Energy Conversion and Management 95 (2015) 406–413

Contents lists available at ScienceDirect

Energy Conversion and Management

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

A hybrid method for forecasting the energy output of photovoltaic systems

Pamela Ramsami, Vishwamitra Oree *

Electrical and Electronic Engineering Department, Faculty of Engineering, University of Mauritius, Reduit, Mauritius

ARTICLE INFO

Article history:
Received 27 August 2014
Accepted 16 February 2015

Keywords:
Prediction
Solar irradiance
Artificial neural network
Stepwise regression
Meteorological

ABSTRACT

The intermittent nature of solar energy poses many challenges to renewable energy system operators in terms of operational planning and scheduling. Predicting the output of photovoltaic systems is therefore essential for managing the operation and assessing the economic performance of power systems. This paper presents a new technique for forecasting the 24-h ahead stochastic energy output of photovoltaic systems based on the daily weather forecasts. A comparison of the performances of the hybrid technique with conventional linear regression and artificial neural network models has also been reported. Initially, three single-stage models were designed, namely the generalized regression neural network, feedforward neural network and multiple linear regression. Subsequently, a hybrid-modeling approach was adopted by applying stepwise regression to select input variables of greater importance. These variables were then fed to the single-stage models resulting in three hybrid models. They were then validated by comparing the forecasts of the models with measured dataset from an operational photovoltaic system. The accuracy of the each model was evaluated based on the correlation coefficient, mean absolute error, mean bias error and root mean square error values. Simulation results revealed that the hybrid models perform better than their corresponding single-stage models. Stepwise regression-feedforward neural network hybrid model outperformed the other models with root mean square error, mean absolute error, mean bias error and correlation coefficient values of 2.74, 2.09, 0.01 and 0.932 respectively. The simplified network architecture of the hybrid schemes suggests that they are promising photovoltaic output prediction tools, particularly in locations where few meteorological parameters are monitored.

1. Introduction

Increasing energy demands combined with rising conventional fuel costs and environmental awareness have contributed to the emergence of renewable energy sources during the last decade. Photovoltaic (PV) systems have sustained a remarkable annual growth rate, driven by several factors including technological innovation, improved cost effectiveness and government incentives. From 2000 to 2011, the International Energy Agency (IEA) reports that global PV installed capacity increased from 1 GW to 67 GW [1]. Nevertheless, significant constraints still hinder the large-scale integration of PV in the electricity mix. In particular, the unpredictability and variability of the solar energy cause major problems to the reliability and stability of existing grid-connected power systems. PV output forecasting is therefore essential for utility companies to plan the operations of power plants properly so as to ensure the stability, reliability and cost effectiveness of the system [2].

The energy generated by a PV system depends on the atmospheric conditions prevailing at the site of the installation, the geographical siting, the tilt and configuration, the ancillary electrical components used in the installation and the environment. For a fixed PV array which is already operational, all these parameters are constant except for the atmospheric factors. The electrical output parameters of PV systems are rated under standard test conditions (STC) corresponding to a PV cell temperature of 25 °C, a solar intensity of 1000 W/m^2 and an air mass of 1.5. However, the variable nature of the atmospheric conditions implies that the STC are rarely met in practice so that the output of PV panel in real operating conditions can differ considerably from the rated value.

Recently, the prediction of the energy generated by PV systems using climatic parameters has been thoroughly investigated. While the techniques applied to forecast the PV power production are extensions of those used for predicting the solar irradiance [3], the following literature review will focus mostly on prediction

* Corresponding author at: Electrical and Electronic Engineering Department, 5th Floor, Sir E. Lim Fat Engineering Tower, Faculty of Engineering, University of Mauritius, Reduit, Mauritius. Tel.: +230 57536381; fax: +230 4657144.

E-mail address: v.oree@uom.ac.mu (V. Oree).

http://dx.doi.org/10.1016/j.enconman.2015.02.052
0196-8904/

models dealing with PV power output. Based on the type of inputs employed, the forecasting techniques can be broadly classified as physical or statistical approaches. Physical methods model the PV as a function of some independent variables such as the PV cell characteristics, solar irradiance and cell temperature. Many physical models are derived from conventional solar cell equivalent circuits which aim at achieving the closest possible local solar radiation. Thus, Ayompe et al. [4] evaluated the PV output power forecasting accuracy by studying different cell temperatures and efficiency models. Zhou et al. [5] introduced five parameters in a model based on the *I–V* curves of a PV module to account for the complex output dependence.

Statistical prediction methods, on the other hand, rely on past historical data for prediction. Techniques such as regression analysis, time-series analysis and artificial intelligence analyze the historical dataset to forecast the PV power. Bacher et al. [6] used past average power measurements in an autoregressive model with exogenous inputs to predict the average output power of rooftop PV systems in Denmark. Da Silva Fonseca et al. [7] developed an approach based on support vector machines (SVMs) and numerically predicted weather variables to forecast the power production of a 1 MW PV plant in Japan. Normalized temperature, relative humidity and cloudiness at three different altitudes were used as inputs to the system. The authors observed that the numerically predicted cloud amount had a significant influence on the accuracy of the power output forecasts. Shi et al. [8] also suggested a methodology based on SVMs combined with a classification of weather prevailing on a typical day into four categories: clear sky, cloudy day, foggy and rainy day. The authors obtained promising results when testing their technique to forecast the power output of a 20 kW_p PV station in China.

