumulative variance of 83.429%. This high cumulative percentage reflecting strong data structure and high efficiency of dimensionality reduction, see Table 3. Varimax rotation method was calibrated, the matrix components with high loadings classify into three classes: the first class namely; Thermal and Evapotranspiration Component with high loading controlled by air temperatures, strong UV, CS and evaporation, inversely associated with relative humidity and surface pressure at clear sky condition. Second class dominated by; Wind Speed component characterized with high loading both average and maximum wind speed, while the Third class reveals; Atmospheric Moisture and Cloud Cover component

reflects saturated atmospheric moisture conditions and the presence of cloud cover. Three components resulting in 56.365%;17.111%; and 9.953% of the total variance and explained 83.429% of a cumulative variance. Al Qurnah shows relatively stronger loadings of cloud cover, a pattern likely linked to its southern geographic location and increased exposure to humid air masses originating from the Gulf region.

3.1.3. Salah El-din Governorate /Balad and Tikrit Stations

Only three PCs in Balad and Tikrit station were retained for further analysis with initial eigenvalues of balad station 7.320;1.879; and 1.301 which explain 56.305%;14.456%; and 10.009% of initial variance respectively yielding a cumulative variance of 80.770%. while, Tikrit station captured initial eigenvalues of 9.011; 1.507; and 1.174 which explain 64.363% ;10.763%; and 8.383% of initial variance respectively yielding a cumulative variance of 83.509 %; see Table 3.

Table 4. shows Rotated Component Matrix of Balad and Tikrit highlighted high loadings classify into three classes: the first class namely; Thermal and Evapotranspiration Component with high loading temperatures, UV and evapotranspiration, inversely related with relative humidity and surface pressure. Clear sky conditions also show a significant positive loading emphasizing the strong radiative control typical of central Iraq. Second class namely; Wind Speed factor characterized by high loading both average and maximum wind speed, third class namely; Atmospheric Moisture and Cloud Cover component reflects saturated atmospheric moisture conditions and the presence of cloud cover. PCs resulting in 55.413%;14.507%; and 10.851% of the total variance and explained 80.770% of a cumulative variance. While Tikrit represented high loadings with three components resulting in 59.688 %;15.244%; and 8.577% of the total variance and explained 83.509%of a cumulative variance.

3.1.4. Diyala Governorate/Khanaqin station

Unrotated component matrix calibrated only three scores factor were retained for further analysis, with initial eigenvalues of 7.854;1.674;1.240, respectively. These factor scores explain 60.412%;12.874%;9.539% of the total variance, yielding a cumulative variance of 82.825%, see Table 3. This high cumulative percentage reflecting strong data structure and high efficiency of dimensionality reduction. While, Rotated Component Matrix with high loadings calibrated three classes: the first class namely; Thermal and Evapotranspiration Component which reflects a climatic dimension characterized by high temperatures, strong SR, and evaporation, in contrast to relative humidity and surface pressure. Second class shows; Wind Speed component exclusively represents the wind factor, characterized by both average and maximum wind speed. Third class namely; Atmospheric Moisture and Cloud Cover component reflects saturated atmospheric moisture conditions and the presence of cloud cover, Table 4, indicating potential condensation and atmospheric stability changes. Khanaqin shows the strongest positive loading of clear sky conditions, which, together with high solar input, emphasizes the dominance of radiation processes in this station. Being located in the eastern region near the Zagros foothills, this pattern reflects a mixed continental influence, where radiation dominates but cloudiness and humidity still

appear as a separate controlling mode. Three PCs resulting in 59.262%;13.510%;10.053% of the total variance and explained 82.825% of a cumulative variance.

Table 4. Rotated Component Matrix across all station over period 2014-2023.

