Solar Irradiance Forecasting using Dynamic Mode Decomposition

Olusegun Abel Odejobi[1, 2, *], Kehinde Olukunmi Alawode[1] and Muyideen Olalekan Lawal[1]
[1] Department of Electrical Electronics Engineering, Faculty of Engineering,
Osun State University, Osogbo, Osun, 210001, Nigeria
[2] Prototype Engineering Development Institute, Ilesa, Osun, 233252, Nigeria
***Corresponding Author E-mail:** segodej@gmail.com
Received: Oct 25, 2024; **Revised:** Dec 23, 2024; **Accepted:** Jan 03, 2025

Abstract

Reliable solar irradiance prediction is necessary for an easier transition from dependence on fossil fuels to renewable energy sources. The features of solar irradiance, such as its non-linearity and high variability, make predicting it a challenging task. This challenge is traditionally addressed by using regression and other ensemble models that require significantly large historical data to adequately train and rely on domain-specific knowledge. In this study, a data-driven framework that employed dynamic mode decomposition for solar irradiance forecasting was proposed. The efficiency of the dynamic mode decomposition-based framework was verified by employing it for short-term forecasting using two distinct datasets from geographically diverse locations. The comparative advantage over traditional regression was confirmed using performance assessment measures, including mean absolute error, mean bias error, and root mean square error. The resulting forecasts significantly outperformed the benchmark models, demonstrating that the proposed model could effectively forecast short-term solar irradiance with improved accuracy.

Keywords: Data-driven approaches, Dynamic mode decomposition, Forecasting, Solar Irradiance

1. Introduction

The use of solar energy as a substitute to hydrocarbon fuel has increased significantly in recent years due to the environmental benefits it presents [1]. This increase in adoption has also been accompanied by improvements in solar photovoltaic technology due to a reduction in material costs and governmental support for sustainable development [2]. While research is still ongoing to enhance the efficiency of solar cells, accurate forecasts of solar irradiance variation is critical to the integration of existing solar energy infrastructure into the electric grid [3].

Solar energy is not a resource that is easily predictable due to its dependence on a myriad of factors including the location, orientation of the sun, aerosol, time of day, and weather patterns, among many other variables. Various methods for solar irradiance prediction have been proposed in recent studies. Numerical weather prediction models simulate the dynamic states of the atmosphere to generate predictions [4]. Image-based models such as sky camera image-based models and satellite image-based models utilize predictions of cloud movements and their effects on solar irradiance [5]. Machine learning models and statistical models employ meteorological parameters such as pressure, humidity, and temperature, together with historical observations of solar irradiance as inputs.

Ensemble models such as the Weather Research and Forecasting Solar Ensemble Prediction System [6] and the Conditional Entropy Embedded Wasserstein Generative Adversarial Network with Long Short-Term Memory model [7] utilized a combination of different models to predict solar irradiance. Other hybrid models also combined the advantages of various methods to enhance forecast accuracy [8],[9]. Furthermore, among models that utilize artificial intelligence, specifically artificial neural networks [10–12] were the machine learning methods that are most often utilized for solar irradiance forecasting. Several studies showed that Long Short-Term Memory (LSTM) is one of the best forecasting techniques [13–16] but to achieve this optimal performance, the availability of extensive training datasets is crucial [17]. While sufficient data might be available for certain regions, many other areas lacked sufficient data, particularly for longer forecasting horizons. Furthermore, the execution of these models requires substantial computational power, which may not be readily accessible in all locations. Therefore, the need for forecasting methods that are not only highly effective but also computationally efficient is evident.

Dynamic mode decomposition (DMD) was considered for its ability to extract system dynamics directly from limited data without the need for extensive training [18]. Unlike LSTM and Convolutional Neural Network, which require large datasets and significant computational resources [19], DMD efficiently identifies key spatial and temporal patterns using dynamic modes and eigenvalues [19]. Its reliance on mathematical decomposition rather than iterative training allows it to work effectively with sparse data, making it well-suited for solar irradiance forecasting in scenarios where data is limited.

Solar irradiance demonstrates significant variability influenced by several external factors, including seasonal fluctuations and geographic location [20]. This inherent variability necessitates the use of advanced forecasting techniques to enhance the accuracy of predictions. In their study, Mohan et al. [21] proposed a data-driven method for short-term electric load prediction utilizing DMD. Building on this approach, the present study proposes the application of dynamic mode decomposition for short-term solar irradiance forecasting. By adopting this technique, we aim to improve the precision of solar

irradiance predictions using relatively little training data and limited computational cost employing truncated singular value decomposition (SVD) [22].

2. Model Framework

2.1 DMD Mathematical Background

A DMD is a matrix decomposition procedure advanced from linear Koopman operator concept [23]. It is a data-driven method that probes the fundamental dynamics of a given system. The DMD algorithm can extract both temporal and spatial patterns where other methods are limited to either pattern [24]. The proficiency of dynamic mode decomposition in forecasting future solar irradiance by learning the historical solar irradiance data characteristics is leveraged in this paper. The dynamic modes and associated eigenvalues are sufficient to characterize the non-stationary and non-linear behaviour of solar irradiance data. If measurements M is taken over time, for a particular system, two "snapshot" matrices X_1 and X_2 expressed by Eq. (1)–(2) are created, separated by an interval of Δt [21].

$$X_1 = [x_1, x_2, x_3, \dots, x_{M-1}] \in R^{N\times(M-1)} \quad (1)$$

$$X_2 = [x_2, x_3, x_4, \dots, x_M] \in R^{N\times(M-1)} \quad (2)$$

The measurement vectors have a size of N, and the two observation matrices overlap throughout time. By assuming that the system is undergoing gradual changes, it is possible to represent the M th snapshot with Eq. (3) which is a linear combination of the preceding M - 1 snapshots, together with a residual error (r) [21].

$$x_M = a_1x_1 + a_2x_2 + \cdots + a_{M-1}x_{M-1} + r \quad (3)$$

The goal of the algorithm is to calculate the eigen decomposition of the linear operator A using the Koopman approximation [23]. It is expressed as:

$$AX_1 \approx X_2 \Rightarrow A = X_2X_1^{\dagger} \quad (4)$$

where † represents the pseudo-inverse operation. The operator matrix A in Eq. (4) is a time-independent linear approximation to the inherent dynamics of the system. The dynamic modes denoted as ϕ_i , refer to the eigenvectors of matrix A that correspond to a single eigenvalueλ_i. However, in numerous practical scenarios, matrix A will have a substantial number of dimensions, making its eigen decomposition a computationally demanding task. Therefore, Eq. (5)–(6) is used to address the problem, a matrix Z with reduced rank is introduced, which has the same non-zero eigenvalues as matrix A. The following expression can be used to obtain matrix Z.

