*Clean Energy*, 2024, Vol. 8, No. 1, 100–110
https://doi.org/10.1093/ce/zkad075
Advance access publication 10 January 2024
**Research Article**

# A hybrid machine-learning model for solar irradiance forecasting

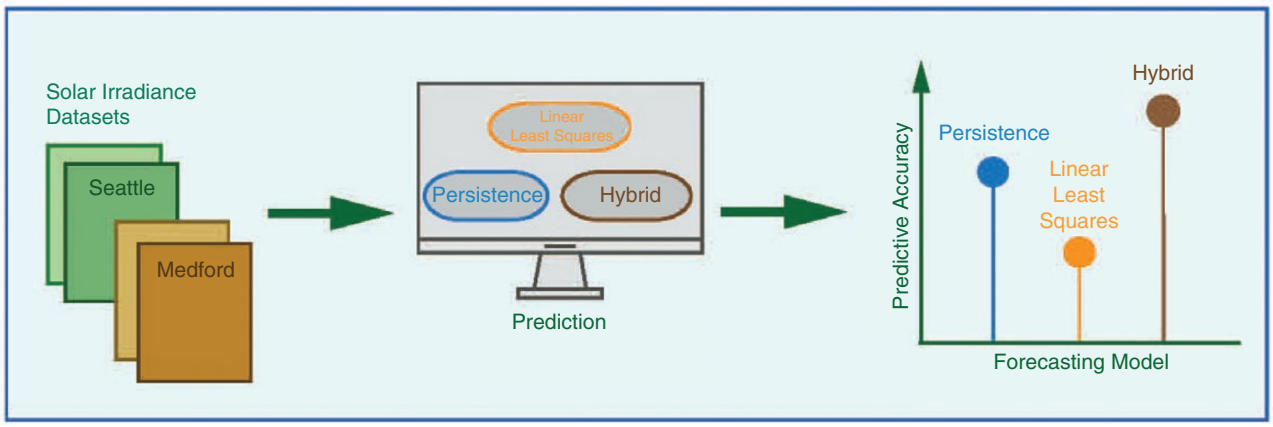
Ameera M. Almarzooqi[1], Maher Maalouf[1,*], Tarek H.M. El-Fouly[2], Vasileios E. Katzourakis[3], Mohamed S. El Moursi[2] and Constantinos V. Chrysikopoulos[3,4]

[1]Department of Industrial and Systems Engineering, Khalifa University, Abu Dhabi, PO Box 127788, UAE
[2]Electrical Engineering and Computer Science Department, Khalifa University Abu Dhabi, UAE
[3]Department of Civil Infrastructure and Environmental Engineering, Khalifa University, Abu Dhabi, PO Box 127788, UAE
[4]School of Chemical and Environmental Engineering, Technical University of Crete, Chania, Greece
[*]Corresponding author. E-mail: maher.maalouf@ku.ac.ae

## Abstract

Nowcasting and forecasting solar irradiance are vital for the optimal prediction of grid-connected solar photovoltaic (PV) power plants. These plants face operational challenges and scheduling dispatch difficulties due to the fluctuating nature of their power output. As the generation capacity within the electric grid increases, accurately predicting this output becomes increasingly essential, especially given the random and non-linear characteristics of solar irradiance under variable weather conditions. This study presents a novel prediction method for solar irradiance, which is directly in correlation with PV power output, targeting both short-term and medium-term forecast horizons. Our proposed hybrid framework employs a fast trainable statistical learning technique based on the truncated-regularized kernel ridge regression model. The proposed method excels in forecasting solar irradiance, especially during highly intermittent weather periods. A key strength of our model is the incorporation of multiple historical weather parameters as inputs to generate accurate predictions of future solar irradiance values in its scalable framework. We evaluated the performance of our model using data sets from both cloudy and sunny days in Seattle and Medford, USA and compared it against three forecasting models: persistence, modified 24-hour persistence and least squares. Based on three widely accepted statistical performance metrics (root mean squared error, mean absolute error and coefficient of determination), our hybrid model demonstrated superior predictive accuracy in varying weather conditions and forecast horizons.

## Graphical Abstract




**Keywords:** solar power generation; kernel ridge regression; hybrid model; forecasting

## Introduction

The application of renewable energy sources is receiving significant attention in the twenty-first century. Renewable and sustainable energy technologies are becoming increasingly prevalent worldwide [1]. One of the most popular forms of clean energy generation is solar photovoltaic (PV). According to the International Renewable Energy Agency [2], the global total capacity of solar PV is expected to reach 14 000 gigawatts (GW) by 2050. As PV power plants are increasingly contributing to global power grids, they not only diversify energy sources, but also mitigate environmental concerns. However, this transition also introduces

**Received:** 12 July 2023. **Accepted:** 14 September 2023
© The Author(s) 2024. Published by Oxford University Press on behalf of National Institute of Clean-and-Low-Carbon Energy
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

technical challenges, including issues related to grid stability, power quality, generation dispatches and the determination of spinning reserves (SRs) [3, 4]. These challenges are due to the variable operating conditions in PV power systems, which include factors such as intermittent weather [5, 6].

Increased power transients result in fluctuating power output, which can have a major impact on grid stability and make it difficult to schedule and dispatch power generation economically [7, 8]. Traditionally, such fluctuations require an increase in the SR of the system, thereby increasing the overall cost [9, 10]. Therefore, it is essential to predict the power production of renewable energy resources for accurate planning of SR requirements. This enables system operators to optimize the required SR and efficiently schedule power dispatch [11]. Forecasting power output allows system planners to explore other ways of mitigating fluctuations, such as installing energy storage systems [12, 13]. Predicting the generated power for PV installations can be performed mainly by: (i) directly using historical data of generated power or (ii) using historical data of influencing parameters (solar irradiance and ambient temperatures) with an accurate physical model for the PV installation. In the former, data collected on the basis of historical time series for actual power generation from a PV facility were used directly to predict future PV production. In the latter method, solar irradiance and ambient temperature data were first predicted up to the target horizons and then the predicted values were used as inputs for a physical model of the PV facility, which related the solar irradiance and ambient temperature to the temperatures of the PV cells and the output power, in order to predict the generated power of the PV facility [14, 15]. Furthermore, the latter method allowed the identification of different factors and parameters that improve the accuracy of the prediction models, including, but not limited to, the effect of scheduled maintenance/cleaning in large PV farms, detailed physical parameters of PV cells considering conversion efficiency and photonic characteristics [16].

The autoregressive integrated moving average (ARIMA) model has been effective in predicting solar radiation on a 24-hour horizon with high resolution [17, 18]. In contrast, advanced machine-learning methods such as artificial neural networks (ANN) have shown promise in both short-term (1–3 hours) and medium-term (12–24 hours) solar irradiance prediction [19]. For example, ANN achieved high coefficient of determination ($R^2$) values of 0.96 and 0.97 for cloudy and sunny days, respectively [19]. Another ANN model, designed to forecast 24-hour-ahead solar irradiance for a grid-connected solar power plant in Italy, reported prediction accuracies of 94% for cloudy days and 98% for sunny days [20].

In addition to ARIMA and ANN, the support vector machine (SVM) method has shown effectiveness in forecasting the PV output. Shi *et al.* [21] applied SVM as a classification problem using four classifications for days, namely cloudy, foggy, sunny and rainy, to classify the model. SVM performed poorly using the cloudy-data model with a mean relative error (MRE) of 12.42% compared with the sunny-data model with an MRE of 4.85%.

When it comes to regression-based forecasting, SVM surpassed ANN, achieving an $R^2$ value of 0.9562 compared with 0.9264 for the ANN on a data set that combines cloudy and sunny days [22]. Furthermore, least-squares support vector machines (LS-SVM) used a variety of meteorological factors as input and achieved impressive 10-minute solar irradiance forecasts with an $R^2$ of 0.94 [23]. Ekici [24] also applied LS-SVM for 24-hour-ahead predictions, achieving a high $R^2$ value of 0.9929. Furthermore, SVM was also combined with other techniques to improve prediction accuracy compared with the ANN method, from a value of $R^2$ of 0.8564 to 0.942 [25].