Variables		AT _{Max}	AT _{Min}	AT _{Avg}	RH _{Max}	RH _{Min}	RH _{Avg}	WS _{Avg}	WS _{Max}	ET	2 m dewpoint	SP	CS	UV	CC
Station 1: Abu Ghraib															
Components	1	0.964	0.926	0.965	-0.931	-0.869	-0.947	0.122	0.148	0.928	0.401	-0.896	0.841	0.882	-0.236
	2	0.026	0.103	0.062	-0.043	-0.040	-0.047	0.973	0.964	0.320	-0.030	-0.134	0.244	0.208	0.027
	3	0.089	0.211	0.146	0.002	0.095	0.044	-0.007	0.016	-0.004	0.710	-0.184	-0.154	-0.214	0.698
Station 2: Al Qurnah															
Components	1	0.960	0.930	0.958	-0.814	-	-0.850	0.197	0.238	0.828	0.141	-0.899	0.862	0.799	0.497
	2	0.127	0.221	0.172	-0.108	-	-0.099	0.928	0.911	0.448	-0.057	-0.242	0.237	0.309	-0.073
	3	0.067	0.173	0.115	0.157	-	0.177	0.02	-0.041	-0.042	0.916	-0.134	-0.196	-0.147	-0.520
Station 3: Balad															
Components	1	0.958	0.943	0.969	-0.692	-	-0.687	0.034	0.116	0.947	0.413	-0.922	0.864	0.892	-0.172
	2	-0.099	0.018	-0.046	-0.213	-	-0.198	0.919	0.903	0.184	-0.167	-0.063	0.109	0.024	0.221
	3	0.030	0.168	0.091	0.475	-	0.501	-0.007	0.034	-0.139	0.690	-0.128	-0.188	-0.233	0.542
Station4: Tikrit															
Components	1	0.957	0.928	0.956	-0.935	-0.907	-0.943	-0.144	0.268	0.864	0.303	-0.861	0.794	0.849	-0.242
	2	0.151	0.224	0.185	-0.043	-0.091	-0.057	0.903	0.869	0.449	-0.014	-0.265	0.308	0.256	0.102
	3	0.025	0.154	0.083	-0.042	0.019	-0.025	0.079	0.026	-0.013	0.725	-0.144	-0.178	-0.244	0.723
Station 5: Khanaqin															
Components	1	0.975	0.944	0.978	-0.907	-	-0.897	0.024	0.246	0.928	0.128	-0.900	0.836	0.887	-0.209
	2	0.011	0.059	0.031	-0.061	-	-0.079	0.917	0.855	0.280	-0.093	-0.158	0.115	0.028	0.210
	3	-0.013	0.17	0.066	0.199	-	0.222	0.018	0.087	-0.087	0.876	-0.145	-0.040	-0.150	0.597

3.2. Multi-linear Regression analysis

An enter regression analysis was performed using PCs scores as independent factors. Table 5; PCR models presents highly significant across all stations ,with PC1 exerting the dominate positive influence on daily TSLR specifically, VIF is exactly 1.000 for all predictors, showing that orthogonal PC scores completely eliminated multicollinearity in regression framework. According to equation 1 :Five regression models for (Abu Ghraib, Al Qurnah, Balad, Tikrit, and khanqah) have been produced algorithms for evaluate TSLR are provided in Equation (3) respectively:

$$\text{Pre_TSLR_mod1} = 18.082 + 5.803\text{PC1} + 0.984\text{PC2} - 1.155\text{PC3}$$

$$\text{Pre_TSLR_mod2} = 16.348 + 4.554 \text{PC1} + 1.32 \text{PC2} - 1.093\text{PC3}$$

$$\text{Pre_TSLR_mod3} = 16.266 + 5.201\text{PC1} + 0.132\text{PC2} - 1.379\text{PC3}$$

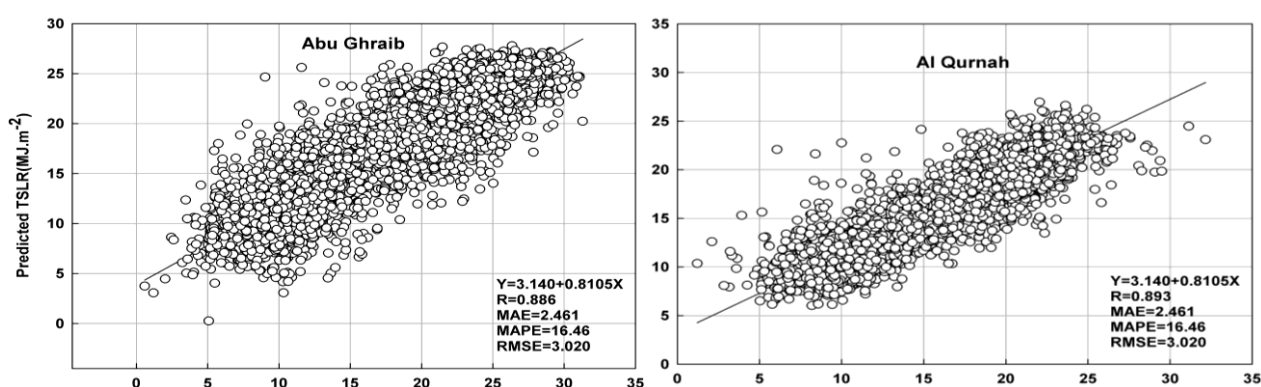
$$\text{Pre_TSLR_mod4} = 17.802 + 6.087\text{PC1} + 1.481\text{PC2} - 1.784\text{PC3}$$

$$\text{Pre_TSLR_mod5} = 16.727 + 5.885\text{PC1} + 0.133 \text{PC2} - 1.203\text{PC3} \quad \dots (3)$$

Table 5: Standardized regression coefficients(β), standard errors, VIF, and p-values for the PCR models across all study stations 2014–2023.