$$X_2 \approx X_1Z \quad (5)$$

The purpose of the relation above is to show the columns of X_2 as the linear combination of the columns of X_1 and Z. The low-rank matrix Z can be expressed as a companion-type matrix with undefined coefficients $[a_1, a_2, \dots, a_{M-1}]$ as follows,

$$Z = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & a_1 \\ 1 & 0 & \cdots & 0 & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & a_{M-2} \\ 0 & 0 & \cdots & 0 & 1 & a_{M-1} \end{bmatrix} \in R^{(M-1)\times(M-1)} \quad (6)$$

2.2 DMD Algorithm

The DMD algorithm is explained in the following major steps [24],

i. Perform singular value decomposition (SVD) on the observation matrix X_1:

$$X_1 \approx U\Sigma V^H \quad (7)$$

where, $U \in C^{N\times K}$, $\Sigma \in \mathrm{C}^{K\times K}$, $V \in C^{M\times K}$, K is the rank of the reduced SVD approximation to X_1

ii. Calculate the companion-type matrix Z from the reduced SVD components

$$X_2 \cong X_1Z \Rightarrow U\Sigma V^HZ \quad (8)$$

$$\mathrm{Z} = \mathrm{V}\Sigma^{\dagger}\mathrm{U}^{\mathrm{H}}\mathrm{X}_2 \quad (9)$$

iii. Derive the matrix $\tilde{\mathrm{Z}}$, that is analogous to Z,

$$\tilde{Z} = U^HX_2V\Sigma^{\dagger} \quad (10)$$

Since analogous matrices have the identical eigenvalues, $\tilde{Z}$ is used to approximate the eigen decomposition of A.

iv. Carry out the computation of the eigen decomposition of $\tilde{Z}$ to determine the system's dynamic modes and eigen values.

$$AU\Sigma V^H{=}\mathrm{X}_2 \Rightarrow A\mathrm{U}{=}\mathrm{X}_2V\Sigma^{\dagger} \quad (11)$$

Multiply first by U^{H} on both sides to yield

$$U^HA\mathrm{U}{=}\mathrm{U}^HX_2V\Sigma^{\dagger}{=}\tilde{\mathrm{Z}} \quad (12)$$

$$A\mathrm{U}{=}\mathrm{U}\tilde{\mathrm{Z}}{=}\mathrm{U}(W\Omega W^{\dagger}) \Rightarrow A(\mathrm{UW}) = (\mathrm{UW})\Omega \quad (13)$$

where the eigen decomposition of $\tilde{Z}$ is given by $W\Omega W^{\dagger}$. We have the eigenvector matrix expressed by Eq. (14), W given by

$$W = [w_1w_2w_3 \quad \dots w_{M-1}] \quad (14)$$

Ω in Eq. (15) is the diagonal matrix of the eigenvalues which is expressed as

$$\Omega = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{M-2} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_{M-1} \end{bmatrix} \quad (15)$$

v. Formulate the dynamic mode matrixFrom

$$\Phi = UW \Rightarrow \Phi = X_2 V \Sigma^{\dagger} W \qquad (16)$$

The matrix Φ consists of M - 1 columns, each representing an eigenvector of matrix A, with the diagonal matrix Ω containing the eigenvalues of A. Each column of the matrix represents a dynamic mode associated with a unique eigenvalue, λ_i.

3. Methodology

3.1 DMD-Based Irradiance Forecasting Model

This study seeks to employ the DMD model to predict solar irradiance using limited historical data to train. Short-term irradiance prediction using DMD is particularly advantageous due to the model's ability to take into account the underlying multi-level temporal patterns of the system. The proposed DMD-based forecasting method involves the following key steps:

i. Normalizing the dataset

The input solar irradiance data is normalised using Eq. (17). It compresses the data into a consistent range between 0 and 1. It is performed using the formula

$$x_j = \frac{x_i - x_{min}}{x_{max} - x_{min}} \qquad (17)$$

where, $i, j \in [1,2,\ldots,M], x_j \in R$ represents the normalized value, x_{max} and x_{min} represent the maximum and minimum values in x.

ii. Transforming the normalized linear dataset into multi-dimensional data using Hankelization.

To effectively capture the temporal dynamics inherent in solar irradiance time series data, Hankelization is employed to convert the normalized one-dimensional dataset into a structured matrix. This transformation facilitates the capture of time-dependent patterns, such as trends, seasonality, and temporal correlations, by embedding historical observations in overlapping sequences. Each row of the resulting Hankel matrix in Eq. (18) serves as a snapshot of consecutive data points, allowing the model to infer relationships between past and future values. This approach is particularly advantageous for solar irradiance forecasting, where understanding temporal dependencies is critical for achieving accurate predictions. The Hankel matrix, X defined as,

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_L \\ x_2 & x_3 & \cdots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_S & x_{S+1} & \cdots & x_{S+L-1} \end{bmatrix} \qquad (18)$$

Where S = M - L + 1 is the number of rows in the resulting Hankel matrix, M is the total length of the time series data and L is the window length. For this study, this was set as $\frac{M}{2}$. **Figure 1** provides a simplified illustration of this transformation.

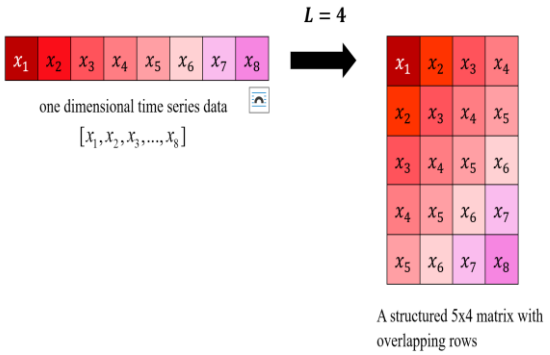


Figure 1. Illustration of Hankelization, employed in the transformation of a one-dimensional time series $[x_1, x_2, x_3, \ldots, x_8]$ into a structured matrix with overlapping rows. Adapted from [25] with modifications.

iii. Performing eigendecompostion and estimating the dynamic modes

The dynamic modes are computed using eigen decomposition using Eq. (7)–(13). Eq. (16) is used to estimate the dynamic mode matrix, with each column ϕ_i representing a dynamic mode as described earlier. By understanding the cyclicity, seasonality, and intrinsic trends in the solar irradiance data, the non-orthogonal dynamic modes are able to describe the behaviour of the system [24]. Visualizing the eigenvalues in the complex plane [24] can be seen to better identify the dynamics, and this is depicted in **Figure 2**. All the eigenvalues lie perfectly across the unit circle indicating the stable dynamic modes. The time-series solar irradiance data's trend is captured by the modes.

iv. Initial prediction and data rearrangement

The forecast of the solar irradiance series data vector, x_P obtained by solving Eq. (19) is based on the calculated dynamic mode matrix Φ,

$$x_P = \Phi\, exp(\Omega t_F) b, x_P \in R^S \qquad (19)$$

where, Ω is the diagonal matrix of eigenvalues, $b = \Phi^{\dagger}\bar{x}_1$ is the initial amplitude of dynamic modes, and t_F is the time of future prediction.