Despite their success, SVM and ANN have limitations. ANN require large data sets, while SVM and LS-SVM are computationally expensive and lack interpretability, especially in probabilistic forecasts [26].

Recently, hybrid models have emerged as powerful solutions that combine the strengths of multiple methods. For example, Castillo-Rojas *et al.* [27] developed an efficient hybrid model using recurrent neural networks and ANN (RNN–ANN) for short-term PV power generation forecasts. This model integrated a RNN with long short-term memory (LSTM) and ANN using a multilayer perceptron (MLP), resulting in improved predictive accuracy. In follow-up work by Castillo-Rojas *et al.* [28], the authors demonstrated the usefulness of a hybrid model (RNN–ANN) with multiple input variables through two types of models: one that used both energy records and meteorological data, and the other that used only meteorological data. The authors showed that the best of both models efficiently predicted the weekly energy generation of the solar plant.

Another study, by Xu *et al.* [29], highlighted the sequential LSTM ensemble approach with optimized topology for short-term multivariate wind power forecasting and evaluated various methods for wind and PV generation. Li *et al.* [30] developed a gated recurrent unit (GRU)–convolutional neural network (CNN)-based scenario forecast method to generate weekly PV power scenarios. Carrera *et al.* [31] proposed three networks: a deep feedforward network (DFN) for weather forecasting (fcst-DFN), an LSTM for weather observations (obs-LSTM) and a hybrid network for both (PVHybNet). The authors demonstrated the effectiveness of a combined network that used both weather forecasts and recent weather observations. They also demonstrated that the hybrid model outperformed several machine-learning models. Zang *et al.* [32] developed a hybrid CNN model combined with LSTM to predict global horizontal irradiance 1 hour ahead. The authors demonstrated that the CNN–LSTM model outperformed CNN and LSTM separately.

Despite these advances, forecasting solar irradiance remains challenging due to its sensitivity to various meteorological factors [33]. Many of these machine-learning models are also commonly combined with statistical models for better performance, while some other studies used numerical weather prediction (NWP) approaches to improve accuracy, which starts to decrease with increasing time horizons [34]. However, a common limitation of the aforementioned machine-learning models is associated with unexplainable and unreliable forecasts during cloudy weather, when solar irradiance fluctuates greatly [35–37].

Maalouf and Homouz [38] developed the truncated-regularized kernel ridge regression (TR–KRR) method that could achieve the same or better accuracy than the SVM, while significantly reducing computational time and improving the interpretability of the predictions. TR–KRR has been shown to be effective in various applications [39–44].

Drawing from the computational efficiency of the TR–KRR model, this study introduces a novel hybrid method for time-series solar irradiance forecasting. We validated our approach using data sets from Seattle (latitude: 47.6° N, longitude: 122° W, elevation: 49 metres) and Medford (latitude: 42.3° N, longitude: 122.8° W, elevation: 405 metres), USA, encompassing both sunny and cloudy conditions. Providing both computational speed and predictive accuracy, the model was particularly designed for short- to medium-term forecasts. It successfully addressed the long-standing challenge of accurately predicting solar irradiance

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

during fluctuating cloudy weather. Consequently, the model can facilitate more efficient power output scheduling and hence can reduce both reservation and operational costs.

In this study, we address the pressing need for a robust, accurate and efficient forecasting model for PV power generation. Our hybrid framework integrates the TR–KRR algorithm with conjugate gradient optimization and grid search methods, thereby forming a scalable solution for solar irradiance forecasting. To validate its effectiveness, the proposed model was compared with traditional persistence and LS regression-based forecasting approaches. Our results establish a reliable and accurate method for solar irradiance forecasting. To our knowledge, such an investigation has not been previously presented in the literature.

# 1 Weather factors affecting PV power generation

PV power generation is determined by multiple intermittent weather factors, including solar irradiance, ambient temperature, humidity, wind speed and cloud cover [45]. An increase in solar irradiance contributes to an increase in the generated power with a strong correlation, although the temperature is inversely proportional to the generated power. Cloud cover is a significant factor affecting PV power generation. The type and motion of the cloud affect the amount of solar energy that passes through the PV modules. Due to the complex motion patterns of clouds, cloud data usually show a higher correlation with solar irradiance for short-term forecast horizons [46].

Humidity reduces the level of solar irradiance that reaches the PV module due to the refraction, reflection or diffraction caused by particles of water vapour in the air [47]. Wind speed can influence PV performance either positively or negatively. It provides natural cooling and convection, thereby reducing in the effect of humidity. The inclusion of wind speed parameters in PV performance prediction models has been shown to improve accuracy [48]. However, the wind also spreads sand and dust that could accumulate on the surface of the PV panel. The higher the wind speed, the larger the decrease in the performance of PV cells due to dust accumulation [49]. Soiling is a term used to describe the effect of the accumulation of dust and dirt on PV modules. It reduces the generation power of PV modules and its effect is magnified where there is no rainfall [50]. The reduction in power generation caused by soiling in the gulf climate can reach 45% if the module is cleaned every 2 months [51]. However, the correlation of meteorological parameters and PV power output greatly depends on the geographical location and will vary between different locations.

# 2 Methods

In this research, the following forecasting models will be employed: persistence, linear LS regression, ridge regression and KRR. Furthermore, accurate solar irradiance forecasts can be obtained using the proposed hybrid TR–KRR framework.

## 2.1 Persistence model

The *persistence* model, commonly known as the Naïve method, is based on the concept that present and past values are highly connected. It is the simplest time-series forecasting model, where the assumption is that the predicted value is equal to the value of the previous point, such that:

$$G(T) = G(T-1), \tag{1}$$

where $G(T)$ is the solar irradiance at time point $T$. In PV forecasting, the persistence model is used as a baseline and modified to reflect the day ahead, where the prediction is set as being equal to the value of the same time on the day before [52], which can be expressed by:

$$G(T) = G(T-24), \tag{2}$$

and is referred to as the modified 24-hour persistence model.

## 2.2 LS regression model

The LS regression is the most fundamental regression model. Linear regression uses the simplest form of the LS method, where the relationship between the variables is described by a line, which in matrix form can be expressed as:

$$\mathbf{y} = \boldsymbol{X\beta} + \epsilon, \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^N$ is the dependent variable vector and $\boldsymbol{X} \in \mathbb{R}^{N\times d}$ is the data matrix representing $d-1$ independent variables and $N$ rows, $\boldsymbol{\beta}$ is the coefficient vector and $\epsilon \in \mathbb{R}^N$ is the random error vector [42, 53]. The solution to the minimization of the sum of squared errors (SSE) with respect to $\boldsymbol{\beta}$ that is given by:

$$SSE = (\mathbf{y} - \boldsymbol{X\beta})^T (\mathbf{y} - \boldsymbol{X\beta}), \tag{4}$$

is

$$\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{y}. \tag{5}$$

## 2.3 KRR model

Ridge regression addresses the problem of multicollinearity of linear models by adding a regularization parameter $\lambda \geq 0$ to the SSE function, as follows:

$$SSE = f(\beta) = \frac{1}{2}(\mathbf{y} - \boldsymbol{X\beta})^T (\mathbf{y} - \boldsymbol{X\beta}) + \frac{\lambda}{2}\left|\left|\boldsymbol{\beta}^T\boldsymbol{\beta}\right|\right|^2. \tag{6}$$

The solution to Equation (6) is given by:

$$\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{\lambda I}_d\right)^{-1}\boldsymbol{X}^T\mathbf{y}, \tag{7}$$

where $\boldsymbol{I}$ is a $d \times d$ identity matrix. The vector $\boldsymbol{\beta}$ can be expressed as a linear combination of the data points so that $\beta = X^T\alpha$. Now, rewriting the model in Equation (3) results in:

$$\mathbf{y} = \boldsymbol{XX}^T\boldsymbol{\alpha} + \epsilon = \boldsymbol{G\alpha} + \epsilon, \tag{8}$$

where the matrix $G = \boldsymbol{XX}^T$ is a Gram matrix.