Station Name	β coefficients				Standard error			VIF	p- value	Durbin- Watson
	PC1	PC2	PC3	Constant	PC1	PC2	PC3			
Abu Ghraib	0.857	0.145	-0.171	0.052	0.052	0.052	0.052	1.000	0.000	0.918
Al Qurnah	0.864	0.234	-0.176	0.037	0.037	0.037	0.037	1.000	0.000	1.121
Balad	0.843	0.021	-0.223	0.050	0.050	0.050	0.050	1.000	0.000	0.884
Tikrit	0.854	0.208	-0.250	0.048	0.048	0.048	0.048	1.000	0.000	0.995
Khanaqin	0.856	0.019	-0.175	0.055	0.055	0.055	0.055	1.000	0.000	0.779

In this study, the performance of TSLR model for five stations was assessed using statistical metrics as shown in Table 2. The models demonstrate a strong relationship with TSLR, as indicated by correlation coefficient R ranging from 0.872 to 0.893 and an adjusted R^2 value between 0.761 and 0.798 indicating a fit model framework. The model's F-statistic values ranging from 3862.389 to 4806.653 associated with a p-value ≤ 0.001 , confirm that the variance of independent variable TSLR over the observed period 2014–2023 is substantially explained by the predictor variables. As shown in Fig 3, the scatter plots of predicted and observed TSLR in 2024, the models yielding moderate to good predictive performance. The error metrics vary across station with MAE ranging 2.461 to 3.01 MJ.m^{-2} , RMSE 3.020 to 3.53 MJ.m^{-2} , and MAPE 16.46% to 21.23% with Residual Stander Deviation ranging from 2.2 to 3.02 reflecting its level of precision. Notably, the highest prediction error were found at Khanaqin MAPE=21.23% and Balad MAPE =17.00% are attributed to their complex micro-climates, terrain-induced atmospheric changes, and transitional dust or cloud cover variations, which are challenging for regional PCR models to capture perfectly.



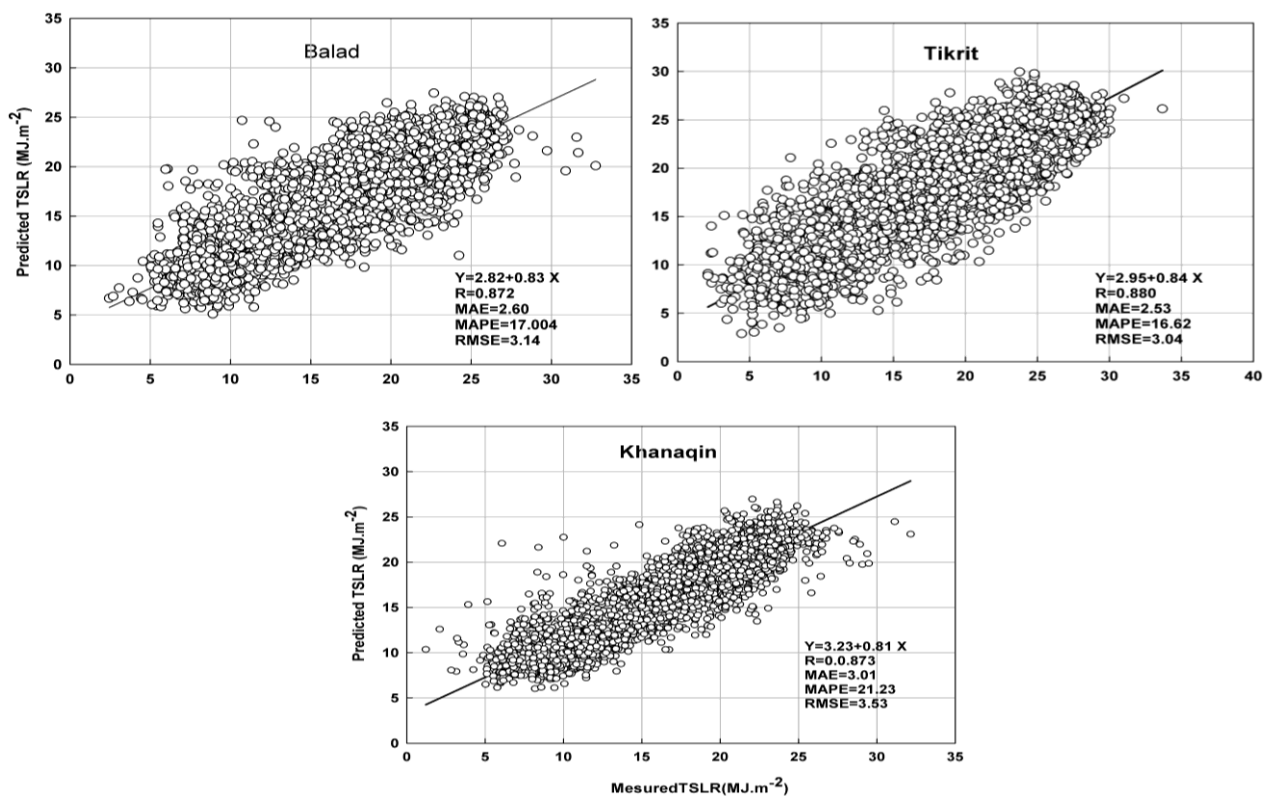


Fig. 3: Scatter plot between the predicted and measured daily TSLR in 2024.