The resulting matrix X_P, converted into a linear dataset P, that integrates both the reconstructed training data and the predicted data by performing the inverse Hankelization process. The forecasted values commence at the index, $i = L + S$. For short-term predictions, the first six values from this index are selected.

v. De-normalization and final solar irradiance forecast

Finally, P is denormalised to the original data scale using Eq. (20)

$$X_{(j)} = \mathrm{P}_{(j)}(P_{max} - P_{min}) + P_{min} \qquad (20)$$

Where $j \in [1,2,\ldots,M]$, $P_{(j)}$represents the elements of the dataset P, and $X_{(j)}$ denotes the de-normalized solar irradiance value.

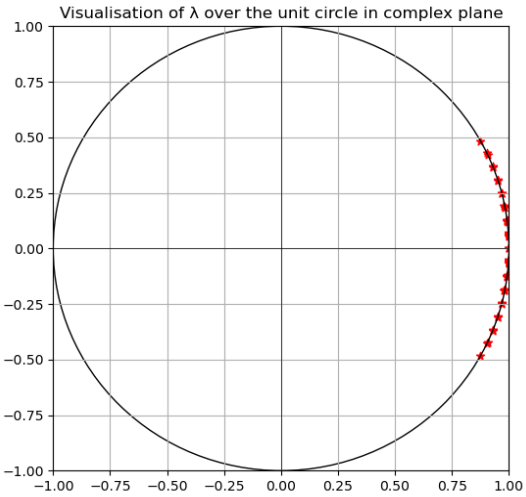


Figure 2. Visualization of the eigenvalues of intra-hour data over the unit circle in the complex plane

3.2 Data Description and Preprocessing

For this study, datasets obtained from two different public sources were utilized:

i. Folsom Dataset

Pedro et al. [26] made the dataset available via the Zenodo repository. It comprises three years (2014–2016) of high-resolution ground-based measurements, including direct normal irradiance and global horizontal irradiance, recorded at 1-minute intervals. In addition, commonly used exogenous variables such as Numerical Weather Prediction forecasts, satellite imagery, sky images, and weather data were also provided. The irradiance data was collected in Folsom (38.642°, -121.148°), a city located in Sacramento County within California's Central Valley shown on the map in **Figure 3**. The region is characterized by a temperate climate with dry, hot summers, classified as Csa under the Köppen climate system [26].

Pedro et al. [26] generated secondary datasets from the primary dataset, featuring irradiance data tailored to various time horizons. For the Folsom dataset, the secondary data utilized included clear sky irradiance for intra-hour (5-minute intervals), intra-day (30-minute intervals), and day-ahead (26-hour intervals) time horizon. This choice of time horizons was made to correspond to the forecasting time horizons used by Pedro et al. [26], making it become more convenient to compare the results obtained.

For intra-hour forecasting, 4000 data points, (about one month of data, 2015-11-27 23:20:00 to 2015-12-31 23:55:00), were used for training the model and a forecast of 6 future data points (2016-01-01 16:00:00 to 2016-01-01 16:25:00) were made. For intra-day forecasting, 1000 data points (about one month, 2015-11-14 20:00:00 to 2015-12-31 23:00:00) were utilized for training and testing the model.

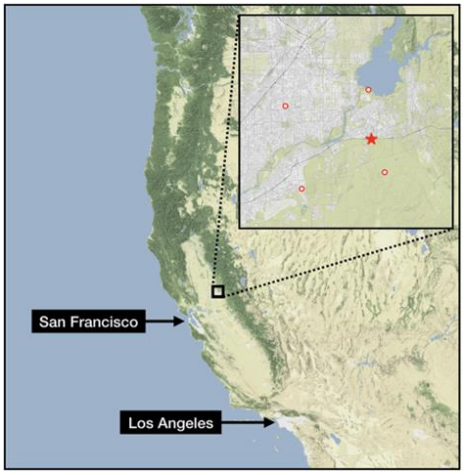


Figure 3. Map showing the area around the Folsom, California site [18].

Six data points from 2016-01-01 16:00:00 to 2016-01-01 18:30:00 at 30-minute intervals were used to validate the performance of the model used for the prediction.

This forecasting horizon is particularly beneficial to utility companies and system operators in system control and real-time electricity bidding [27].

For day-ahead forecasting, 800 data points, 2014-03-30 12:00:00 to 2016-06-08 12:00:00, were used to train and test the model. The short-term irradiance forecasting capability of the model was assessed using the predictions of 6 data points, 2016-01-01 to 2016-01-06 based on all the prior data points.

The Folsom site is specified by the star marker in the insert plot while the other circle markers indicate the four nearest North American Mesoscale (NAM) model grid points [26]. The Folsom dataset presented some challenges, most notably the occurrence of missing data points across various forecast horizons. The missing values, denoted as 'NaN', were particularly evident in the intra-hour dataset, resulting in data gaps that required attention. There were also negligible outliners.

These challenges were addressed through the implementation of adequate preprocessing techniques. The outliners were retained, missing values were excluded by filtering out 'NaN' entries, and the remaining data was normalized using the MinMaxScaler to achieve a standardized range. Additionally, seasonal decomposition was utilized to extract and visualize the underlying trends and periodic patterns in the data. These steps ensured the dataset's integrity and reliability for forecasting, despite its inherent limitations. It is advisable to account for the documented uncertainties when utilizing datasets for forecasting.

ii. National Aeronautics and Space Administration (NASA) Prediction of Worldwide Energy Resources (POWER) Project Dataset

To further verify the results of the study, data from the NASA POWER Project was employed [28]. The utilized dataset consists of Clear Sky Surface Shortwave Downward Irradiance and All Sky Surface Shortwave

Downward Irradiance for hourly and daily time horizons from 2014 to 2023 for Osogbo, Osun state, Nigeria (7.78°N, 4.55°E) shown on the map in **Figure 4**. For the hourly time horizon, the DMD model was trained and tested using 3,656 data points from 2022-01-01 06:00:00 to 2022-10-18 06:00:00. The other benchmarking models were trained and tested using 40,544 entries, 2014-01-01 06:00:00 to 2022-10-18 06:00:00. For the daily time horizon, the DMD model was trained and tested using 725 data points, 2014-01-02 to 2015-12-28 at 24-hour intervals. The other benchmarking models were trained and tested using 11684 entries, from 1984-01-01 to 2015-12-28. The dataset exhibits challenges similar to those encountered with the Folsom dataset, and they were addressed using the same mitigation strategies applied to Folsom's data.

iii. Data Pre-processing

In addition to the Clear Sky Surface Shortwave Downward Irradiance and All Sky Surface Shortwave Downward Irradiance that were used, the clear sky index K_c was computed using Eq. (21).

