We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Abstract

As intermittent renewable energy sources such as wind and solar proliferate, the power systems operation and planning become more complicated due to increased uncertainties and variabilities. Accurate forecasting of these sources facilitates planning and operating the electric grid to integrate wind/solar power more reliably and efficiently. The neural network learning process can be disrupted by anomalies of wind/solar time-series data, which results in less accurate forecasting. By processing and analyzing wind/solar time-series data, machine learning and pattern recognition methods such as data clustering and classification can significantly enhance the forecast accuracy. This chapter reviews the various machine learning and pattern recognition methods proposed in the literature for time-series forecasting of solar radiation.

Keywords: forecasting, pattern recognition, renewable energy, solar, time series

1. Introduction

Pattern recognition is the analysis of data to detect their patterns and arrangements [1]. Pattern recognition provides a signal identification technique that is popularly applied in time-series forecasting. Pattern recognition is particularly essential for highly fluctuating and volatile time series such as solar radiation with irregular patterns. The chaotic nature of solar radiation time-series data disrupts the neural network learning process and imposes high errors on forecasting. By better detecting anomalous data points (outliers) and irregular patterns, pattern recognition and machine learning-based techniques characterize the training data more accurately and provide better learning results for neural networks [2]. In this chapter, an overview of the most commonly used pattern recognition and machine learning methods for solar radiation forecasting is presented.

2. Pattern recognition methods in solar forecasting

Several pattern recognition methods have been used to identify patterns and provide a pattern-based prediction technique for solar irradiance [4, 6–8, 10, 13, 14, 16–21, 23–25]. A brief description is provided for each method in the following sections.
2.1 Combination of SOM, SVR, and PSO

Self-organizing map (SOM) is an unsupervised learning approach for data classification and clustering [2]. Support vector machines (SVMs) provide supervised learning for data classification and regression analysis [3]. Dong et al. [4] propose a hybrid forecasting method combining SOM, support vector regression (SVR), and particle swarm optimization (PSO) methods. The SOM is used to divide the input space into disjointed regions with different characteristic information on the correlation between the input and the output. Then, the SVR is applied to model each disjointed region to identify the characteristic correlation. Finally, the PSO is applied to reduce the performance volatility of SVR with different parameters. The hybrid SOM-SVR-PSO method is rigorously tested and compared with several well-known time series forecasting algorithms. The comparison demonstrates higher accuracy of the proposed forecasting method.

2.2 Fuzzy c-means-based method

Fuzzy c-means (FCM) clustering is a well-known data clustering approach that allows each data element to belong to multiple clusters with varying degrees of membership [5]. Boata and Gravila [6] propose a novel method to forecast the stochastic component of the solar irradiation by the sky condition. By adopting the fuzzy inference system (FIS) principles, the proposed model forecasts daily clearness index. The FIS uses fuzzy logic to map any given input (features in the case of fuzzy clustering) to an output (clusters in the case of fuzzy clustering). In the proposed method, fuzzy c-means clustering is used to establish the membership functions (MFs) from the input variable's attributes. The MFs constitute the building blocks of the fuzzy set theory in fuzzy logic and characterize the fuzziness in a fuzzy set. The proposed method is evaluated and concluded to produce highly accurate results for practical applications.

2.3 Combination of fuzzy logic and neural networks

Fuzzy logic provides a powerful pattern recognition tool mainly because of its capability to characterize imperfect or noisy information and to measure data resemblance for clustering [7]. Chen et al. [8] proposes a new technique for solar radiation forecasting by combining the fuzzy logic and neural networks, to achieve a good accuracy at different conditions. In this method, the future sky conditions and temperature information are clustered to different fuzzy sets using a fuzzy logic-based clustering algorithm. The results demonstrate that the hybrid of fuzzy logic and neural network enhances the forecast accuracy for different sky and temperature conditions.

2.4 Spectral clustering

A new time-series clustering technique is proposed in Ref. [9] to reduce the computational complexity of smart grid optimization problems. Spectral clustering is used in this chapter to cluster different profiles for N days, where each day's profile is a time series over T slots. The data are clustered with respect to three different features including time attribute, frequency attribute, and weighted average of time-frequency attributes. The results show that clustering time-series data to provide two or more sub time series can improve the optimization performance when compared to using just a single typical time series.
2.5 Combination of k-means and NAR

K-means algorithm is a partition-based data clustering approach, which iteratively relocates data points among clusters for clustering optimization [10]. Nonlinear autoregressive (NAR) forecasting is based on the nonlinear relation between the past outputs and the predicted outputs, which can be defined by a high-order diversity measure [11].

