Solar Energy 158 (2017) 595–609

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Minimum redundancy – Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series

Hassen Bouzgou[a,*], Christian A. Gueymard[b]

[a] Department of Industrial Engineering, Faculty of Technology, University of Batna 2 (Mustapha Benboulaid), Avenue Boukhlouf Med El Hadi, 05000 Batna, Algeria
[b] Solar Consulting Services, Colebrook, NH, USA

ARTICLE INFO

Keywords:
Global irradiance
Mutual information
Minimum redundancy - maximum relevance
Sequential forward selection
Extreme learning machine
Time series

ABSTRACT

Solar energy is expected to provide a major contribution to the future global energy supply, while helping the transition toward a carbon-free economy. Because of its variable character, its efficient use will necessitate trustworthy forecast information of its availability in a variety of spatial and time scales, depending on application. This study proposes a new forecasting approach for irradiance time series that combines mutual information measures and an Extreme Learning Machine (ELM). The method is referred to as Minimum Redundancy – Maximum Relevance (MRMR). To assess the proposed approach, its performance is evaluated against four scenarios: long window (latest 50 variables), short window (latest 5 variables), standard Principal Components Analysis (PCA) and clear-sky model. All these scenarios are applied to three typical forecasting horizons (15-min ahead, 1-h ahead and 24-h ahead). Based on measured irradiance data from 20 sites representing a variety of climates, the test results reveal that the selection of a good set of relevant variables positively impacts the forecasting performance of global solar radiation. The present findings show that the proposed approach is able to improve the accuracy of solar irradiance forecasting compared to other proposed scenarios.

1. Introduction

Modern electric grids become “smarter”, but face the need of unavoidable changes to successfully tackle the increased variability introduced by irregular generation of renewable energies (RE)—wind and solar most particularly. Several approaches have been proposed to reduce the unwanted effects of their integration into the grid, e.g. by increasing storage capacity, improving resource and load forecasting, developing appropriate demand response, etc. (Kaur et al., 2016). Although power systems have been designed to handle the variable nature of loads, the extra supply variability and uncertainty can cause many difficulties for system operators and utilities in general. Nevertheless, different operational and technical solutions exist to facilitate the integration of RE generation (Bird et al., 2013; Kalogirou, 2014). Still, more flexibility is required in the grid system to accommodate supply-side variability and the relationship between generation levels and loads. Under ideal conditions, the RE generation varies in the same direction as the load, and can be completely used to meet it. In the real world, further actions are required to balance the system if non-ideal situations occur, which constitutes a risk at any moment. System operators need to guarantee that they have satisfactory resources to contain fast ramps in solar generation to keep system balance. They also want to keep generation costs at a minimum, which means they have direct interest in using RE generation as much as possible at any time. One more challenge arises when RE generation is available during low load levels; in some cases, usual generators may switch to their minimum generation levels (Bird et al., 2013). When RE generation represents a large fraction of the total generation, the former may even surpass the load during at least parts of the day, thus creating difficulties. Such adverse conditions need to be forecasted and prevented.

A good assessment of the solar resource available to a solar energy conversion system starts by characterizing the amount of solar irradiance that was available over a specific region or location during a recent time period of typically 15–20 years. This provides the necessary data to design the system and obtain its financing. After construction and commissioning, both the system and grid operators typically need to rely on solar radiation forecasting for the routine operations of the plant and electrical grid. Most importantly, precise solar energy forecasting is vital to guarantee the smooth operation of the grid, lower risks of curtailment from RE generation (energy dumping), lower

* Corresponding author.
E-mail addresses: bouzgou@gmail.com, h.bouzgou@univ-batna2.dz (H. Bouzgou), Chris@SolarConsultingServices.com (C.A. Gueymard).

http://dx.doi.org/10.1016/j.solener.2017.10.035
Received 25 July 2017; Received in revised form 17 September 2017; Accepted 9 October 2017

Table 1
Solar forecasting techniques: characteristics and inputs; adapted and augmented from Coimbra et al. (2013).

Technique	Sampling rate	Spatial resolution	Spatial extent	Suitable forecast horizon	Application
Persistence	High	Single point	Single point	Minutes	Baseline
Total-sky imagery (camera)	30 s	10–100 m	2–5 m radius	10–20 min	Short-term ramps, regulation
Satellite imagery	15 min	1 km	Country	5 h	Load following
Numerical Weather Prediction model	1 h	12 km	Country	10 days	Unit commitment
Artificial intelligence	1 time step	Single point	Single point	A few time steps	Any

energy costs, and/or maximize revenue of RE power plants (Kleissl, 2013).

Diverse time horizons have to be considered with respect to the forecasting application. Hence, selecting an appropriate solar forecasting model depends primarily on the timescale involved. Depending on application and energy market, the forecasting time horizon typically varies from a few seconds or minutes (intrahour), a few hours (intraday), to one or a few days ahead (intraweek). Intraday forecasts may be less important than day-ahead forecasts in terms of economic value. Nevertheless, with increasing solar penetration and the improvement of intraday accuracy, large market opportunities have started to materialize. Accordingly, medium-term (< 48 h) forecasts are relevant for the scheduling and planning of RE generation, whereas intraday forecasts are helpful for load following and predispatch, thus alleviating the necessity to control the frequency in real time (regulation) (Coimbra et al., 2013). The typical forecast horizons are summarized in Table 1.