Figure 4. Map showing Osogbo. The circle marker in the insert plot shows the zoomed in map of Osogbo.

$$K_c = \mathrm{I}/\mathrm{I}_{\mathrm{clear}} \tag{21}$$

where, I is the All-Sky Surface Shortwave Downward Irradiance, and I_{clear} is the Clear Sky Surface Shortwave Downward Irradiance. The computed clear sky index is then added as an additional feature for training the benchmarking models. To ensure uniformity and facilitate further analysis, the input solar irradiance series data $x \in R^M, x_i \in R,\ i = 1,2,\ldots,M$ undergoes a normalization process as described in Eq. (17).

3.3 Benchmark Models Description

Pedro et al. [26] presented some models for benchmarking the performance of solar irradiance forecasting models. In this present work, to evaluate the performance of the proposed DMD model, three well-established regression techniques,Ordinary Least Squares (OLS), Least Absolute Shrinkage and Selection Operator (LASSO), and Ridge regression described by Eq. (22)–(24) were utilized as benchmark models. The performance of the proposed DMD technique is compared against these benchmark models. Each of these models is summarily presented below:

i. Ordinary Least Squares (OLS)

Ordinary Least Square (OLS) is one of the most frequently used methods of regression studies. It aims to minimize the sum of squared residuals between the forecasted values $\hat{y}_i$and the observed values y_i. The OLS objective function is given by,

$$min_{\boldsymbol{\beta}}\left(\sum_{\mathrm{i}=1}^{N}(y_i - x_i^T\beta)^2\right) \tag{22}$$

Where, x_i denotes the feature vector for the i-th observation, β is the vector of coefficients, and N is the number of observations. Although OLS gives unbiased estimates of β, it sometimes exhibits high variance, particularly in the presence of multicollinearity [29].

ii. Least Absolute Shrinkage and Selection Operator (LASSO) Model

The LASSO is a regularization method that performs both variable selection and coefficient shrinkage. It introduces an L1 penalty term that penalizes the sum of the absolute values of the coefficients, encouraging some coefficients to be exactly zero. The objective function for LASSO is expressed as:

$$min_{\boldsymbol{\beta}}\left(\sum_{\mathrm{i}=1}^{N}(y_i\text{-}x_i^T\beta)^2 + \lambda\sum_{\mathrm{j}=1}^{p}\left|\beta_j\right|\right) \tag{23}$$

where β is the vector of coefficients, and β_j represents the j -th coefficient, which quantifies the contribution of the j -th predictor to the response variable. Here, y_i is the observed value, x_i is the predictor vector, N is the number of observations, and p is the number of predictors. The regularization parameter $\lambda \geq 0$ controls the sparsity level, with larger values of λ shrinking moreβ_j values to zero. The optimal value of λ is determined through a grid search combined with k-fold cross-validation on the training dataset.

iii. Ridge

Ridge regression - also known as L2 regularization - is a method that reduces errors caused by overfitting training data. It does this by introducing a penalty term which reduces the coefficient estimates. The objective function for Ridge regression is given by,

$$min_{\boldsymbol{\beta}}\left(\sum_{\mathrm{i}=1}^{N}(y_i - x_i^T\beta)^2 + \lambda\sum_{\mathrm{j}=1}^{p}\beta_j^2\right) \tag{24}$$

where, $\lambda \geq 0$ is the regularization parameter and p is the number of predictors.

The penalty term $\lambda\sum_{\mathrm{j}=1}^{\mathrm{p}}\beta_{\mathrm{j}}^2$ constrains the coefficients β_j to be smaller, thereby reducing overfitting by discouraging large coefficients. This becomes especially useful when the predictors are highly correlated, as it stabilizes the estimation process [29].

3.4 Performance Assessment

The performance of the proposed DMD-based forecasting model is evaluated using three error metrics described with Eq. (25)–(27), selected for their ability to capture distinct aspects of forecast accuracy:

i. Mean Absolute Error (MAE)

This Measures the average magnitude of errors without considering their direction. This metric provides an overall measure of model accuracy and is simple to interpret, making it useful for general performance comparison.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (25)$$

ii. Mean Bias Error (MBE)

This Captures the bias in the model's predictions, indicating whether forecasts are systematically overestimated (positive MBE) or underestimated (negative MBE). This metric is essential for identifying and correcting directional errors in solar irradiance forecasts.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right) \quad (26)$$

iii. Root Mean Square Error (RMSE)

This Highlights the magnitude of larger errors by squaring the residuals, thus penalizing significant deviations more heavily. This makes RMSE particularly useful for applications where large errors have a critical impact.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \quad (27)$$

In all the cases described above, Y_i is the actual value (original data), $\hat{Y}_i$ is the forecasted value while n is the total number of observations.

These metrics collectively provide a comprehensive evaluation of model performance, addressing overall accuracy (MAE), bias (MBE), and error magnitude (RMSE), ensuring a robust assessment of the forecasting model.

4. Results and Discussions

4.1 Folsom Dataset

To demonstrate the effectiveness of the proposed DMD-based solar irradiance forecasting model, the Folsom dataset as described in section 3.2 was utilized. Rather than train with all the data available from 2014 to 2015 as done by Pedro et al. [26], data from within a limited timeframe was selected to serve as training data, depicted by **Figure 5**. This is because the selected data describes the system's dynamics, emphasizes the main trends and periodicities, and sufficiently represents the complexity of the system being studied without introducing unnecessary noise or being too computationally intensive, making it easy for the DMD model to decompose more accurately, especially for the intra-hour and intra-day data.