Benmouiza [12] proposes a hybrid of k-means clustering and NAR neural networks for short-term solar radiation forecasting. In this method, k-means algorithm is applied to extract useful information from the input data. By clustering the data, k-means recognizes patterns of the input space, which provides better training for the neural network and improves the forecasting results. The proposed hybrid method first uses the mutual information (MI) approach to identify the suitable time delay and reconstitutes the phase space of solar radiation time series. The false nearest neighbors (FNN) algorithm is then used to determine the minimum embedding dimension for reconstructing the nonlinear dynamics from a time series. The next phase involves the k-means clustering to cluster the input patterns into k groups with similar characteristics. The silhouette method is applied to select the most suitable number of clusters. A NAR neural network is then trained for each cluster as a sub-predictor for the corresponding subset of the input pattern. Finally, another NAR network is applied as a global predictor for the solar radiation time-series data. Taking advantage of both k-means and NAR, the proposed forecasting method provides satisfactory results.

2.6 K-fold-based method

K-fold cross validation approach is a resampling procedure used to evaluate machine learning models over a sample data. The bootstrap is another resampling technique that generates multiple datasets by sampling from the original single dataset [13]. A combination of bootstrap sampling, k-fold cross validation, mutual information, and stationary time-series processing with clear sky model is applied in Ref. [13] for solar radiation forecasting. In this method, the cumulative distribution function (CDF) and matching quintile estimation (MQE) are used to determine the prediction interval. Based on the evaluation results, the proposed method demonstrates higher accuracy for forecasting horizons varying between 1 and 6 hours.

2.7 Combination of ANN with fuzzy logic pre-processing

Sivaneasan [14] develops an improved solar forecasting algorithm by combining fuzzy logic pre-processing and artificial neural network (ANN) model. The fuzzy pre-processing is used to calculate the correlation between temperature, cloud cover, wind speed, and direction with irradiance values. In this method, a three-layer feed forward with back-propagation model is applied as the neural network training algorithm. An error correction factor is proposed to reduce forecast errors by combining the error from the previous 5-min estimated output in the input layer. The evaluation results demonstrate that the error correction factor combined with the ANN approach improves solar irradiance forecasting accuracy due to its adaptive error correction capability. The forecasting accuracy of the proposed method is compared with other ANN forecasting algorithms. The results show that the proposed method has a better performance.
2.8 Semi-empirical methods

Ångström-Prescott (AP) is a model used to estimate the global solar irradiation as a function of solar radiation at the earth’s surface and daily sunshine duration [15]. Akarslan et al. [16] presents five semi-empirical solar radiation forecasting models based on the AP approach. These models utilize solar irradiance historical data along with the extra-terrestrial irradiance and the clearness index for hourly forecasting of solar radiation. Models 1–2 use historical samples of solar radiation and clearness index, whereas models 3–5 utilize the extra-terrestrial irradiance in addition to historical values of solar radiation and clearness index. The evaluation results show that including the extra-terrestrial irradiance data in the model enhances the forecasting accuracy.

2.9 Combination of k-means, DT, and SVM-C

A hybrid framework is proposed in Ref. [17] to model and forecast hourly global solar radiation data. This approach includes two different phases and uses data mining techniques in each step. K-means clustering technique is used in the first phase to identify the type of days. In the second phase, the decision trees (DT), artificial neural networks, support vector machine regression (SVM-R), and support vector machine classification algorithms (SVM-C) are combined with regression algorithms to obtain the daily clearness index and the meteorological parameters to forecast hourly global solar radiation. The results of the evaluation indicate that the proposed method satisfies the desired accuracy.

2.10 Machine learning-based methods

Five machine learning models, including adaptive forward-backward greedy algorithm (Fo-Ba), leap-Forward, forecast and variable selection by spike and slab regression (spikeslab), bagging wrapper for multivariate adaptive regression splines (Bagged MARS) using generalized cross validation (gCV) pruning, Cubist, and bagEarthGCV are presented in Ref. [18] for the solar irradiance prediction. The Fo-Ba algorithm is an efficient sparse learning and feature selection method with applications in optimization problems. The spikeslab model is a prediction and variable selection approach based on spike and slab regression. Bagged MARS approach computes an Earth model for each bootstrap sample of the original training set. In addition, the generalized cross validation approach is used for regularizing parameter selection in geophysical diffraction tomography (GDT). The above-mentioned models are evaluated and compared for different forecasting horizons from 1 h ahead to 48 h ahead. The evaluation results show that the spikeslab and Cubist models have more accurate and consistent performance for different forecasting horizons.