KRR addresses nonlinearity in data by applying a mapping function $\phi(.)$ that maps the data into a higher-dimensional space. The kernel function uses the dot products such that $K = k\left(\mathbf{x}_i,\mathbf{x}_j\right) = \phi(\mathrm{x}_i),\ \phi(\mathrm{x}_j)$ [38, 54]. The most widely used kernels are linear, polynomial and radial basis function (RBF) kernels. For this study, we use the RBF kernel, which is given by:

$$k\left(\mathrm{x}_i,\ \mathrm{x}_j\right) = e^{-\frac{1}{2\sigma^2}||\mathrm{x}_i - \mathrm{x}_j||^2}, \tag{9}$$

where $\sigma > 0$ is the kernel width. Replacing the Gramian matrix $\boldsymbol{G}$ in Equation (8) with the kernel $K$, the KRR model is then:

$$\mathrm{y} = \boldsymbol{K\alpha} + \epsilon. \tag{10}$$

The KRR function to be minimized with regard to the dual variable $\alpha$ becomes:

$$f(\alpha) = \frac{1}{2}(\mathbf{y} - \boldsymbol{K\alpha})^T (\mathbf{y} - \boldsymbol{K\alpha}) + \frac{\lambda}{2}||\boldsymbol{\alpha}||^2, \tag{11}$$

to which the solution would be:

$$\boldsymbol{\alpha} = (\boldsymbol{K} + \boldsymbol{\lambda I}_N)^{-1}\mathbf{y}. \tag{12}$$

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

Now, if the matrix $(K + \lambda I_N)$ is dense, then the computational time would have a complexity of $O(N^3)$, and hence iterative methods would be the most suitable to compute $\alpha$.

## 2.4 Proposed hybrid TR–KRR framework

Maalouf and Homouz [38] applied the linear conjugate gradient (CG) method to the Newton step, leading to the development of the TR–KRR algorithm. This TR–KRR algorithm (see Algorithm 1) has been shown to be much faster than SVM without compromising accuracy [38].

**Algorithm 1:** Linear CG algorithm to compute the optimal $\boldsymbol{\alpha}$

Data: $A = (\boldsymbol{K} + \lambda \boldsymbol{I}_N), \boldsymbol{b} = \boldsymbol{y}$
Output: optimal alpha: $\hat{\alpha}$
$r^{(0)} = \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{\alpha}}^{(0)}$ — Initialize the residual
$c = 0$
while $||r^{(c+1)}||^2 > tolerance$ and $c \leq max \quad iterations$
if $c = 0$, then $z = 0$
else $z^{(c)} = \frac{r^{T(c+1)}}{r^{T(c)}} \frac{r^{(c+1)}}{r^{(c)}}$ — Update A-conjugacy
$d^{(c+1)} = \boldsymbol{r}^{T(c+1)} + z^{(c)}\boldsymbol{d}^{(c)}$ — Update search direction
$s^{(c)} = -\frac{\boldsymbol{r}^{T(c)}}{\boldsymbol{d}^{T(c)}\boldsymbol{A}} \frac{\boldsymbol{r}^{(c)}}{\boldsymbol{d}^{(c)}}$ — Optimal step length
$\hat{\alpha}^{(c+1)} = \hat{\boldsymbol{\alpha}}^{(c)} - s^{(c)}\boldsymbol{d}^{(c)}$ — Approximate solution
$r^{(c+1)} = \boldsymbol{r}^{(c)} - \boldsymbol{A}\hat{\boldsymbol{\alpha}}^{(c)}$ — Update residual
$c = c + 1$

The proposed model can be broken down into multiple stages. The first stage after receiving the data stream describes the definition of the train and test matrices using a moving (forward-rolling) window. In the second stage, the algorithm parameters are set and optimized using a grid search. The execution of TR–KRR is where the prediction matrix is obtained. Finally, the last stage is where the error calculations are made for further investigation and analysis. A detailed flowchart of the hybrid TR–KRR model is presented in Fig. 1.

The model was observed to perform best with a training set of 1 week. The testing range is equal to the forecasting range, which includes forecasts of 1, 12 and 24 hours ahead. Fig. 2 illustrates an example of a 1-week training and 1-day forecasting scenario. In this case, the algorithm trains on a contiguous block of 7 days and then forecasts 1 day ahead. Subsequently, the window is moved 1 day into the future and the process is repeated. Like many machine-learning algorithms, TR–KRR relies on specific parameters: sigma (σ) and lambda (λ). These can be optimized using a variety of techniques, the selection of which is contingent upon the application. When forecasting solar irradiance, the data are time series, thus the rolling window validation serves as the primary method for validation.

Moving-window cross validation is a suitable method for time-series data sets in which the data are validated by rolling over the data set without mixing the instances up [55]. After selecting the train and test sizes, the matrices are sent into the validation function, where a grid search is performed to optimize the parameter selection. A grid search is used to optimize the parameters, sigma (σ) and lambda (λ). In this study, the optimization process seeks to minimize the mean squared error (MSE) by systematically evaluating all combinations of σ spanning the range from 1 to 7 at increments of 0.1 and λ in the range from 0.1 to 1 at increments of 0.01. Model selection for TR–KRR is based on the lowest MSE. This process is repeated for every forecasting iteration. To maintain a consistent size of the training set, the approach is to remove the oldest data points, equivalent to the size of the moving window, as suggested by [56].

Because the algorithm is continuously optimized for forecasting, heavy computing is required. The program should run through the entire range of combinations of σ and λ, to optimize the parameters. Thus, a high-performance computing (HPC) cluster is used to effectively reduce the time requirement.

After the parameters are set, the next stage is to execute the TR–KRR function for the actual forecasts. The kernel function, $\boldsymbol{K}$, is first calculated using the $\mathbf{x}_{train}$ matrix and the sigma value:

$$K = k\left(\boldsymbol{x}_{train(i)}, \boldsymbol{x}_{train(j)}\right) = e^{-\frac{1}{2\sigma^2}||\mathbf{x}_{train(i)} - \mathbf{x}_{train(j)}||^2}. \quad (13)$$

This is followed by the calculation of the alpha vector using $\boldsymbol{K}$, $\boldsymbol{y}_{train}$ and λ as:

$$\alpha = (\boldsymbol{K} + \lambda \boldsymbol{I}_N)^{-1}\boldsymbol{y}_{train} \quad (14)$$

Then, the kernel is calculated again using $\boldsymbol{x}_{train}$, $\boldsymbol{x}_{test}$ and alpha $\boldsymbol{\alpha}$ as:

$$K = k\left(\boldsymbol{x}_{test}, \boldsymbol{x}_{train}\right) = e^{-\frac{1}{2\sigma^2}||\mathbf{x}_{train} - \mathbf{x}_{test}||^2}. \quad (15)$$

The forecast vector is then computed in dual form as:

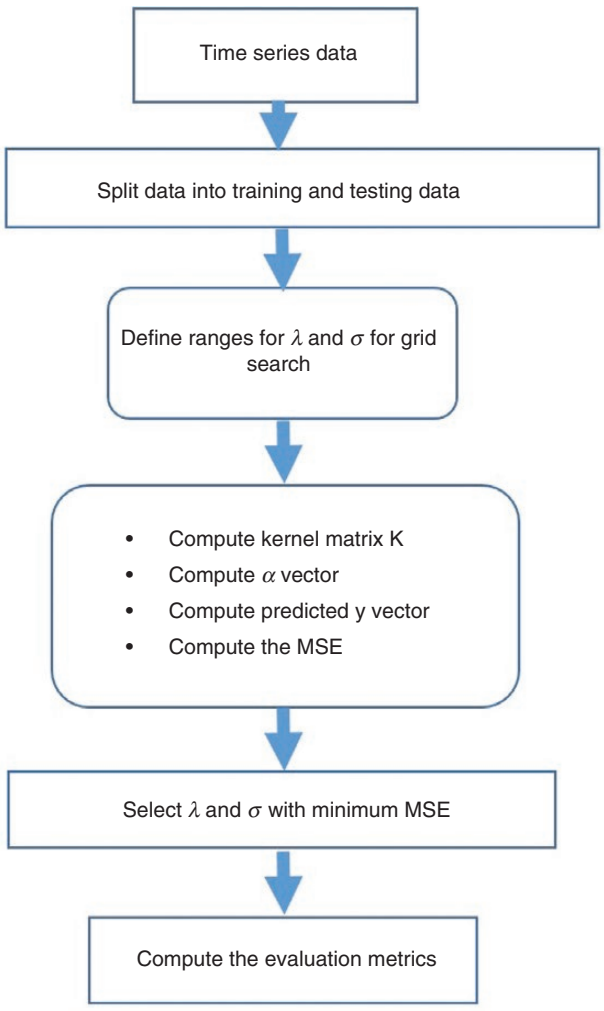



**Fig. 1:** Flowchart of the proposed hybrid TR–KRR model

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

$$\hat{y} = \boldsymbol{K}^{-1}\boldsymbol{\alpha}_{train}. \quad (16)$$

## 3 Experimental analysis

Each forecast horizon was tested over cloudy and sunny periods. For the Seattle data, the cloudy period is between 14 and 18 October, while the sunny period is from 7 to 11 July. On the other hand, the Medford data for cloudy and sunny periods are from 13 to 17 January and from 9 to 13 July, respectively. The choice of investigating these periods in particular is based on two extreme behaviours: extremely intermittent solar irradiance and uniform irradiance for cloudy and sunny periods, respectively.

### 3.1 Data sets and software use

As mentioned above, in this study, two data sets are used from the USA, specifically from Seattle and Medford, for the year 2018. The data sets were obtained from the National Solar Radiation Database [57, 58]. The hourly data sets comprise timestamp (hours), cloud type (discrete values 0–9), ambient temperature (degrees Celsius), relative humidity (%), wind speed (m/s) and solar global horizontal irradiance (GHI) (W/m$^2$). With the exception of timestamp and cloud type, each array of the data set was normalized to a mean of zero and a standard deviation of 1 before executing the forecasting models. The forecasting models were developed using MATLAB® and an HPC cluster server. The HPC cluster was used to process the large data stream and the online grid search, as discussed in the model framework.

### 3.2 Data preparation

To determine the optimal combination of meteorological input, both training and testing sizes were fixed while varying the model parameters. This is essential to achieve optimal results because proper selection of inputs determines the strength of an effective forecasting method. Data sets from different geographical locations will respond differently to the inputs. Furthermore, different forecast horizons also affect the accuracy of the results with different input variables. Therefore, a combination of multiple inputs was used to confirm the most suitable combination. All input combinations were tested for a train set size of 1 week and a test set size of 1 day for the entire year using the grid search optimization technique. The choice of weather parameters and training size was optimized based on the trial-and-error method to observe the duration required for the most accurate forecast [34, 59]. It was revealed that the combination that yields the highest accuracy is:

(i) short-term horizon: cloud type, ambient temperature, humidity, wind speed and solar irradiance;
(ii) medium-term horizon: timestamp, ambient temperature, humidity, wind speed and solar irradiance.

Solar irradiance, ambient temperature and wind speed showed significant correlations with the forecast output. However, cloud type showed a strong correlation for the short-term horizon and a very weak correlation for the medium-term horizon due to the nature of the cloud changes in motion. Because timestamp data are in units of hours, it introduces uncertainty in short-term forecasting due to fluctuations. However, it shows a strong correlation for medium-term horizons [59].

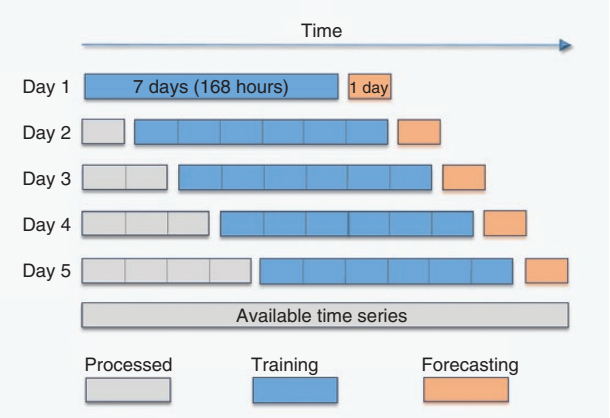



**Fig. 2:** Forecast using 1 week as training and 1 day ahead as testing with rolling window

### 3.3 Performance evaluation measures

To effectively evaluate and compare models, standardized error indices have been used to establish common ground. Various evaluation measures have been used in the literature for time-series forecasting. In this study, the following measures were used for performance evaluation:

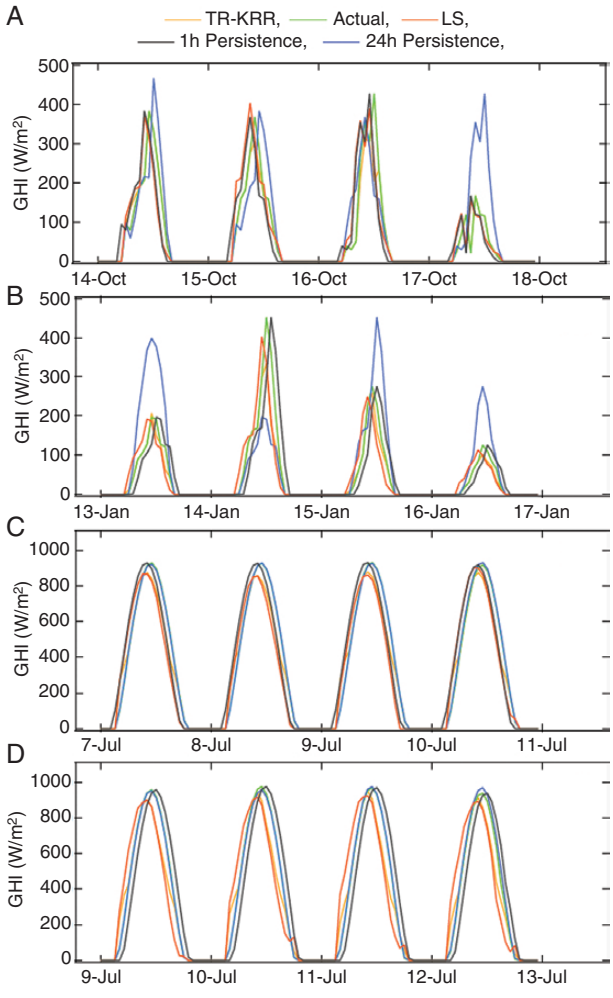



**Fig. 3:** Short-term forecast results (1-hour horizon) for (a) cloudy days in Seattle, (b) cloudy days in Medford, (c) sunny days in Seattle and (d) sunny days in Medford