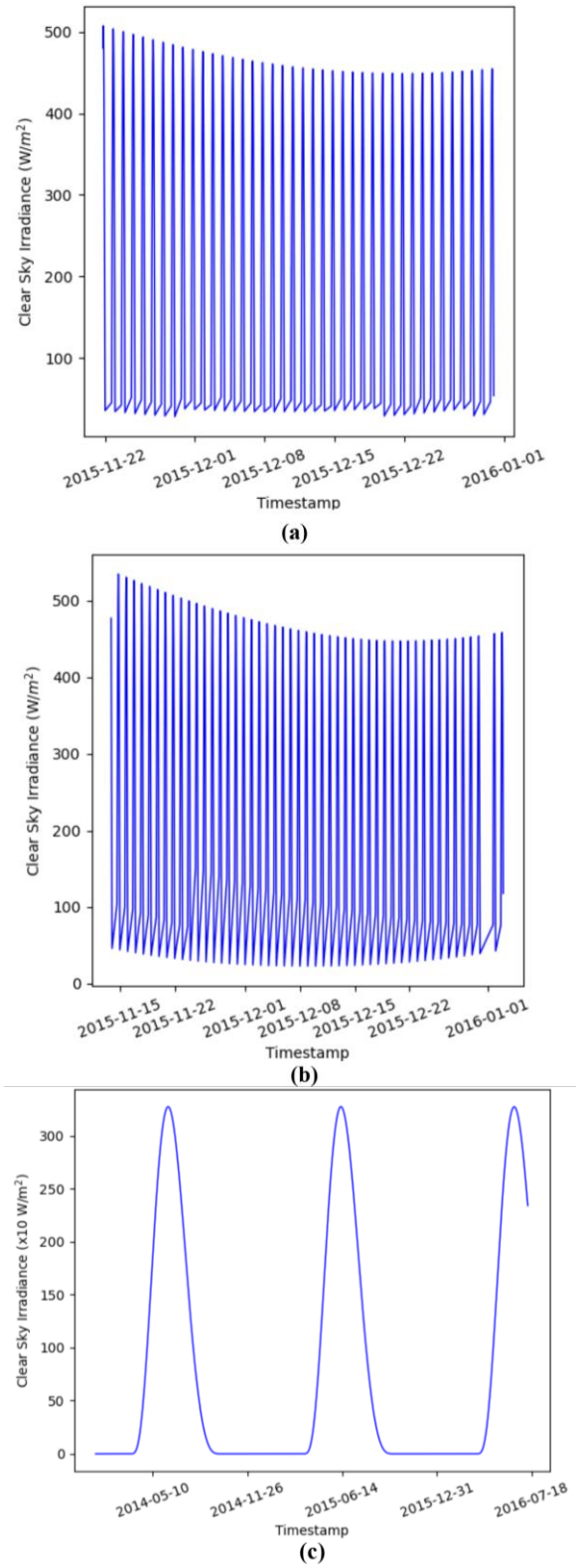


Figure 5. Illustrations of selected solar irradiance data from the Folsom dataset used to train the model. **(a)** Intra-hour **(b)** Intra-day **(c)**Day-ahead

The results obtained for short-term intra-hour solar irradiance forecasting (5–30 minutes) and intra-day forecasting (30 minutes–3 hours) for January 1, 2016, are shown in **Figure 6(a)–(b)** while the results for day-ahead forecasting (1–6 days) are presented in **Figure 6(c)**. The

results obtained are compared with that of linear regression models utilized by Pedro et al. [18] for benchmarking such as OLS, LASSO, and Ridge [18]. The proposed approach outperformed the other models in predicting future solar irradiance as reflected by the significantly lower errors.

As shown in **Table 1**, the DMD model achieved exceptional accuracy, recording a MAE of 12.68 ± 2.6 W/m², which was significantly lower than all the benchmark models. This result highlights the model's consistent ability to provide precise predictions. The MBE of 12.68 ± 2.6 W/m² further confirms its minimal forecasting bias, while the RMSE of 12.95 ± 8.4 W/m² underscores its robustness in handling short-term solar irradiance variability effectively.

For intra-day forecasts, the DMD model exhibited a remarkable performance. With MAE and MBE values of 5.46 ± 2.3 W/m², it achieved an approximate 90% improvement in accuracy compared to benchmark models. The significantly lower RMSE further emphasizes the model's reliability in managing intra-day solar irradiance predictions under varying conditions

Day-ahead results also showed the DMD model to be better than the benchmark models. DMD model recorded MAE and RMSE values being approximately 0.1% those of the benchmark models. While OLS model exhibits a slightly lower MBE of 0.7W/m^2 implying that it slightly overpredicted when compared to -0.90 of the DMD model which slightly underpredicts the GHI. However, the DMD model is still better because it exhibits a lower level of variability with a standard deviation of just 0.3 W/m^2 compared to 7.9 W/m^2 of OLS.