2.11 Combination of SOM and wavelet neural networks

Wavelet neural network is the combination of wavelet analysis and neural networks to address the over fitting issue of single network models. Reference [19] develops a novel short-term PV generation forecasting method by combining SOM algorithm and wavelet neural networks. The SOM method is applied for clustering the weather data and recognizing the future weather type. The wavelet neural network is utilized to build the prediction models for each cluster sample. The proposed method is proven to culminate in highly accurate forecasting.
2.12 Adaptive neuro-fuzzy system

Mohammadi et al. [20] propose an adaptive neuron-fuzzy inference system (ANFIS)-based model for daily solar radiation forecasting. The ANFIS method is a hybrid artificial intelligence method combining learning and generalization ability of neural network with characteristics of fuzzy inference system. The results demonstrate that the ANFIS model is an effective approach in the field of solar radiation forecasting.

2.13 SVR-based method

Support vector regression (SVR) is a variant of SVM that is used to compute the linear regression function of a nonlinear mapping [3]. SVR methodology is used in Ref. [21] for horizontal global solar radiation forecasting based on the daily hours of sunshine and daylight. The obtained results show that this method forecasts the monthly average daily global solar radiation with higher accuracy over the PSO-based models.

2.14 Combination of SVM, WT, ANN, and GA methods

Wavelet transform (WT) is a time-series analysis tool that decomposes a signal into a representation demonstrating time-series details and trends as a function of time [22]. Mohammadi et al. [23] uses a hybrid of support vector machine and wavelet transform algorithm (SVM-WT) for extracting features of solar radiation data. In addition, the artificial neural network (ANN) and genetic algorithm (GA) are combined to forecast daily global solar radiation. The results show that the SVM-WT method improves the forecasting accuracy by recognizing the patterns and providing better training for the neural network.

2.15 Clustering-based multi-model method

Wu and Chan [24] propose a novel multi-model framework for solar radiation forecasting. The framework is based on the assumption that there are several patterns in the stochastic component of solar radiation series. The time series data are first classified into multiple subsequences. The k-means algorithm is then applied to group the subsequence into different clusters. Finally, the time-delay neural network (TDNN) is trained to model a specific pattern in each cluster. The pattern corresponding to the current time is then determined for the forecasting purpose. This process is followed by selecting the appropriate trained TDNN model. The comparison analysis of the proposed forecasting method with autoregressive-moving-average (ARMA) model shows that the proposed model provides superior performance.

2.16 Data-driven method

Aerosol Optical Depth (AOD) and the Angstrom Exponent data are used in [25] and are included in several data-driven models for hourly solar radiation forecasting. Several machine learning methods including multilayer perceptron (MLP), SVR, k-nearest neighbors (kNN), and decision tree regression are used as the data-driven models and evaluated for their forecasting accuracies. The evaluation results show that the MLP method outperforms other data-driven forecasting models.
3. Performance metrics

Several indexes are presented for evaluating the performance of time-series prediction methods. These indexes are used to represent the forecast error. Lower error values correspond to more accurate forecasting. This section provides the definitions and mathematical representations for most commonly used error indexes in forecasting.

Note that \( S_{\text{Actual}} \) is the observed value, \( \hat{S} \) is the predicted value (forecast), and \( N \) is the total number of observations.