Figure 4 compares the observed TSLR values with those predicted from January 2014 to January 2023. At the five stations, which include Abu Ghraib, Qurnah, Balad, Tikrit, and Khanaqin, both the predicted (red line) and observed (black line) TSLRs show comparable and slick seasonal behaviors, tapering in winter (December–February) and elevating during the summer (June–August). This suggests that the model adequately represents the general seasonal fluctuations in TSLR. There are small discrepancies between the observational and model-derived TSLR, especially at specific times. Although small differences remain, in general agreement is strong, as shown by the close co-variation of the two lines. Observed and modeled TSLR remain closely aligned entire study period 2014–2023, implies that the model reliably represents the temporal evolution of TSLR. The diagnostic analysis of PCR model residual revealed highly acceptable behavior, validating the model's assumption. Although the Anderson–Darling normality test indicated statistically significant departure from the normal distribution ($p < 0.001$), this is expected due to the large sample size where formal tests are overly sensitive to minor deviations. Fig 5. Confirm that the residuals are approximately normally distributed. Additionally, Durbin–Watson values ranging from 0.779 to 1.121 as shown in Table 5, show some positive serial correlation, this is normal for daily weather data which naturally change slowly from day to day. However, this does not affect the model's ability to predict daily TSLR, as proven by the good error results.

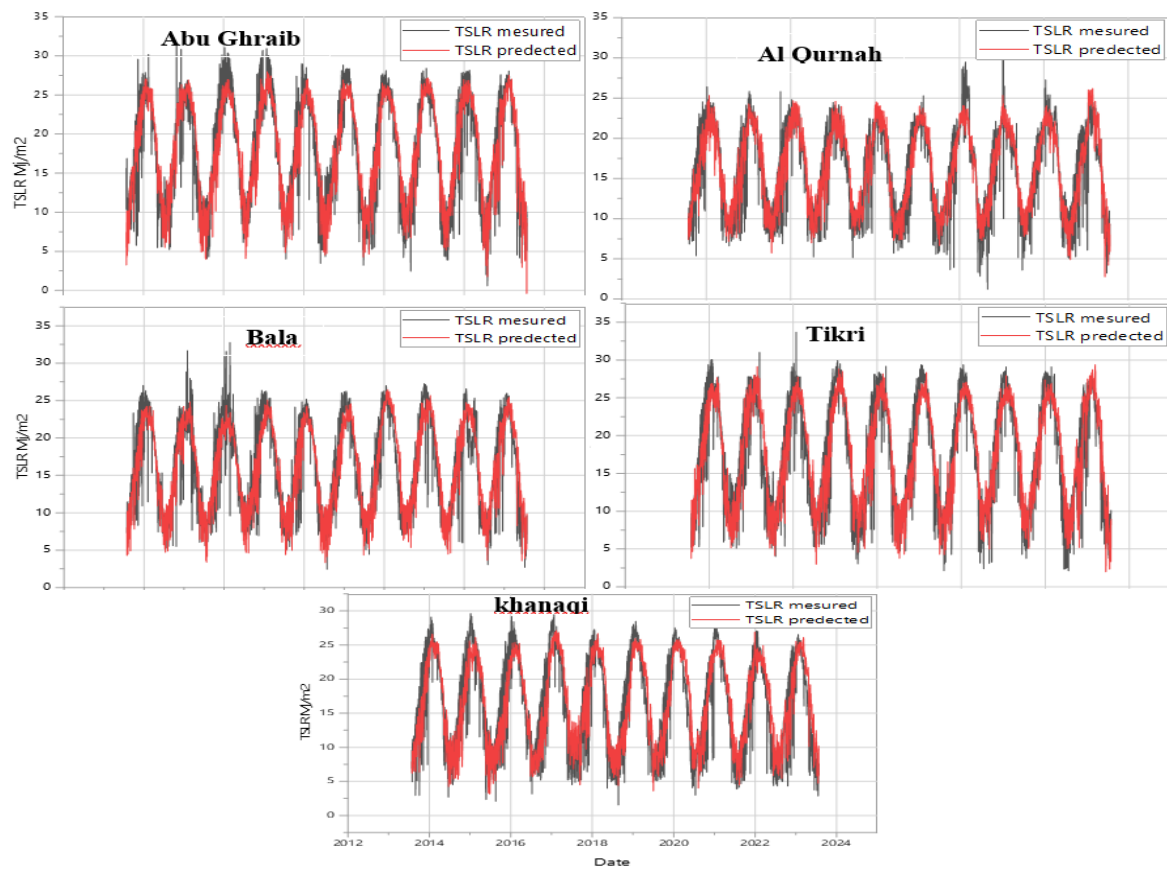


Fig 4. Comparison estimated and measured daily TSLR for five stations over period 2014-2024.