Artificial neural networks (ANN) have been extensively used to predict solar output from meteorological variables due to their inherent ability to model non-linear, dynamic, noisy data and complex systems [3]. Thus, Chen et al. [9] developed an online 24-h ahead forecasting model which uses a radial basis function network (RBFN) as well as a self-organized map to classify the input variables of numerical weather predictions, consisting of average daily values of solar irradiance, air temperature, wind speed and humidity. Using daily values of measured and theoretical sunshine duration, maximum temperature as well as the month number during the period 1986 to 1992 in Cyprus, Tymvios et al. [10] trained seven ANN models. The best results were obtained when all the climate inputs were applied to the system. Saberian et al. [11] used generalized regression neural network (GRNN) and feed-forward neural network (FFNN) to predict the power output of a PV model using minimum temperature, maximum temperature, mean temperature and irradiance as inputs. It was found that while both FFNN and GRNN performed satisfactorily, FFNN produced more accurate results. Izgi et al. [12] attempted to exploit the learning ability of an ANN to minimize prediction errors in the energy generation of a 750 W_p PV system in Istanbul, Turkey. The input parameters to the model consisted of cell temperature, ambient temperature together with global and diffuse solar irradiance. The authors observed that accurate prediction results were obtained in the month of August when the weather conditions were stable. Mellit et al. [13] developed three distinct ANN models for the short-term forecasting of the power produced by a 1 MW_p grid-connected PV plant in Southern Italy using future values of solar irradiance and cell temperature as well as the present value of power output as input parameters. The different ANNs corresponded to three databases classified on the basis of typical daily weather conditions: sunny, partly cloudy and overcast. Results proved that the model developed for overcast days was more accurate than the one built on the entire dataset, with the correlation coefficient (R) of all ANN models exceeding 0.98.

Several studies have also been performed to compare the performances of different statistical techniques that predict solar power output using climatic variables. In particular, Almonacid et al. [14] compared the results of three conventional mathematical models with those of an ANN for predicting the annual energy output of a PV generator. The authors concluded that the predictions of the ANN were far more accurate than the results of the classical models as the former takes into account the second-order energy losses due to factors like low irradiance, shading and spectral effects in addition to the usual temperature and irradiance losses considered by the mathematical models. Oudjana et al. [15] found that an ANN model outperformed regression techniques as a forecasting tool for the power output of a PV generator in the Ghardaia province of Algeria with solar radiation and temperature as independent variables. By means of an approximated function for the clear-sky model, Pedro and Coimbra [3] assessed five techniques, namely the Persistent model, the Autoregressive Integrated Moving Average (ARIMA), the k-Nearest-Neighbour, ANNs and the hybrid Genetic Algorithm-ANN, for short-term forecasting of the power output of a 1 MW_p PV plant in California. The error statistics indicated that the ANN-based forecasting models performed better than the other forecasting techniques, with Root Mean Square Error (RMSE) and R values of 11.42 and 0.97 respectively for 1-h ahead predictions. Furthermore, the prediction accuracy could be enhanced by optimizing the ANN parameters with the genetic algorithm.

A key strength of ANN-based forecasting methods is that the designer can leverage the ability to select several inputs in order to improve the forecast accuracy. For instance, Azadeh et al. [16] considered seven meteorological input parameters to an ANN model so as to obtain better prediction capability of the solar output in various regions of Iran as compared to conventional prediction tools. Integrating too many input variables in an ANN model can however give rise to many implementation issues. The most obvious consequence of using too many variables is a significant increase in the computation time devoted to training and querying the network. This is particularly the case in a Multilayer Perceptron (MLP) network where each additional input parameter is multiplied by the number of hidden neurons. Moreover, applying many input parameters to the model increases the risk of having redundant variables and further complicates the training stage. Redundant variables increase the number of local optima in the error function, resulting in higher risks of suboptimal convergence and slowing down of the learning process [17]. They also contribute negligibly to the performance of the output variable while enhancing the complexity of the model. Finally, from a statistical perspective, each additional variable in the input vector inserts another dimension to the output space. The training stage then requires sufficient data to populate the space densely enough to represent the mapping relationship [18]. More input parameters therefore imply that a bigger training dataset is required to keep an equivalent accuracy in the mapping. Datasets of sufficiently big size might not be available at many locations.

A few studies have been conducted to determine the most influential meteorological parameters that will serve as inputs in ANN-based forecast models. Sfetsos and Coonick [19] investigated the use of additional meteorological variables as potential inputs to solar output prediction models. For this purpose, the authors used a two-step technique based on trial and error. Initially, the model was trained with each candidate variable until the minimum training error was achieved. Subsequently, the influence of the variable was removed by replacing it with its mean value or zero. In another study performed by Marquez and Coimbra [20], a strategy based on a Gamma test was used to identify the most relevant input variables among numerous meteorological parameters. A genetic algorithm search was also included in the selection procedure in order