3.1 MSE

The mean square error (MSE) is used as the accuracy performance indicator given by:

\[
\text{MSE} = \frac{1}{N} \sum_{n=1}^{N} \left( \hat{S}(n) - S_{\text{Actual}}(n) \right)^2
\]

(1)

3.2 nMSE

Normalized mean square error (nMSE) is an estimator of overall deviations between the forecasted and measured samples. It is defined as follows:

\[
\text{nMSE} = \frac{N \sum_{n=1}^{N} \left( \hat{S}(n) - S_{\text{Actual}}(n) \right)^2}{\sum_{n=1}^{N} \hat{S}(n) \cdot \sum_{n=1}^{N} S_{\text{Actual}}(n)}
\]

(2)

3.3 RMSE

The square root of MSE is given by:

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left( \hat{S}(n) - S_{\text{Actual}}(n) \right)^2}
\]

(3)

3.4 nRMSE

Normalized root mean square error (NRMSE) is calculated as follows:

\[
\text{nRMSE} = \sqrt{\frac{\frac{1}{N} \sum_{n=1}^{N} \left( \hat{S}(n) - S_{\text{Actual}}(n) \right)^2}{\frac{1}{N} \sum_{n=1}^{N} S_{\text{Actual}}(n)}}
\]

(4)

3.5 MAE

Mean absolute error is calculated as follows:

\[
\text{MAE} = \frac{1}{N} \sum_{n=1}^{N} |\hat{S}(n) - S_{\text{Actual}}(n)|
\]

(5)
3.6 nMAE

Normalized mean absolute error (nMAE) is given by:

\[
\text{nMAE} = \frac{1}{N} \sum_{n=1}^{N} \frac{|\hat{S}(n) - S_{\text{Actual}}(n)|}{\max(S_{\text{Actual}}(n))}
\]  

(6)

3.7 MAPE

Mean absolute percentage error is calculated as follows:

\[
\text{MAPE} = \frac{1}{N} \sum_{n=1}^{N} \frac{|\hat{S}(n) - S_{\text{Actual}}(n)|}{S_{\text{Actual}}(n)}
\]  

(7)

3.8 MBE

Mean bias error (MBE) is given by:

\[
\text{MBE} = \frac{1}{N} \sum_{n=1}^{N} \hat{S}(n) - S_{\text{Actual}}(n)
\]  

(8)

3.9 nMBE

Normalized mean bias error (nMBE) is calculated by normalizing the MBE as follows:

\[
\text{nMBE} = \frac{\sum_{n=1}^{N} \hat{S}(n) - S_{\text{Actual}}(n)}{\sum_{n=1}^{N} S_{\text{Actual}}(n)}
\]  

(9)

3.10 MABE

The MABE represents the mean absolute bias error as follows:

\[
\text{MABE} = \frac{1}{N} \sum_{n=1}^{N} |\hat{S}(n) - S_{\text{Actual}}(n)|
\]  

(10)

3.11 SMAPE

The SMAPE represents symmetric mean absolute percentage error as follows:

\[
\text{SMAPE} = \frac{1}{N} \sum_{n=1}^{N} \frac{|\hat{S}(n) - S_{\text{Actual}}(n)|}{\left(\hat{S}(n) + S_{\text{Actual}}(n)\right)} \times 100
\]  

(11)

3.12 Forecast skill

The forecast skill is an error index which provides accuracy comparison between any given forecasting model and the persistence forecasting method. The forecast skill is calculated by:
Forecast Skill = \left( 1 - \frac{\text{RMSE}}{\text{RMSE (Persistence)}} \right) \quad (12)

A forecast skill of 1.0 implies an unattainable perfect forecasting, and a forecast skill of 0.0 indicates a performance similar to the persistence method. Besides, negative values show lower forecasting accuracies compared to the persistence method. Detailed explanations and applications of the forecast skill can be found in [26].

Table 1 provides a summary of the solar radiation forecasting methods explained in subsections 2.1–2.16, along with the error metrics used for each method.