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

**Table 1:** Evaluation metrics for short-term forecasting in Seattle and Medford

| Duration | Metric | Forecasting model | | | |
|---|---|---|---|---|---|
| | | Hybrid TR–KRR | 24-Hour persistence | LS | 1-Hour persistence |
| **Seattle (1-hour horizon)** | | | | | |
| **Cloudy (14–17 October)** | $R^2$ | 0.9485 | 0.6416 | 0.7578 | 0.7388 |
| | MAE | 0.0053 | 0.0342 | 0.271 | 0.0301 |
| | RMSE | 0.0261 | 0.0754 | 0.0558 | 0.0577 |
| **Sunny (7–10 July)** | $R^2$ | 0.9663 | 0.9997 | 0.9045 | 0.9175 |
| | MAE | 0.0448 | 0.0038 | 0.0840 | 0.0811 |
| | RMSE | 0.0778 | 0.0064 | 0.1199 | 0.1090 |
| **Medford (1-hour horizon)** | | | | | |
| **Cloudy (13–16 January)** | $R^2$ | 0.9329 | 0.5618 | 0.7917 | 0.7799 |
| | MAE | 0.0073 | 0.0343 | 0.0197 | 0.0205 |
| | RMSE | 0.0258 | 0.0732 | 0.0385 | 0.0408 |
| **Sunny (9–12 July)** | $R^2$ | 0.9569 | 0.9986 | 0.8473 | 0.9132 |
| | MAE | 0.0458 | 0.0069 | 0.0983 | 0.0782 |
| | RMSE | 0.0830 | 0.0139 | 0.1419 | 0.1077 |

**Table 2:** Percentage improvements in evaluation metrics for short-term forecasting in Seattle and Medford relative to the least-squares model

| Duration | Metric | Forecasting model percentage improvement (%) | | | |
|---|---|---|---|---|---|
| | | Hybrid TR–KRR | 24-Hour persistence | LS | 1-Hour persistence |
| **Seattle (1-hour horizon)** | | | | | |
| **Cloudy (14–17 October)** | $R^2$ | 25.2 | –15.3 | – | –2.5 |
| | MAE | 98.0 | 87.4 | – | 88.9 |
| | RMSE | 53.2 | –35.1 | – | –3.4 |
| **Sunny (7–10 July)** | $R^2$ | 6.8 | 10.5 | – | 1.4 |
| | MAE | 46.7 | 95.5 | – | 3.5 |
| | RMSE | 35.1 | 94.7 | – | 9.1 |
| **Medford (1-hour horizon)** | | | | | |
| **Cloudy (13–16 January)** | $R^2$ | 17.8 | –29.0 | – | –1.5 |
| | MAE | 62.9 | –74.1 | – | –4.1 |
| | RMSE | 33.0 | –90.1 | – | –6.0 |
| **Sunny (9–12 July)** | $R^2$ | 12.9 | 17.9 | – | 7.8 |
| | MAE | 53.4 | 93.0 | – | 20.4 |
| | RMSE | 41.5 | 90.2 | – | 24.1 |

(i) root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(G_{pred(i)} - G_{actual(i)}\right)^2}; \quad (17)$$

(ii) mean absolute error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|G_{pred(i)} - G_{actual(i)}\right|; \quad (18)$$

(iii) coefficient of determination ($R^2$):

$$R^2 = 1 - \frac{Var\left(G_{actual} - G_{pred}\right)}{Var\left(G_{pred}\right)}, \quad (19)$$

where $N$ is the total number of test samples, $G_{pred}$ is the predicted solar irradiance and $G_{actual}$ is the actual solar irradiance (test samples) [45].

## 3.4 Results of short-term forecasting

In our study on short-term solar forecasting, we focused on a 1-hour prediction horizon. The hybrid TR–KRR approach we introduce excels particularly during cloudy conditions, outperforming both the persistence and the LS models, as evidenced in Fig. 3. This graphical representation underscores the ability of the TR–KRR model to track inherent variability in actual solar irradiance data—a crucial feature for advancing clean energy forecasting. In this context, the TR–KRR model addresses the challenges that other models frequently encounter on cloudy days characterized by high solar irradiance fluctuations—conditions that often lead them to yield imprecise results. In contrast, on sunny days, the modified 24-hour persistence model delivers the most favourable

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

outcomes. However, the TR–KRR maintains a commendable performance, with an $R^2$ value ranging from 0.93 to 0.96, as detailed in Table 1. The proposed TR–KRR model consistently delivered predictions with low error values, specifically ranging from 0.0053 to 0.0458 for MAE and from 0.0258 to 0.083 for RMSE, respectively. Table 2 lists the percentage improvements in the evaluation metrics ($R^2$, MAE and RMSE) for all models compared with the LS model. The results show that the TR–KRR model outperformed all other models for cloudy-day test samples with percentage improvements, relative to the LS model, of up to 25%, 98% and 53% for $R^2$, MAE and RMSE, respectively. For sunny-day samples, the TR–KRR exhibited good performance, with improvements of up to 13%, 53% and 41% in $R^2$, MAE and RMSE, respectively, relative to the LS model; however, it did not surpass the 24-hour persistence model.

## 3.5 Results of the medium-term forecast

The same process was repeated for medium-term forecasts for 12-hour and 24-hour horizons; however, the input variables to the model were: timestamp (in hours), ambient temperature, humidity, wind speed and solar irradiance. TR–KRR showed good performance with an average $R^2$ of 0.915 and 0.996 for cloudy and sunny days, respectively.

Based on the results generated, the proposed model provides an accurate solution for the forecasting of solar irradiance in winter and cloudy days. The model shows better forecasting performance with smaller horizons due to the more frequent updates in the tuning parameters. For medium-term forecasting, the model proved to be a suitable tool for all weather conditions. The results are presented in Figs 4 and 5. These figures show a great tracking feature for the proposed TR–KRR model for the fluctuation in the actual solar irradiance data for cloudy weather, which outperformed the other two models (LS and 24-hour persistence). In general, the results show the superiority of the proposed hybrid TR–KRR model over traditional forecasters, as shown by the RMSE, MAE and $R^2$ values presented in Table 3. The proposed TR–KRR model produced predictions with $R^2$ ranging from 0.93 to 0.99 for all weather conditions. Furthermore, the model also produced relatively low error values ranging from 0.01 to 0.0302 and from 0.0177 to 0.0475 for MAE and RMSE, respectively. The percentage improvements in the evaluation metrics ($R^2$, MAE and RMSE) for all models relative to the LS model are presented in Table 4. This table reveals that the TR–KRR model outperformed all other models for cloudy-day test samples with percentage improvements, relative to the LS model, of up to 97%, 66% and 65% for $R^2$, MAE and RMSE, respectively. For sunny-day test samples, the TR–KRR produced good performance and percentage improvement, relative to the LS model, of up to 28%, 91% and 90% for $R^2$, MAE and RMSE, respectively, but once again not as good as the 24-hour persistence

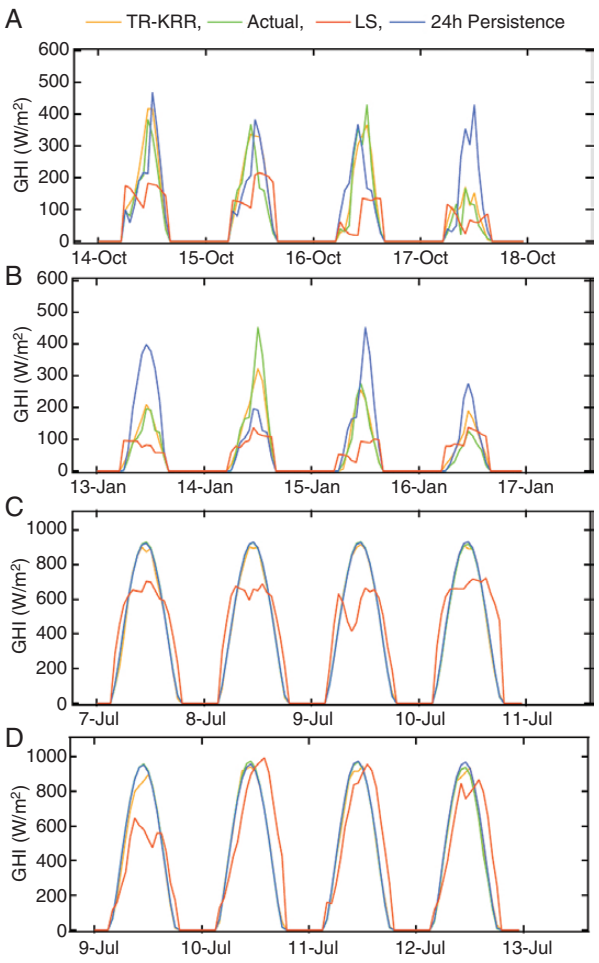



**Fig. 4:** Medium-term forecast results (12-hour horizon) for (a) cloudy days in Seattle, (b) cloudy days in Medford, (c) sunny days in Seattle and (d) sunny days in Medford