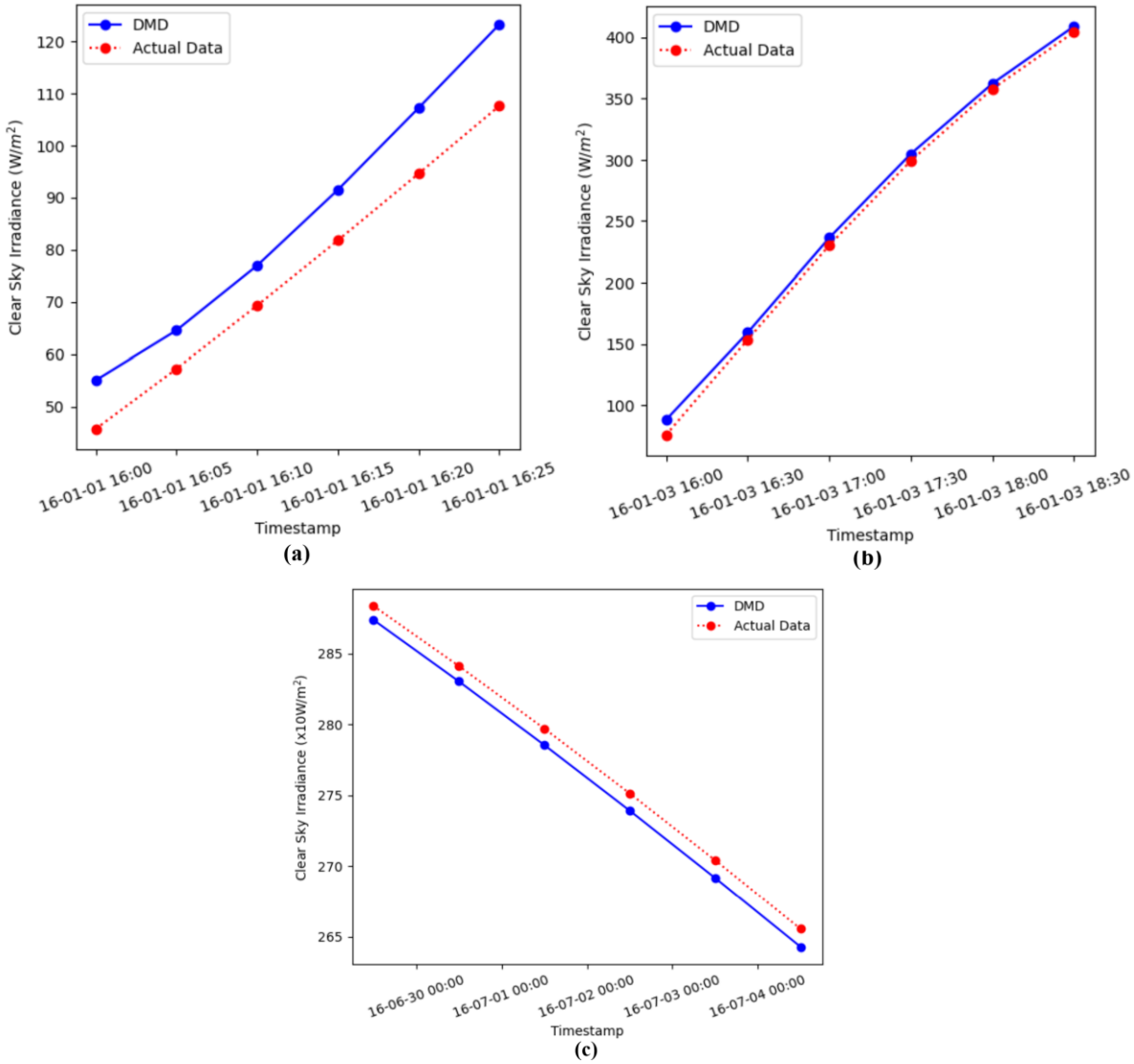


Figure 6. Illustration of the proposed DMD approach forecasting results using data from the Folsom dataset for (**a**) Intra-hour data for 01-Jan-2016 (**b**) Intra-day data for 01-Jan-2016 (**c**) Day-ahead data for July 2016

Table 1 GHI forecast results for the Folsom dataset by horizon and model. For each error metric, the mean ± the standard deviation over all horizons is reported. Adapted from Pedro et al. [26] with modifications

Time Horizon	Model	MAE (W/m²)	MBE (W/m²)	RMSE (W/m²)
Intra-hour	**DMD**	**12.68 ± 2.6**	**12.68 ± 2.6**	**12.95 ± 8.4**
	Pers.	32.3 ± 7.1	0.9 ± 0.5	73.2 ± 11.9
	OLS	33.0 ± 6.9	-1.9 ± 2.6	67.5 ± 9.8
	Ridge	33.0 ± 6.9	-1.9 ± 2.6	67.5 ± 9.8
	Lasso	33.0 ± 6.9	-1.9 ± 2.6	67.5 ± 9.8
Intra-day	**DMD**	**5.46 ± 2.3**	**5.46 ± 2.3**	**5.92 ± 5.5**
	Pers.	49.9 ± 13.7	8.0 ± 6.0	89.6 ± 20.3
	OLS	50.1 ± 11.1	-16.8 ± 14.4	89.2 ± 20.6
	Ridge	50.1 ± 11.1	-16.8 ± 14.4	89.2 ± 20.6
	Lasso	50.1 ± 11.1	-16.9 ± 14.4	89.2 ± 20.6
Day-ahead	**DMD**	**1.49 ± 0.18**	**-1.49 ± 0.18**	**1.50 ± 0.72**
	NAM	85.1 ± 21.6	-20.5 ± 62.3	110.0 ± 29.3
	OLS	72.0 ± 42.2	0.7 ± 7.9	101.0 ± 56.7
	Ridge	70.4 ± 40.9	0.9 ± 8.2	98.5 ± 54.6
	Lasso	70.9 ± 41.4	2.5 ± 9.3	96.9 ± 53.2

4.2 Osogbo Dataset

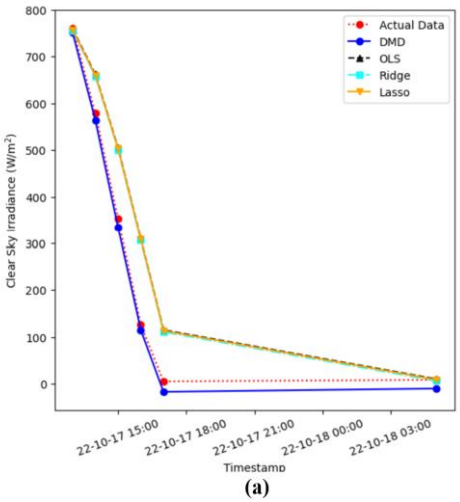
To further evaluate the effectiveness of the DMD model, data from Osogbo (7.78°N, 4.55°E), Nigeria, obtained through the NASA POWER Project, was utilized. The dataset comprises hourly and daily measurements of clear-sky surface shortwave downward irradiance for this location.

To train the model for intra-day irradiance prediction, data spanning from 2022-01-01 06:00:00 to 2022-10-17 06:00:00 was utilized while the model was trained using data collected between 2014-01-02 and 2015-12-22 for day-ahead solar irradiance forecasting. The benchmark the models - OLS, LASSO, and Ridge, were trained on a larger dataset for better performance. The benchmark models were trained with data from 2014-01-01 06:00:00 to 2022-10-17 13:00:00 for intra-day forecasting, while for day-ahead irradiance forecasting, the training period spanned from 2020-07-01 to 2022-10-01.

The comparative results obtained for intra-day and day-ahead irradiance forecasting for the Osogbo dataset were presented visually in **Figure 7**. As observed by Pedro et al. [26] in their analysis of the Folsom, dataset, the OLS, Ridge, and LASSO models produced similar results.This is most likely due to the high correlation between the features used to train the model. Though Ridge and Lasso are expected to deal with the multicollinearity and the shrinkage penalty of models, λ even after cross-validation was likely very low or equal to zero, resulting in similar predictions across all three models.

It can be observed that for the intraday Folsom forecast, the DMD model also performed well, yielding near-perfect predictions. The other models, however, fall short of the mark more significantly. The extremity of the variance between the proposed DMD model and the other models could be observed in **Table 1**, which for MAE, the DMD model performs significantly better with values less than 10% of those of the other models. MAE is about 25% of the values obtained from the benchmark models as shown in **Table 2**, which can generally be described as similar to the results obtained when comparing the error metrics of the models provided by Pedro et al. [26] with those of the DMD model. The results further validated the model's superiority.

The forecasts of all the benchmark models for the day-ahead irradiance showed more visual discrepancy from the actual data, particularly after the third day as illustrated in **Figure 7(b)**. However, The DMD model outperforms the benchmark models across the time horizons and datasets, signifying its competence in short duration solar irradiance predictions.