<table>
<thead>
<tr>
<th>References</th>
<th>Forecast variable</th>
<th>Forecast horizon</th>
<th>Error metric</th>
<th>Method</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>[4]</td>
<td>Hourly GHI</td>
<td>1 h</td>
<td>nRMSE, nMBE</td>
<td>Hybrid SOM-SVR-PSO</td>
<td>Colorado, USA and South Park, Singapore</td>
</tr>
<tr>
<td>[6]</td>
<td>Daily global solar irradiation</td>
<td>1 day</td>
<td>RMSE, MBE</td>
<td>Functional fuzzy approach</td>
<td>World Radiation Data Center (WRDC), St. Petersburg, Russia</td>
</tr>
<tr>
<td>[8]</td>
<td>Hourly solar radiation</td>
<td>1 h</td>
<td>nRMSE, MAPE</td>
<td>A decomposition-clustering-ensemble learning approach</td>
<td>Beijing, China</td>
</tr>
<tr>
<td>[10]</td>
<td>Hourly solar radiation</td>
<td>1 h</td>
<td>nRMSE, MAPE</td>
<td>Hybrid k-means-NAR</td>
<td>National Meteorological Office of Algeria for the site of Oran</td>
</tr>
<tr>
<td>[13]</td>
<td>Hourly GHI</td>
<td>1 h</td>
<td>nRMSE, forecast skill</td>
<td>Combination of bootstrap sampling, k-fold approach, mutual information, and stationary processing</td>
<td>Global Earth Observation System of Systems (GEOSS)</td>
</tr>
<tr>
<td>[14]</td>
<td>Very short term GHI</td>
<td>5-min</td>
<td>MAPE</td>
<td>Combination of ANN with fuzzy logic pre-processing</td>
<td>Singapore</td>
</tr>
<tr>
<td>[16]</td>
<td>Hourly solar radiation</td>
<td>1 h</td>
<td>MBE, RMSE, nRMSE</td>
<td>AP-based semi-empirical models</td>
<td>Solar observation station at Afyon Kocatepe University</td>
</tr>
<tr>
<td>[17]</td>
<td>Hourly GHI</td>
<td>1 h</td>
<td>MAE, rMAE, nRMSE</td>
<td>Combination k-means, DT and SVM-C</td>
<td>Photovoltaic Systems Laboratory of the University of Malaga</td>
</tr>
<tr>
<td>[18]</td>
<td>Hourly solar radiation</td>
<td>1 h–48 h</td>
<td>RMSE</td>
<td>Machine learning-based models</td>
<td>National renewable energy laboratory (NREL)</td>
</tr>
<tr>
<td>[19]</td>
<td>Daily global solar irradiation</td>
<td>1 day</td>
<td>nMAPE, nRMSE</td>
<td>Combination of SOM and wavelet neural networks</td>
<td>Zhejiang Global Photovoltaic Technology</td>
</tr>
</tbody>
</table>


### 4. Conclusion

This chapter provides a comprehensive literature review to demonstrate the application of pattern recognition and machine learning techniques for solar radiation forecasting. The results of this survey show that identifying the irregular patterns of solar radiation time series by data clustering and/or classification provides better training for neural networks and enhances the forecast accuracy. However, computational complexities of hybrid forecasting methods utilizing multiple pattern recognition and machine learning techniques render their applications inefficient for online predictions or very short-term forecasting.

### Table 1.

Summary of the pattern recognition-based solar forecasting methods.

<table>
<thead>
<tr>
<th>References</th>
<th>Forecast variable</th>
<th>Forecast horizon</th>
<th>Error metric</th>
<th>Method</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>[20]</td>
<td>Daily global solar irradiation</td>
<td>1 day</td>
<td>RMSE, MAPE</td>
<td>ANFIS-based approach</td>
<td>Tabas County, South Khorasan Province, Iran</td>
</tr>
<tr>
<td>[21]</td>
<td>Daily global solar irradiation</td>
<td>1 day</td>
<td>RMSE, MAPE, MABE</td>
<td>SVR-based methodology</td>
<td>Arid desert cold climate of Isfahan, Iran</td>
</tr>
<tr>
<td>[23]</td>
<td>Daily global solar irradiation</td>
<td>1 day</td>
<td>RMSE, MAPE, MABE</td>
<td>Hybrid SVM-WT-ANN-GA</td>
<td>Bandar Abbas, Iran</td>
</tr>
<tr>
<td>[24]</td>
<td>Hourly solar radiation</td>
<td>1 h</td>
<td>SMAPE, RMSE</td>
<td>Clustering-based multi-model method</td>
<td>Nanyang Technological University, Singapore</td>
</tr>
</tbody>
</table>

*Global horizontal irradiance (GHI).  
**Direct normal irradiance (DNI).  
***Diffuse horizontal irradiance (DHI).
Author details

Mahmoud Ghofrani¹*, Rasool Azimi² and Mastaneh Youshi³

1 School of Science, Technology, Engineering and Mathematics (STEM), University of Washington, Bothell, USA

2 Young Researchers and Elite Club, Qazvin Branch, Islamic Azad University, Qazvin, Iran

3 Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

*Address all correspondence to: mrani@uw.edu

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
References


[18] Sharma A, Kakkar A. Forecasting daily global solar irradiance generation using machine learning. Renewable and
Sustainable Energy Reviews. 2018;82:2254-2269