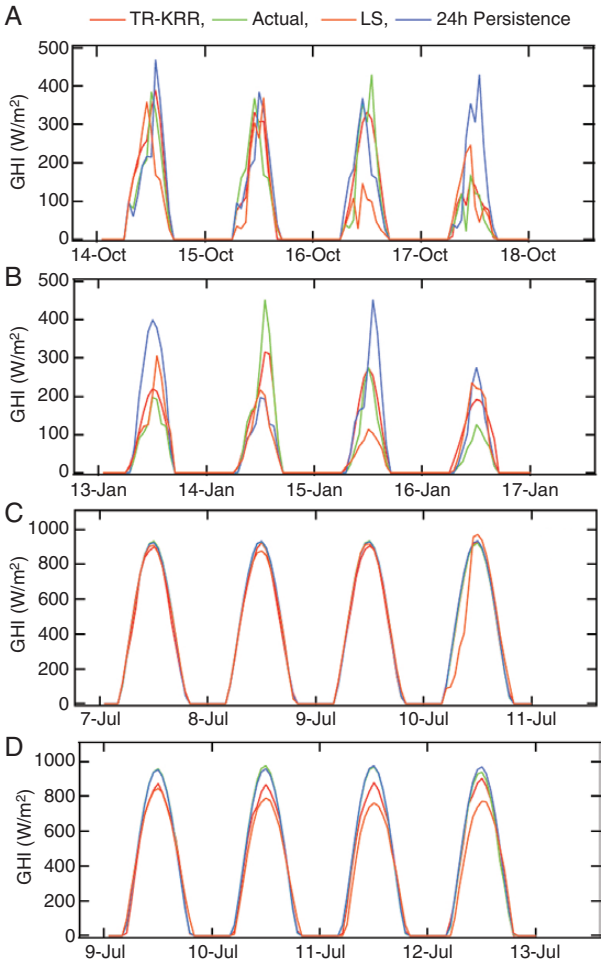



**Fig. 5:** Medium-term forecast results (24-hour horizon) for (a) cloudy days in Seattle, (b) cloudy days in Medford, (c) sunny days in Seattle and (d) sunny days in Medford

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

**Table 3:** Evaluation metrics for medium-term forecasting in Seattle and Medford

| Duration | Metric | Forecasting model | | |
|---|---|---|---|---|
| | | Hybrid TR–KRR | 24-Hour persistence | LS |
| **Seattle (12-hour horizon)** | | | | |
| **Cloudy (14–17 October)** | $R^2$ | 0.9349 | 0.6416 | 0.4736 |
| | MAE | 0.0150 | 0.0342 | 0.0379 |
| | RMSE | 0.0324 | 0.0754 | 0.0804 |
| **Sunny (7–10 July)** | $R^2$ | 0.9982 | 0.9997 | 0.7929 |
| | MAE | 0.0106 | 0.0038 | 0.1200 |
| | RMSE | 0.0177 | 0.0064 | 0.1732 |
| **Seattle (24-hour horizon)** | | | | |
| **Cloudy (14–17 October)** | $R^2$ | 0.9127 | 0.6416 | 0.5483 |
| | MAE | 0.0158 | 0.0342 | 0.0343 |
| | RMSE | 0.0331 | 0.0754 | 0.0747 |
| **Sunny (7–10 July)** | $R^2$ | 0.9979 | 0.9997 | 0.9723 |
| | MAE | 0.0146 | 0.0038 | 0.0263 |
| | RMSE | 0.0234 | 0.0064 | 0.0636 |
| **Medford (12-hour horizon)** | | | | |
| **Cloudy (13–16 January)** | $R^2$ | 0.9356 | 0.5618 | 0.4978 |
| | MAE | 0.0100 | 0.0343 | 0.0291 |
| | RMSE | 0.0217 | 0.0732 | 0.0625 |
| **Sunny (9–12 July)** | $R^2$ | 0.9959 | 0.9986 | 0.7758 |
| | MAE | 0.0150 | 0.0069 | 0.1165 |
| | RMSE | 0.0261 | 0.0139 | 0.1717 |
| **Medford (24-hour horizon)** | | | | |
| **Cloudy (13–16 January)** | $R^2$ | 0.8778 | 0.5618 | 0.5801 |
| | MAE | 0.0157 | 0.0343 | 0.0243 |
| | RMSE | 0.0316 | 0.0732 | 0.0550 |
| **Sunny (9–12 July)** | $R^2$ | 0.9926 | 0.9986 | 0.9753 |
| | MAE | 0.0302 | 0.0069 | 0.0616 |
| | RMSE | 0.0475 | 0.0139 | 0.0901 |

model, especially for 24-hour prediction horizon. However, the 24-hour persistence model performed the worst for cloudy samples.

## 4 Conclusion and discussion

The disadvantage of PV power generation is the uncertainty and instability of the power output. Therefore, this paper presents a novel hybrid TR–KRR framework for forecasting solar irradiance, which is directly correlated with solar power generation. The algorithm used was selected for its computational speed and accuracy with highly fluctuating data to overcome the limitations faced by some forecasting models to predict solar irradiance on cloudy days. In this study, two data sets from Medford and Seattle were used. The proposed model uses a moving window for training, testing and validation. The tuning parameters, sigma and lambda, are updated for each data stream to obtain the most accurate model. The analysis was performed considering cloudy and sunny days. Three evaluation metrics were used to evaluate the performance of the model: RMSE, MAE and $R^2$. The results show that the hybrid TR–KRR model overcomes the limitation of forecasting on cloudy days faced by the modified persistence, one-step persistence and LS models, where the $R^2$ values increased from an average of 0.6274 to an average of 0.9237 for cloudy days. The model also performed with high accuracy for sunny days with an $R^2$ average of 0.9766 and 0.9961 for short- and medium-term horizons, respectively.

However, the proposed hybrid TR–KRR model still has some limitations and room for improvement that could be studied in future work:

(i) Although the proposed model shows improved computational time compared with kernel-based algorithms, it still requires a considerable amount of time compared with simpler algorithms such as the LS and persistence models.
(ii) The model has the limitation that it is not applicable to long-term forecasting due to its statistical nature. NWP methods are commonly used for long-term forecasting.
(iii) The data modification process currently requires to be performed separately before running the model. This could be improved by adding an interpolation section within the algorithm to estimate missing data points and make direct online forecasts. Future work would focus on comparing TR–KRR with various novel approaches, such as LSTM or hybrid CNN–LSTM, to further validate the efficacy and efficiency of the algorithm.