(a)

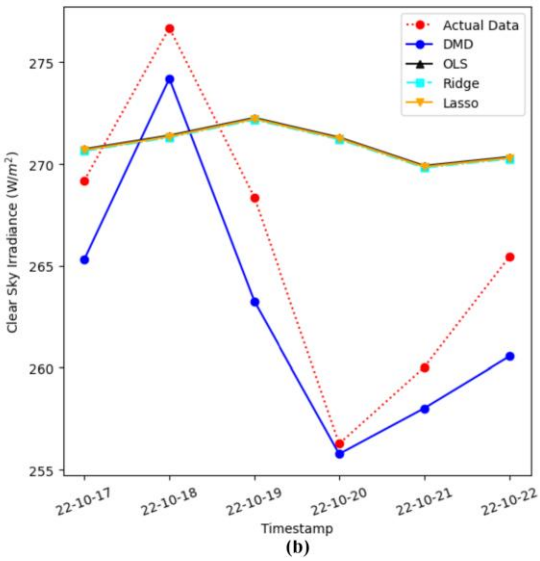


Figure 7. Illustration of forecasting results using the proposed DMD approach compared with other models on NASA POWER Project data for **(a)** hourly irradiance data for 17-October 2022 and **(b)** daily irradiance data for 17–22 October 2022.

Figure 8. present a comparison of the error metrics—MAE, MBE, and RMSE across the DMD, OLS, Ridge, and LASSO models for different forecasting horizons. **Figure 8(a)**. illustrates the results for intra-day solar irradiance forecasting, while **Figure 8(b)**. focuses on day-ahead forecasts. The DMD model demonstrated superior performance, achieving significantly lower error values across all metrics when compared to the benchmark models. These results highlight the robustness and reliability of the DMD model in handling diverse solar irradiance forecasting scenarios.

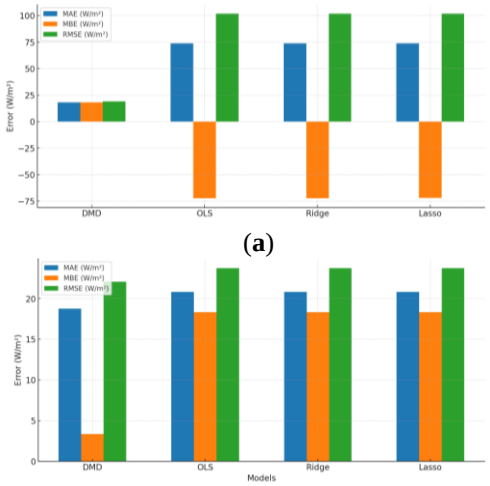


Figure 8 A bar chart illustrating error metrics (MAE, MBE, and RMSE) for the forecasting models: **(a)** intra-day solar irradiance forecasting and **(b)** day-ahead solar irradiance forecasting. The DMD model demonstrates superior performance across both horizons compared to the benchmark models

Table 2 GHI forecast results by horizon and model. For each error metric, we report the mean ± the standard deviation over all horizons for the NASA POWER Project

Time Horizon	Model	MAE (W/m²)	MBE (W/m²)	RMSE (W/m²)
Intra-day	**DMD**	**18.18 ± 5.9**	**18.18 ± 5.9**	**19.10 ± 14.5**
	OLS	73.80 ± 70.1	-72.24 ± 71.7	101.79 ± 113
	Ridge	73.80 ± 70.1	-72.24 ± 71.7	101.79 ± 113
	Lasso	73.91 ± 70.1	-72.10 ± 71.9	101.85 ± 113
Day-ahead	**DMD**	**18.75 ± 2.8**	**3.33 ± 2.8**	**22.08± 7.28**
	OLS	21.25 ± 4.9	18.33 ± 6.45	23.75 ± 8.69
	Ridge	21.25 ± 4.9	18.33 ± 6.45	23.75 ± 8.69
	Lasso	21.25 ± 4.9	18.33 ± 6.45	23.75 ± 8.69

5. Conclusion

A data-driven approach for short-term solar irradiance prediction using DMD has been presented. The intrinsic capability of the DMD-based algorithm to extract the underlying dynamics of the time-reliant solar irradiance data was utilized for precise future solar irradiance forecasting. To validate the efficiency of the presented framework, an extensive comparison with other benchmarking models, including OLS, LASSO, and Ridge model was conducted using data from two public data sources. The precision and efficiency of the proposed model when compared to the benchmark models are satisfactory. The model's implementation is straight forward. The conspicuously lower values of the employed error metrics (MAE. MBE and RMSE) points to DMD's robustness and efficiency, thereby validating the forecasting performance of the presented model. Therefore, the presented DMD-based framework is demonstrated to be suitable for short term solar irradiance forecasting, even when utilizing relatively limited training data.

However, DMD has some practical limitations that should be considered. First, it is highly dependent on data quality; missing values or noise in the dataset can significantly affect its accuracy, therefore, proper data preprocessing is often required to address this issue. Second, the model requires careful tuning of parameters, such as the embedding window length and the number of dynamic modes retained, which can be computationally intensive. Additionally, DMD is most

effective for short- to medium-term forecasting, as its reliance on linear dynamics can limit its ability to capture complex, nonlinear behaviors in very long-term predictions. Finally, while DMD is computationally efficient for small- to moderate-sized datasets, scaling to high-dimensional or extremely large datasets can lead to significant computational demands during singular value decomposition. Addressing these limitations in future research, such as through hybrid approaches that combine DMD with machine learning techniques, may further enhance its applicability to diverse forecasting scenarios.

6. References

[1] R. K. Koech, M. Kigozi, A. Bello, P. A. Onwualu and W. O. Soboyejo, "Recent advances in solar energy harvesting materials with particular emphasis on photovoltaic materials," in *2019 IEEE PES/IAS Power Africa*, Abuja, Nigeria, Aug. 20–23, 2019, pp. 627–632, doi: 10.1109/PowerAfrica.2019.8928859.

[2] V. Sekhar and P. Pradeep, "A review paper on advancements in solar PV technology, environmental impact of PV cell manufacturing," *International Journal of Advanced Research in Science, Communication and Technology*, vol. 8, no. 1, pp. 485–492, 2021, doi: 10.48175/IJARSCT-1887.

[3] D. K. Chaturvedi and I. Isha, "Solar power forecasting: A review," *International Journal of Computer Applications*, vol. 145, no. 6, pp. 28–50, 2016, doi: 10.5120/ijca2016910728.

[4] M. Díaz, O. Nicolis, J. Marín and S. Baran, "Post-processing methods for calibrating the wind speed forecasts in central regions of Chile," *Annales Mathematicae Et Informaticae*, vol. 53, pp. 93–108, 2021, doi:10.33039/ami2021.03.012.