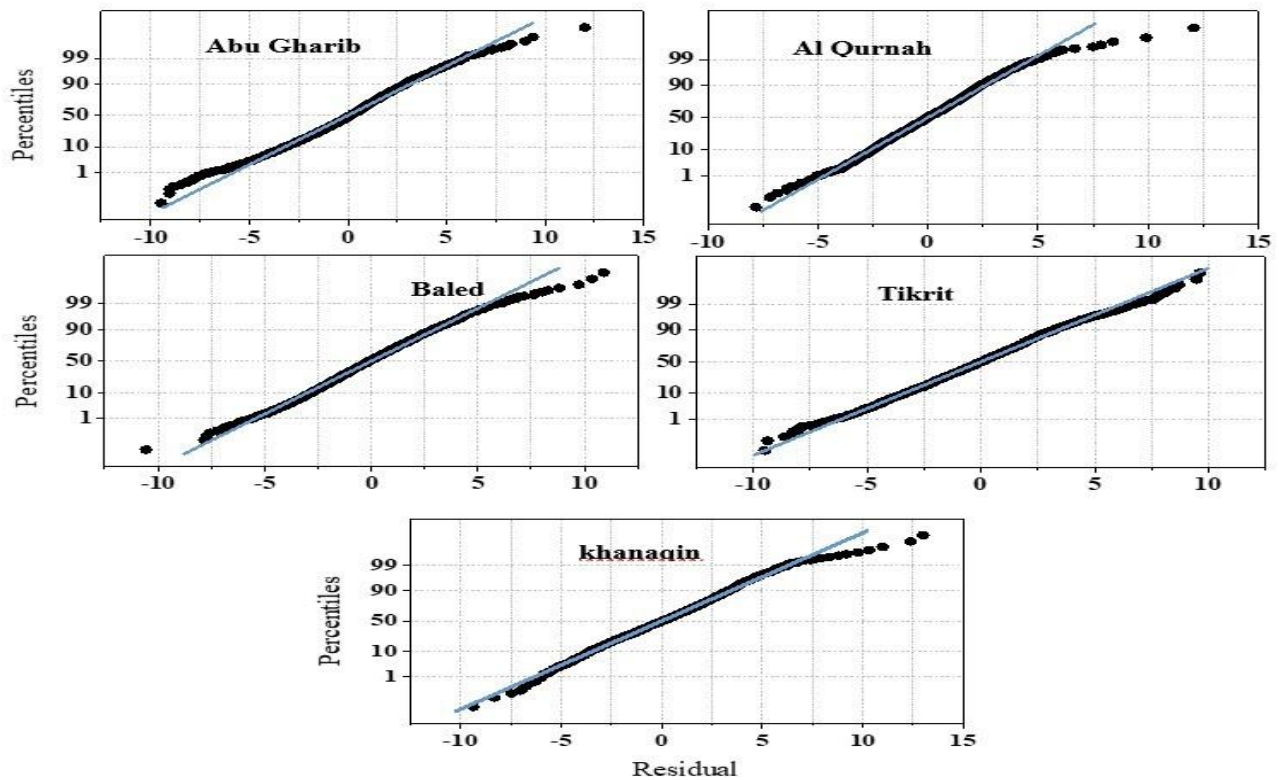


Fig 5. Residual normality test for five stations over period 2014-2024.

4. DISCUSSION

The results were validated and compared using data of 2024 have been shown in Table 6; for five stations: Abu Ghraib, Al Qurnah, Balad, Tikrit, and Khanaqin. The models achieve a relatively Residual Mean Absolute MAE is (2.459, 1.75, 2.18, 2.41, and 1.83) MJ.m⁻², MSE (6.407, 4.60, 5.69, 7.186, and 3.92) MJ.m⁻², RMSE (3.144, 2.24, 2.77, 2.95, and 2.24) MJ.m⁻², MAPE (16.23%, 13.55%, 18.47%, 18.06%, and 15.08%) with Residual Stander Deviation (2.54, 2.14, 2.39, 2.68, and 1.98) reflecting its precision.

Table 6. The validations and comparisons for five stations data.

Station Name	Comparison	R	R ²	MAE	MSE	RMSE	MAPE
Abu Ghraib	Prediction	0.886	0.785	2.63	7.38	3.42	19%
	validation	0.884	0.780	2.46	6.41	3.14	16.23%
Al Qurnah	Prediction	0.893	0.798	2.46	4.89	3.02	16.46%
	validation	0.889	0.711	1.75	4.60	2.24	13.55%
Balad	Prediction	0.872	0.761	2.60	6.61	3.14	17%
	validation	0.871	0.761	2.18	5.69	2.77	18.47%
Tikrit	Prediction	0.880	0.775	2.53	8.41	3.04	16.62%
	validation	0.879	0.741	2.41	7.19	2.95	18.06%
Khanaqin	Prediction	0.873	0.763	3.01	9.13	3.53	21.23%
	validation	0.871	0.712	1.83	3.92	2.24	15.08%

These results imply that model's predictions closely match the observed values, yielding a more accurate regression equation. The algorithm model with observed TSLR need to verified to test the TSLR. The validations were conducted with a linear regression correlation for one year using the data of 2024. The TSLR algorithm that used fourteen variables, (ATmax, ATmin, ATavg, WSavg, WSmin, ET, RHmax, RHmin, RHavg, 2m dewpoint temperature, SP, CS, UV downward, and CC) were yielded a strong correlation coefficient (R) with observed TSLR from for 2024 represented by values (0.884, 0.889, 0.871, 0.879, and 0.871). In addition, the asset values of R² were (0.780, 0.711, 0.761, 0.741, and 0.712) for considered station respectively.