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

**Table 4:** Percentage improvements of evaluation metrics for medium-term forecasting in Seattle and Medford relative to the least-squares model

| Duration | Metric | Forecasting model | | |
|---|---|---|---|---|
| | | Hybrid TR–KRR | 24-Hour persistence | LS |
| **Seattle (12-hour horizon)** | | | | |
| **Cloudy (14–17 October)** | $R^2$ | 97.4 | 35.5 | – |
| | MAE | 60.4 | 9.8 | – |
| | RMSE | 59.7 | 6.2 | – |
| **Sunny (7–10 July)** | $R^2$ | 25.9 | 26.1 | – |
| | MAE | 91.2 | 96.8 | – |
| | RMSE | 89.8 | 96.3 | – |
| **Seattle (24-hour horizon)** | | | | |
| **Cloudy (14–17 October)** | $R^2$ | 66.5 | 17.0 | – |
| | MAE | 53.9 | 0.3 | – |
| | RMSE | 55.7 | –0.9 | – |
| **Sunny (7–10 July)** | $R^2$ | 2.6 | 2.8 | – |
| | MAE | 44.5 | 85.6 | – |
| | RMSE | 63.2 | 89.9 | – |
| **Medford (12-hour horizon)** | | | | |
| **Cloudy (13–16 January)** | $R^2$ | 87.9 | 12.9 | – |
| | MAE | 65.6 | –17.9 | – |
| | RMSE | 65.3 | –17.1 | – |
| **Sunny (9–12 July)** | $R^2$ | 28.4 | 28.7 | – |
| | MAE | 87.1 | 94.1 | – |
| | RMSE | 84.8 | 91.9 | – |
| **Medford (24-hour horizon)** | | | | |
| **Cloudy (13–16 January)** | $R^2$ | 51.3 | –3.2 | – |
| | MAE | 35.4 | –41.2 | – |
| | RMSE | 42.5 | –33.1 | – |
| **Sunny (9–12 July)** | $R^2$ | 1.8 | 2.4 | – |
| | MAE | 51.0 | 88.8 | – |
| | RMSE | 47.3 | 84.6 | – |

## Acknowledgements

The authors are thankful to Tanujit Chakraborty for valuable discussions and inputs.

## Conflict of interest statement

The authors declare no conflict of interest.

## Funding

This work was supported by the Khalifa University of Science and Technology under Award No. RC2 DSO and the Advanced Power and Energy Center.

## Data Availability

The data underlying this article are available in Mendeley Repository at https://dx.doi.org/10.17632/v5h6f8wnkc.1.

## References

[1] Iheanetu KJ. Solar photovoltaic power forecasting: a review. *Sustainability*, 2022, 14:17005. https://doi.org/10.3390/su142417005.

[2] IRENA. *World Energy Transitions Outlook: 1.5° C Pathway*. Abu Dhabi, UAE: International Renewable Energy Agency, 2021.

[3] Teleke S, Baran ME, Bhattacharya S, *et al*. Rule-based control of battery energy storage for dispatching intermittent renewable sources. *IEEE Trans Sustainable Energy*, 2010, 1:117–124. https://doi.org/10.1109/TSTE.2010.2061880.

[4] Chathurika DAH, Perumpuliarachchi LY, Abeyratne SG, *et al*. Managing spinning reserves for power system stability under increased renewable energy penetration. In: *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, Rupnagar, India, 26–28 November 2020, 127–132.

[5] Eltawil MA, Zhao Z. Grid-connected photovoltaic power systems: technical and potential problems—a review. *Renew Sustain Energy Rev*, 2010, 14:112–129. https://doi.org/10.1016/j.rser.2009.07.015.

[6] Smith O, Cattell O, Farcot E, *et al*. The effect of renewable energy incorporation on power grid stability and resilience. *Sci Adv*, 2022, 8:eabj6734.

[7] Varma RK, Salama M. Large-scale photovoltaic solar power integration in transmission and distribution networks. In: *2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, 24–28 July 2011, 1–4. https://doi.org/10.1109/PES.2011.6039860.

[8] Alshahrani A, Omer S, Su Y, *et al*. The technical challenges facing the integration of small-scale and large-scale PV systems into the grid: a critical review. *Electronics*, 2019, 8:1443.

[9] Anwar M, El Moursi M, Xiao W. Novel power smoothing and generation scheduling strategies for a hybrid wind and marine current turbine system. *IEEE Trans Power Syst*, 2016, 32:1315–1326. https://doi.org/10.1109/TPWRS.2016.2591723.

[10] Mohandes B, Wahbah M, El Moursi MS, *et al*. Renewable energy management system: Optimum design and hourly dispatch. *IEEE Trans Sustainable Energy*, 2021, 12:1615–1628.

[11] Pourmousavi SA, Cifala AS, Nehrir MH. Impact of high penetration of PV generation on frequency and voltage in a distribution feeder. In: *2012 North American Power Symposium (NAPS)*, Champaign, IL, USA, 9–11 September 2012, 1–8. https://doi.org/10.1109/NAPS.2012.6336320.

[12] Li X, Hui D, Lai X. Battery energy storage station (BESS)-based smoothing control of photovoltaic (PV) and wind power generation fluctuations. *IEEE Trans Sustainable Energy*, 2013, 4:464–473. https://doi.org/10.1109/TSTE.2013.2247428.

[13] Mahmud K, Azam S, Karim A, *et al*. Machine learning based PV power generation forecasting in Alice Springs. *IEEE Access*, 2021, 9:46117–46128.

[14] Niccolai A, Dolara A, Ogliari E. Hybrid PV power forecasting methods: a comparison of different approaches. *Energies*, 2021, 14:451. https://doi.org/10.3390/en14020451.

[15] Jaber M, Hamid ASA, Sopian K, *et al*. Prediction model for the performance of different PV modules using artificial neural networks. *Applied Sciences*, 2022, 12:3349. https://doi.org/10.3390/app12073349.

[16] Chong KK, Khlyabich PP, Hong KJ, *et al*. Comprehensive method for analyzing the power conversion efficiency of organic solar cells under different spectral irradiances considering both photonic and electrical characteristics. *Appl Energy*, 2016, 180:516–523. https://doi.org/10.1016/j.apenergy.2016.08.002.

[17] Reikard G. Predicting solar radiation at high resolutions: a comparison of time series forecasts. *Sol Energy*, 2009, 83:342–349. https://doi.org/10.1016/j.solener.2008.08.007.

[18] Alsharif MH, Younes MK, Kim J. Time series ARIMA model for prediction of daily and monthly average global solar radiation: the case study of Seoul, South Korea. *Symmetry*, 2019, 11:240.

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

[19] Mellit A, Sağlam S, Kalogirou SA. Artificial neural network-based model for estimating the produced power of a photovoltaic module. *Renew Energy*, 2013, 60:71–78. https://doi.org/10.1016/j.renene.2013.04.011.

[20] El Moursi MS, Al Hinai A, Ssekulima EB, *et al*. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review. *IET Renew Power Gener*, 2016, 10:885–989. https://doi.org/10.1049/iet-rpg.2015.0477.

[21] Shi J, Lee WJ, Liu Y, *et al*. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Trans Ind Appl*, 2012, 48:1064–1069. https://doi.org/10.1109/TIA.2012.2190816.

[22] Bae KY, Jang HS, Sung DK. Hourly solar irradiance prediction based on support vector machine and its error analysis. *IEEE Trans Power Syst*, 2016, 32:1–1. https://doi.org/10.1109/TPWRS.2016.2569608.

[23] Deng F, Su G, Liu C, *et al*. Prediction of solar radiation resources in China using the LS-SVM algorithms. In: *2010 The 2nd International Conference on Computer and Automation Engineering, ICCAE 2010*, Singapore, 26–28 February 2010, 31–35. https://doi.org/10.1109/ICCAE.2010.5451535.

[24] Ekici BB. A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems. *Measurement*. 2014, 50:255–262. https://doi.org/10.1016/j.measurement.2014.01.010.

[25] Jang HS, Bae KY, Park HS, *et al*. Solar power prediction based on satellite images and support vector machine. *IEEE Trans Sustainable Energy*, 2016, 7:1255–1263. https://doi.org/10.1109/TSTE.2016.2535466.