The methodology proposed in this study refers to the last row of Table 1 and utilizes a time series of observations having the appropriate time step to provide accurate forecasts for a few steps ahead. The main objective of the study is to evaluate the feasibility of using the proposed hybrid (MRMR-ELM) model to forecast global solar radiation for three time horizons (intra-hour, intra-day and intra-week). To achieve this, solar irradiance measurements from 20 sites around the world representing four different climates are used. The novelty of the proposed method resides in its efficient selection of temporal variables in the time series through an optimum dimensionality reduction technique.

The remainder of the paper is organized as follows. A literature review is proposed in Section 2. The methodology is described in detail in Section 3. In Section 4, the database used to evaluate the performance of the proposed methodology is presented. Section 5 describes some methods used to construct the different learning models. The experimental results of the different forecasting horizons are presented and discussed in Section 6. Finally, Section 7 draws the conclusion of this work.

2. Literature review

The accurate forecasting of global horizontal irradiance (GHI) is an essential capability for most PV power prediction systems. Solar radiation forecasting approaches may be classified into four main categories: (i) Statistical and Artificial Intelligence (AI) models; (ii) Remote-sensing models using ground-based or satellite-based cloud imagery; (iii) Numerical weather prediction models; and (iv) Hybrid models, using two or more of the above techniques. Some detailed reviews about solar radiation forecasting and power generation methods can be found in the literature (Diagne et al., 2013; Inman et al., 2013; Khatib et al., 2012; Voyant et al., 2017; Yadav and Chandel, 2014). Since a growing interest in research on AI techniques is now noticeable in a variety of scientific fields, the succinct literature review that follows focuses on statistical and AI forecasting methods.

As early as 2000, (Sfetsos and Coonick, 2000) investigated numerous AI-based techniques, including linear, feed-forward, recurrent Elman and Radial Basis neural networks, along with the adaptive neuro-fuzzy inference scheme. They indicated that AI models had the potential to forecast solar radiation time series more efficiently than conventional procedures based on the clearness index. Mihalakakou et al. (2000) proposed a neural network (NN) technique to forecast short-term total solar radiation time series. The future GHI hourly values are predicted for several years by extracting information from past values, using feed-forward back-propagation NN. They concluded that the NN approach performs better than the auto-regressive (AR) model. Dorvlo et al. (2002) proposed an advanced ANN (Radial Basis Functions and Multilayer Perceptron) to estimate solar radiation based on the clearness index; these models were then assessed using long-term data from eight stations in Oman. In (Sözen et al., 2004), the authors proposed to evaluate the solar energy potential of Turkey using ANNs, based on data from 17 stations spread over Turkey (11 stations for training and 6 for testing). Each approach was investigated for each station by using different learning algorithms and a logistic sigmoid transfer function in the ANN. The predicted ANN-based solar potential was represented as monthly maps.

Mellit et al. (2006) proposed an adaptive wavelet-network to forecast the daily solar irradiance, and stated that training wavelet-networks needed a smaller number of iterations compared to other NN techniques, and that the resulting model could be used to fill data breaks in weather databases. In (Hocaoğlu et al., 2008), a two-dimensional (2D) representation model of the hourly solar radiation data was proposed using image-processing methods. Nine different linear filters with various filter-tap configurations were optimized and tested through feed-forward NN. The authors found that NN models performed better than linear prediction filters. Cao and Lin (2008) proposed a model to forecast GHI based on diagonal recurrent wavelet neural network (DRWNN) and a specially designed training algorithm. Experimental results showed that the model was able to effectively map the solar irradiance magnitude by combining both recurrent NN and wavelet NN. Reikard (2009) proposed to forecast GHI using Autoregressive Integrated Moving Average (ARIMA) and unobserved-component models applied to resolutions of 5, 15, 30, and 60 min. He stated that ARIMA captured the diurnal cycle of solar radiation more successfully than any other proposed models. Mellit and Pavan (2010) also proposed a multilayer perceptron (MLP) model to forecast solar irradiance on a 24-h horizon basis using daily-mean solar irradiance and air temperature. Their results indicated that the proposed model performs well, with correlation coefficients in the range 98–99% for sunny days and 94–96% for cloudy days. In parallel, Martín et al. (2010) presented a comparison of various types of models (AR, NN and fuzzy logic) able to forecast half-daily values of GHI with a forecasting horizon of 3 days. They concluded that all these AI models outperformed persistence, and that the best approach was NN with the difference between extraterrestrial solar irradiance and ground measured global solar irradiance (which they called “lost component”) as input. Voyant et al. (2011) proposed to study the contribution of exogenous meteorological data (multivariate method) as time series to an optimized MLP. The method was compared with different forecasting methods: persistence, ARIMA, an ANN with preprocessing using only endogenous inputs (univariate method) and an ANN with preprocessing using endogenous and exogenous inputs. They concluded that exogenous data shows potential interest in winter, whereas inputs of endogenous data to a preprocessed ANN seemed sufficient in summer.