5. CONCLUSIONS

Across five Iraq stations, the first component consistently represent temperature -Evapotranspiration factor confirming solar heating as the primary driver across the country. The second component reflect winds dynamics. while, the third highlights a moisture–cloud factor associated with dewpoint temperature and cloud cover. Al Qurnah exhibits the strong contrast cloud-humidity effects due to Gulf air mass influences, whereas Balad separates humidity as an independent factor, indicating transitional climatic control. In contrast, central stations (Abu Ghraib, Tikrit, and Khanaqin) remain dominated by solar–temperature forcing, consistent with their arid to semi-arid climatic regime. Five models have been developed, have a good relationship with TSLR as reflecting a high correlation coefficient for Abu Ghraib, Al Qurnah, Balad, Tikrit, and Khanqah, were ranged

between (0.872-0.893), and an adjusted R^2 value of (0.761-0.798), demonstrate a good fit model. The model's F-statistic were high associated with a p-value ≤ 0.001 , shows that the variance in the dependent variable TSLR over the observed time period can be substantially explained by the predictor variables.

The models achieve a relatively MAE for Abu Ghraib, Al Qurnah, Balad, Tikrit, and khanqah ranged between (2.461-3.01) MJ.m⁻², MSE (4.89- 9.13) MJ.m⁻², RMSE (3.02-3.53) MJ.m⁻², and MAPE ranged between (16.46% - 21.23%) the validations and comparisons of the data were conducted by using the last year data for five stations exhibit a good relationship with TSLR as evidenced by a high correlation coefficient for Abu Ghraib, Al Qurnah, Balad, Tikrit, and khanqah, were ranged between (0.871-0.889), and an adjusted R^2 value of (0.711-0.780), indicating a good fit model .The models achieve a relatively MAE ranged between (1.75-2.459) MJ.m⁻², MSE (3.92-7.18) MJ.m⁻², and RMSE (2.24-2.95) MJ.m⁻², and MAPE (13.55% - 18.47%). The all result implies that the models' predictions closely match observed values create a regression equation fairly reliable. The investigation was supported by validation and comparison carried out on this study which effectively illustrate high accuracy of regression algorithms.

Author Contributions: Conceptualization, Mervat A. Al-Obaidi, and Jasim M. Rajab; methodology, Mervat A. Al-Obaidi, and Jasim M. Rajab; validation, Mervat A. Al-Obaidi, Jasim M. Rajab, and Ali M. Al-Salihi; formal analysis, Mervat A. Al-Obaidi, and Jasim M. Rajab; writing—original draft preparation, Mervat A. Al-Obaidi, Jasim M. Rajab, and Ali M. Al-Salihi; writing—review and editing, Mervat A. Al-Obaidi, Jasim M. Rajab, and Ali M. Al-Salihi; supervision, Jasim M. Rajab and Ali M. Al-Salihi; All authors have read and agreed to the published version of the manuscript.

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflicts of Interest: All authors have read, understood, and have complied, as applicable, with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

REFERENCES

1. Abdelhafidi, N., Bachari, N. E. I., & Abdelhafidi, Z. 2021. "Estimation of solar radiation using stepwise multiple linear regression with principal component analysis in Algeria". *Meteorology and Atmospheric Physics*, 133(2), 205–216. <https://doi.org/10.1007/s00703-020-00739-0>
2. Abdul-Wahab, S., Bouhamra, W., Ettouney, H., Sowerby, B., & Crittenden, B. D. 1996. "Predicting ozone levels: A statistical model for predicting ozone levels in the Shuaiba Industrial Area, Kuwait". *Environmental Science and Pollution Research*, 3(4), 195–204.
3. Abed, F.G., Rajab, J.M., and Daway, H.G. 2026. "Modeling Surface Ozone Concentrations in Iraq Using Principal Component Analysis and Stepwise Multiple Linear Regression Method". *Ozone Science and Engineering Open-source preview*, 48(1), pp. 66–82.
4. Abed, F. G., Rajab, J. M., & Daway, H. G. 2025. "Integration of Stepwise Multiple Linear Regression and Artificial Neural Network Models for Predicting Surface Ozone Concentration in Iraq". *Papers in Applied Geography*, 1–19.