[26] Youssef Ali Amer A. Global-local least-squares support vector machine (GLocal-LS-SVM). *PLoS One*, 2023, 18:e0285131. https://doi.org/10.1371/journal.pone.0285131.

[27] Castillo-Rojas W, Bekios-Calfa J, Hernández C. Daily prediction model of photovoltaic power generation using a hybrid architecture of recurrent neural networks and shallow neural networks. *Int J Photoenergy*, 2023, 2023:2592405. https://doi.org/10.1155/2023/2592405.

[28] Castillo-Rojas W, Medina Quispe FM, Hernández C. Photovoltaic energy forecast using weather data through a hybrid model of recurrent and shallow neural networks. *Energies*, 2023, 16:5093. https://doi.org/10.3390/en16135093.

[29] Xu D, Shao H, Deng X, *et al*. The hidden-layers topology analysis of deep learning models in survey for forecasting and generation of the wind power and photovoltaic energy. *CMES—Computer Modeling in Engineering and Sciences*, 2022, 131:567–597. https://doi.org/10.32604/cmes.2022.019245.

[30] Li H, Ren Z, Xu Y, *et al*. A multi-data driven hybrid learning method for weekly photovoltaic power scenario forecast. *IEEE Trans Sustainable Energy*, 2021, 13:91–100.

[31] Carrera B, Sim MK, Jung JY. PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data. *IET Renew Power Gener*, 2020, 14:2192–2201.

[32] Zang H, Liu L, Sun L, *et al*. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew Energy*, 2020, 160:26–41.

[33] Benmouiza K, Cheknane A. Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Convers Manage*, 2013, 75:561–569. https://doi.org/10.1016/j.enconman.2013.07.003.

[34] Ahmed R, Sreeram V, Mishra Y, *et al*. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. *Renew Sustain Energy Rev*, 2020, 124:109792. https://doi.org/10.1016/j.rser.2020.109792.

[35] Sobri S, Koohi-Kamali S, Rahim NA. Solar photovoltaic generation forecasting methods: a review. *Energy Convers Manage*, 2018, 156:459–497. https://doi.org/10.1016/j.enconman.2017.11.019.

[36] Yang HT, Huang CM, Huang YC, *et al*. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. *IEEE Trans Sustainable Energy*, 2014, 5:917–926. https://doi.org/10.1109/TSTE.2014.2313600.

[37] Mellit A, Massi Pavan A, Ogliari E, *et al*. Advanced methods for photovoltaic output power forecasting: a review. *Applied Sciences*, 2020, 10:487. https://doi.org/10.3390/app10020487.

[38] Maalouf M, Homouz D. Kernel ridge regression using truncated Newton method. *Knowledge-Based Systems*, 2014, 71:339–344. https://doi.org/10.1016/j.knosys.2014.08.012.

[39] Maalouf M, Homouz D, Abutayeh M. Accurate prediction of preheat temperature in solar flash desalination systems using Kernel Ridge regression. *J Energy Eng*, 2016, 142:E4015017. https://doi.org/10.1061/(asce)ey.1943-7897.0000333.

[40] Maalouf M, Barsoum Z. Failure strength prediction of aluminum spot-welded joints using kernel ridge regression. *The International Journal of Advanced Manufacturing Technology*, 2017, 91:3717–3725.

[41] Gaganis V, Homouz D, Maalouf M, *et al*. An efficient method to predict compressibility factor of natural gas streams. *Energies*, 2019, 12:2577.

[42] Alalami MA, Maalouf M, EL-Fouly THM. Wind speed forecasting using kernel ridge regression with different time horizons. In: Valenzuela O, Rojas F, Herrera LJ, Pomares H, Rojas I (eds). *Theory and Applications of Time Series Analysis. ITISE 2019. Contributions to Statistics*. Cham, Switzerland: Springer, 2020, 191–203. https://doi.org/10.1007/978-3-030-56219-9_13.

[43] Abdulla H, Maalouf M, Barsoum I, *et al*. Truncated Newton kernel ridge regression for prediction of porosity in additive manufactured SS316L. *Applied Sciences*, 2022, 12:4252.

[44] Al Nuaimi H, Abdelmagid M, Bouabid A, *et al*. Classification of WatSan Technologies using machine learning techniques. *Water*, 2023, 15:2829.

[45] Khatib T, Mohamed A, Sopian K. A review of solar energy modeling techniques. *Renew Sustain Energy Rev*, 2012, 16:2864–2869. https://doi.org/10.1016/j.rser.2012.01.064.

[46] Chow CW, Urquhart B, Lave M, *et al*. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol Energy*, 2011, 85:2881–2893. https://doi.org/10.1016/j.solener.2011.08.025.

[47] Gwandu BAL, Creasey DJ. Humidity: a factor in the appropriate positioning of a photovoltaic power station. *Renew Energy*, 1995, 6:313–316. https://doi.org/10.1016/0960-1481(95)00073-S.

[48] Schwingshackl C, Petitta M, Wagner JE, *et al*. Wind effect on PV module temperature: analysis of different techniques for an accurate estimation. *Energy Procedia*, 2013, 40:77–86. https://doi.org/10.1016/j.egypro.2013.08.010.

[49] Goossens D, Van Kerschaever E. Aeolian dust deposition on photovoltaic solar cells: the effects of wind velocity and airborne dust concentration on cell performance. *Sol Energy*, 1999, 66:277–289. https://doi.org/10.1016/S0038-092X(99)00028-6.

[50] Kimber A, Mitchell L, Nogradi S, *et al*. The effect of soiling on large grid-connected photovoltaic systems in California and the Southwest region of the United States. In: *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*, Waikoloa, HI, USA, 7–12 May 2006, 2391–2395. https://doi.org/10.1109/WCPEC.2006.279690.

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024

[51] Guo B, Javed W, Figgis BW, *et al*. Effect of dust and weather conditions on photovoltaic performance in Doha, Qatar. In: *2015 First Workshop on Smart Grid and Renewable Energy (SGRE)*, Doha, Qatar 22–23 March 2015, 1–5. https://doi.org/10.1109/SGRE.2015.7208718.

[52] Zhang J, Hodge BM, Lu S, *et al*. Baseline and target values for regional and point PV power forecasts: toward improved solar forecasting. *Sol Energy*, 2015, 122:804–819. https://doi.org/10.1016/j.solener.2015.09.047.

[53] Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 2021.

[54] Shawe TJ, Cristianini N. *Kernel Methods for Pattern Analysis*. Cambridge, UK: Cambridge University Press, 2004.

[55] Kasemsumran S, Du YP, Li BY, *et al*. Moving window cross validation: a new cross validation method for the selection of a rational number of components in a partial least squares calibration model. *The Analyst*, 2006, 131:529–537. https://doi.org/10.1039/b515637h.

[56] Agarwal S, Vijaya Saradhi V, Karnick H. Kernel-based online machine learning and support vector reduction. *Neurocomputing*, 2008, 71:1230–1237. https://doi.org/10.1016/j.neucom.2007.11.023.

[57] National Renewable Energy Laboratory. NSRDB: National Solar Radiation Database. https://nsrdb.nrel.gov/data-viewer (26 September 2023, date last accessed).

[58] Sengupta M, Xie Y, Lopez A, *et al*. The national solar radiation data base (NSRDB). *Renew Sustain Energy Rev*, 2018, 89:51–60.

[59] Al-Dahidi S, Ayadi O, Alrbai M, *et al*. Ensemble approach of optimized artificial neural networks for solar photovoltaic power prediction. *IEEE Access*, 2019, 7:81741–81758. https://doi.org/10.1109/ACCESS.2019.2923905.

Downloaded from https://academic.oup.com/ce/article/8/1/100/7515046 by guest on 02 April 2024