[5] J. Alonso-Montesinos, F. J. Batlles and C. Portillo, "Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images," *Energy Conversion and Management*, vol. 105, pp. 1166–1177, 2015.

[6] J. -H. Kim, P. A. Jimenez Munoz, M. Sengupta, J. Yang, J. Dudhia, S. Alessandrini and Y. Xie, "The WRF-solar ensemble prediction system to provide solar irradiance probabilistic forecasts," *IEEE Journal of Photovoltaics*, vol. 12, no. 1, pp. 141–144, 2021, doi: 10.1109/JPHOTOV.2021.3117904.

[7] Q. Li, D. Zhang and K. Yan, "A solar irradiance forecasting framework based on the CEE-WGAN-LSTM model," *sensors*, vol. 23, no. 5, 2023, Art. no. 2799, doi: 10.3390/s23052799.

[8] R. Marquez, H. T. C. Pedro, and C. F. M. Coimbra, "Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs," *Solar Energy*, vol. 92, pp. 176–188, 2013, doi: 10.1016/j.solener.2013.02.023.

[9] Y. Wang, Y. X. Shen, S. W. Mao, G. Q. Cao and R. M. Nelms, "Adaptive learning hybrid model for solar intensity forecasting," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1635–1645, 2018, doi: 10.1109/TII.2017.2789289.

[10] P. K. Ray, A. Bharatee, P. S. Puhan and S. Sahoo, "Solar irradiance forecasting using an artificial intelligence model," in *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, Hyderabad, India, Jul. 21–23, 2022, pp. 1–5doi: 10.1109/ICICCSP53532.2022.9862494.

[11] R. Azimi, M. Ghayekhloo, and M. Ghofrani, "A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting," *Energy Conversion and Management*, vol. 118, pp. 331–344, 2016, doi: 10.1016/j.enconman.2016.04.009.

[12] A. K. Yadav and S. S. Chandel, "Solar radiation prediction using artificial neural network techniques: A review," *Renewable & Sustainable Energy Reviews*, vol. 33, pp. 772–781, 2014, doi: 10.1016/j.rser.2013.08.055.

[13] Y. Yu, J. Cao and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019, doi: 10.1109/ACCESS. 2019.2946057.

[14] M. Aslam, J. -M. Lee, H. -S. Kim, S. -J. Lee and S. Hong, "Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study," *energies*, vol. 13, no. 1, 2020, Art. no. 147, doi: 10.3390/en13010147.

[15] B. Farsi, M. Amayri, N. Bouguila and U. Eicker, "On Short-Term Load Forecasting Using Machine Learning Techniques and a Novel Parallel Deep LSTM-CNN Approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021, doi: 10.1109/ACCESS.2021.3060290.

[16] S. Salah, A. ElSayed, O. Khaled, M. Deif and R. Elgohary, "Advancing Space Weather Forecasting: A Comparative Analysis of AI Techniques for Predicting Geomagnetic Storms," *International Integrated Intelligent Systems*, vol. 1, no. 2, 2024, doi: 10.21608/iiis.2024.357835.

[17] E. Doğan, "LSTM training set analysis and clustering model development for short-term traffic flow prediction," *Neural Computing and Applications*, vol. 33, no. 17, pp. 11175–11188, 2021, doi: 10.1007/s00521-020-05564-5.

[18] S. Gugaliya, D. Durgesh, S. Kumar and A. Sheikh, "Data driven approach for long-term forecasting of renewable energy generation," in *2022 4th Global Power, Energy and Communication Conference (GPECOM)*, Nevsehir, Turkey, Jun. 14–17, 2022, pp. 383–388, doi: 10.1109/GPECOM55404.2022.9815777.

[19] S. Akshay, K. P. Soman, N. Mohan and S. Sachin Kumar, "Dynamic mode decomposition and its application in various domains: An overview," in *Applications in Ubiquitous Computing*, R. Kumar and S. Paiva, eds., Cham, Switzerland: Springer

Nature Switzerland AG, 2021, ch. 2, sec. 1, pp. 121–132.

[20] A. Forstinger, S. Wilbert, A. R. Jensen, .B. Kraas, C. F. Peruchena, C. A. Gueymard, D. Ronzio, D. Yang, E. Collino, J. P. Martinez *et al.*, "Worldwide Benchmark of Modelled Solar Irradiance Data," IEA-PVPS, NSW, Australia, Rep. no. IEA-PVPS T16-05: 2023, Jul. 14, 2024, [Online]. Available: https://iea-pvps.org/wp-content/uploads/2023/06/IEA-PVPS-T16-05-2023-Worldwide-Benchmark-Modelled-Solar.pdf.

[21] N. Mohan, K. P. Soman and S. Sachin Kumar, "A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model," *Applied Energy*, vol. 232, pp. 229–244, 2018, doi: 10.1016/j.apenergy.2018.09.190.

[22] S. Hsieh, K. Liu and K. Yao, "Estimation of multiple sinusoidal frequencies using truncated least squares methods," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 990–994, 1993, doi: 10.1109/78.193242.

[23] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, 2010, doi: 10.1017/S0022112010001217.

[24] J. N. Kutz, S. L. Brunton, B. W. Brunton and J. L. Proctor, "Dynamic Mode Decomposition: An Introduction," in *Dynamic Mode Decomposition: Data-Driven Modelling of Complex Systems*, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2016, ch. 1, sec. 1.1, pp. 1–3.

[25] F. Sedighin and A. Cichocki, "Image completion in embedded space using multistage tensorring decomposition," *Frontiers in Artificial Intelligence*, vol. 4, 2021, Art. no. 687176, doi: 10.3389/frai.2021.687176

[26] H.T.C. Pedro, D.P. Larson and C. F. M. Coimbra, "A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods," *Journal of Renewable and Sustainable Energy*, vol. 11, no. 3, 2019, Art. no. 036102, doi: 10.1063/1.5094494.

[27] C. Huang, L. Wang and L. L. Lai, "Data-driven short-term solar irradiance forecasting based on information of neighboring sites," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9918–9927, 2018, doi: 10.1109/TIE.2018.2856199

[28] *NASA Prediction of Worldwide Energy Resources (POWER)*, NASA Langley Research Center, Feb. 27, 2024. [Online]. Available: https://power.larc.nasa.gov/data-access-viewer/.

[29] D. C. Montgomery, E. A. Peck and G. G. Vining, "Multicollinearity," in *Introduction to Linear Regression Analysis*, 5th ed., Hoboken, NJ, USA: Wiley, 2012, ch. 9, sec. 9.4, pp. 292–302