5. Achite, M., Elshaboury, N., Jehanzaib, M., Vishwakarma, D. K., Pham, Q. B., Anh, D. T., Abdelkader, E. M., and Elbeltagi, A. 2023. "Performance of Machine Learning Techniques for Meteorological Drought Forecasting in the Wadi Mina Basin, Algeria". *Water*, 15(4), 765.
6. Akan, R., & Keskin, S. N. 2019. "The effect of data size of ANFIS and MLR models on prediction of unconfined compression strength of clayey soils". *SN Applied Sciences*, 1(8), 843.
7. Al-Obaidi, M. A., & Al-Timimi, Y. K. 2022. "Change detection in Mosul dam lake, north of Iraq using remote sensing and GIS techniques". *Iraqi Journal of Agricultural Sciences*, 53(1), 38–47.
8. Al-Obaidi, M. A., Rajab, J. M., & Al-Salihi, A. M. 2025. "Estimation Global Solar Radiation from Common Meteorological Data Using a Stepwise Multiple Regression Model in Abu Ghraib, Iraq". *IOP Conference Series: Earth and Environmental Science*, 1531(1), 012006.
9. Angstrom, A. 1924. "Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation". *Quarterly Journal of the Royal Meteorological Society*, 50(210), 121–126.
10. Bakken, G. A., Long, D. R., & Kalivas, J. H. 1997. "Examination of criteria for local model principal component regression". *Applied Spectroscopy*, 51(12), 1814–1822.
11. Bezir, N. C., Akkurt, I., & Ozek, N. 2010. "Estimation of horizontal solar radiation in Isparta (Turkey)". *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 32(6), 512–517.
12. Bilgili, M., & Ozgoren, M. 2011. "Daily total global solar radiation modeling from several meteorological data". *Meteorology and Atmospheric Physics*, 112(3–4), 125–138. <https://doi.org/10.1007/s00703-011-0137-9>
13. Demircioğlu, A. 2024. "The effect of feature normalization methods in radiomics". *Insights into Imaging*, 15(1), 2. <https://doi.org/10.1186/s13244-023-01575-7>
14. Duzen, H., & Aydin, H. 2012. "Sunshine-based estimation of global solar radiation on horizontal surface at Lake Van region (Turkey)". *Energy Conversion and Management*, 58, 35–46.
15. Elbeltagi, A., Vishwakarma, D.K., Katipoğlu, O.M. et al. 2025. "Air temperature estimation and modeling using data driven techniques based on best subset regression model in Egypt". *Sci Rep* 15, 20200.
16. El Chaar, L., & Lamont, L. A. 2010. "Global solar radiation: Multiple on-site assessments in Abu Dhabi, UAE". *Renewable Energy*, 35(7), 1596–1601.
17. El-Sebaei, A. A., Al-Hazmi, F. S., Al-Ghamdi, A. A., & Yaghmour, S. J. 2010. "Global, direct and diffuse solar radiation on horizontal and tilted surfaces in Jeddah, Saudi Arabia". *Applied Energy*, 87(2), 568–576.
18. Fariz, T. N., & Basha, S. S. 2024. "Enhancing solar radiation predictions through COA optimized neural networks and PCA dimensionality reduction". *Energy Reports*, 12, 341–359.
19. Guttman, L. 1954. "Some necessary conditions for common-factor analysis". *Psychometrika*, 19(2), 149–161.
20. Héberger, K. 1999. "Evaluation of polarity indicators and stationary phases by principal component analysis in gas–liquid chromatography". *Chemometrics and Intelligent Laboratory Systems*, 47(1), 41–49.
21. Mohammadi, B., & Moazenzadeh, R. 2021. "Performance analysis of daily global solar radiation models in Peru by regression analysis". *Atmosphere*, 12(3), 389.
22. Henderi, H., Wahyuningsih, T., & Rahwanto, E. 2021. "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer". *International Journal of Informatics and Information Systems*, 4(1), 13–20.

23. Hwee S. L., M. Z. MatJafri, K. Abdullah, A. N. Alias, J. M. Rajab and N. M. Saleh 2008. "Algorithm for TSS Mapping Using Satellite Data for Penang Island, Malaysia". Fifth International Conference on Computer Graphics, Imaging and Visualisation, Penang, Malaysia, 2008, pp. 376-379.
24. Jain, A., Nandakumar, K., & Ross, A. 2005. "Score normalization in multimodal biometric systems". Pattern Recognition, 38(12), 2270–2285.
25. Jasim, M. R., MatJafri, M. Z., & Lim, H. S. 2014. "Air surface temperature correlation with greenhouse gases by using airs data over peninsular malaysia". Pure and Applied Geophysics, 171(8), 1993–2011.
26. Kaiser, H. F. 1974. "An Index of Factorial Simplicity". Psychometrika, 39(1), 31–36. <https://doi.org/DOI:10.1007/BF02291575>.
27. Lengyel, A., Héberger, K., Paksy, L., Bánhidi, O., & Rajkó, R. 2004. "Prediction of ozone concentration in ambient air using multivariate methods". Chemosphere, 57(8), 889–896. <https://doi.org/10.1016/j.chemosphere.2004.07.043>
28. Li, H., Ma, W., Lian, Y., Wang, X., & Zhao, L. 2011. "Global solar radiation estimation with sunshine duration in Tibet, China". Renewable Energy, 36(11), 3141–3145. <https://doi.org/10.1016/j.renene.2011.03.019>.
29. Lim HS, Jasim M. Rajab, Al-Salihi A, Zainab S, and MatJafri M 2022. "A statistical model to predict and analyze air surface temperature based on remotely sensed observations". Environ Sci Pollut Res 29, 9755–9765.
30. Maenhaut, W., Cornille, P., Pacyna, J. M., & Vitols, V. 1989. "Trace element composition and origin of the atmospheric aerosol in the Norwegian Arctic". Atmospheric Environment, 23(11), 2551–2569.
31. Marukatat, S. 2023. "Tutorial on PCA and approximate PCA and approximate kernel PCA". Artificial Intelligence Review, 56(6), 5445–5477. <https://doi.org/10.1007/s10462-022-10297-z>
32. Mecibah, M. S., Boukelia, T. E., Tahtah, R., & Gairaa, K. 2014a. "Introducing the best model for estimation the monthly mean daily global solar radiation on a horizontal surface (Case study: Algeria)". Renewable and Sustainable Energy Reviews, 36, 194–202.
33. Mecibah, M. S., Boukelia, T. E., Tahtah, R., & Gairaa, K. 2014b. "Introducing the best model for estimation the monthly mean daily global solar radiation on a horizontal surface (Case study: Algeria)". Renewable and Sustainable Energy Reviews, 36, 194–202.
34. Minh, T. N. L., Van Melkebeke, M., Carrieri, E., Hogue, J., Poelman, H., Michiels, Y., Van Geem, K. M., & De Meester, S. 2025. "Multiple linear regression in adsorption capacity prediction: Application in plastic waste pyrolysis oil purification". Separation and Purification Technology, 134651.
35. Murariu, G., Dinca, L., & Munteanu, D. 2025. "Trends and Applications of Principal Component Analysis in Forestry Research: A Literature and Bibliometric Review". Forests, 16(7), 1155.
36. Nguyen Luu Minh, T., Van Melkebeke, M., Carrieri, E., Hogue, J., Poelman, H., Michiels, Y., Van Geem, K. M., & De Meester, S. 2025. "Multiple linear regression in adsorption capacity prediction: Application in plastic waste pyrolysis oil purification". Separation and Purification Technology, 378, 134651. <https://doi.org/https://doi.org/10.1016/j.seppur.2025.134651>
37. Noi, P. T., J. Degener and M. Kappas 2017. "Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data". Remote Sensing 9(5): 398.
38. Rahimikhoob, A. 2010. "Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment". Renewable Energy, 35(9), 2131–2135.

39. Rajab, J. M., Jafri, M. Z. M., Lim, H. S., & Abdullah, K. 2012. "Regression analysis in modeling of air surface temperature and factors affecting its value in Peninsular Malaysia". *Optical Engineering*, 51(10), 101702.
 40. Rajab, J. M., MatJafri, M. Z., & Lim, H. S. 2013. "Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia ". *Atmospheric Environment*, 71, 36–43. <https://doi.org/10.1016/j.atmosenv.2013.01.019>
 41. Rösler, F., & Manzey, D. 1981. "Principal components and varimax-rotated components in event-related potential research: some remarks on their interpretation". *Biological Psychology*, 13, 3–26.
 42. Shams, S. R., Jahani, A., Kalantary, S., Moeinaddini, M., & Khorasani, N. 2021. "The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO₂ concentration". *Urban Climate*, 37. <https://doi.org/10.1016/j.uclim.2021.100837>
 43. Shrestha, N. 2021. "Factor analysis as a tool for survey analysis". *American Journal of Applied Mathematics and Statistics*, 9(1), 4–11.
 44. Salman, R. K., Farhan, S. S., Al-Falahi, M. N., & Mohammed, T. E. 2025. "Spatial Analyses of Reliability of Solar Power in the Western Part of Iraq". *Nature Environment and Pollution Technology*, 24(2), 1-6.
 45. Tamilselvi, M. 2024. "Significance and Comparison of PCA in Removing Multicollinearity of Variables in Face Detection and Recognition". 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), 1–6.
 46. Maćkiewicz, A., & Ratajczak, W. 1993. "Principal components analysis (PCA)". *Computers & Geosciences*, 19(3), 303-342.
- Wang, M., Gao, G., Huang, H., Heidari, A. A., Zhang, Q., Chen, H., & Tang, W. 2021. "A principal component analysis-boosted dynamic Gaussian mixture clustering model for ignition factors of Brazil's rainforests". *IEEE Access*, 9, 145748–14576