Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 126 (201709) 651–658

Energy Procedia

www.elsevier.com/locate/procedia

72nd Conference of the Italian Thermal Machines Engineering Association, ATI2017, 6-8 September 2017, Lecce, Italy

Forecasting of PV Power Generation using weather input data-preprocessing techniques

Maria Malvoni*, Maria Grazia De Giorgi, Paolo Maria Congedo

Dipartimento di Ingegneria dell'Innovazione, Università del Salento, via per Monteroni - 73100 Lecce , Italy

Abstract

Stochastic nature of weather conditions influences the photovoltaic power forecasts. The present work investigates the accuracy performance of data-driven methods for PV power ahead prediction when different data preprocessing techniques are applied to input datasets. The Wavelet Decomposition and the Principal Component Analysis were proposed to decompose meteorological data used as inputs for the forecasts. A time series forecasting method as the GLSSVM (Group Least Square Support Vector Machine) that combines the Least Square Support Vector Machines (LS-SVM) and Group Method of Data Handling (GMDH) was applied to the measured weather data and implemented for day-ahead PV generation forecast.

Peer-review under responsibility of the scientific committee of the 72nd Conference of the Italian Thermal Machines Engineering Association

Keywords: data-driven forecast; data preprocessing, GLSSVM, wavelet, PCA, day-ahead forecast; photovoltaic; solar power.

1. Introduction

The European policies impose strict targets to reach a significant integration of renewable energy sources (RES) by year 2030 [1]. The development of forecasting models represents an essential requirement to support high sharing of renewable such as solar and wind, allowing to deal with the stochastic nature and the variability of the renewable

* Corresponding author. Tel.: +39 0832 297782.
E-mail address: maria.malvoni@unisalento.it

1876-6102
Peer-review under responsibility of the scientific committee of the 72nd Conference of the Italian Thermal Machines Engineering Association
10.1016/j.egypro.2017.08.293

sources in order to ensure the safe and reliability of the electric grid. In the small scale grid integration as microgrid, the prediction of PV power generation leads also to challenges for innovative energy management systems by the development of an integrated control of energy flows between loads and sources [2].

The PV power generation forecasting is a current research topic and several approaches have been investigated in the literature [3]. The data-driven forecast methods also known as time series forecasting methods need historical power measurements [4]. They include statistical and stochastic machine learning techniques and have already demonstrated high performance for such purpose [5]. On the other hand, the performance of a grid-connected PV system depends on the climate in which it is located [6]-[7], therefore the weather variables have a significant impact on the accuracy of the forecasting systems.

The weather parameters as the solar irradiance on the plane of the array and the ambient temperature can improve the forecast accuracy when actual measurements are used as input data to predict the PV output power [8]-[9]. In addition the wind speed has also a relevant influence on the predictions of PV power, leading more accurate estimations [10]-[11].

The data-driven forecast methods have to deal with a huge amount of data. In the PV power forecasts based on the historical data, redundant information can led more complex modelling. The principal component analysis (PCA) is one of the most popular techniques to remove such redundancy. It has already demonstrated its efficiency in the reduction of the dataset size without loss of information, improving the forecast accuracy with a sustainable computational time [12]-[13]. The weather variables are stochastic and introduce random fluctuations in the predicted PV power when they are used as input data in the forecast systems [8]. The wavelet technique represents an efficient solution to reduce the noise in input datasets before to implement a prediction model. The wavelet decomposition (WD) applied to the input data of a forecasting model allows to deal with the solar irradiance fluctuations, leading to the accuracy improvement [14].

The novel supervised learning algorithm GLSSVM was developed by the authors in [5], combining the group method of data handling (GMDH) and the least squares support vector machine (LS-SVM) and was implemented to predict the hourly PV output power in one-day ahead time frame. Results demonstrated that the GLSSVM outperforms the mother models, allowing to obtain lower forecast errors of the PV output power up to 24 hours ahead than the LS-SVM and the GMDH algorithms.

The purpose of this study is to investigate the performance of the data-driven forecast methods coupled with data pre-processing techniques. The forecast performance of the GLSSVM was evaluated when different preprocessing data techniques are applied to the weather historical data series to forecast the PV power output at several ahead time horizons. The wavelet decomposition and the Principal Component Analysis are the main preprocessing methods adopted in this work to decompose the time series of weather data. The input data decomposition results were applied to train the hybrid forecasting model GLSSVM at different horizons, from 1 h to 24 h ahead. This paper is organized as follows. Section 2 describes the data-preprocessing techniques and the forecasting model GLSSVM which were implemented. The input data description, the error metrics to assess the accuracy performance and the investigation methodology are presented in section 3. Results and discussions are described in section 4, and conclusions of this study are given in section 5.

2. Methods

2.1. PCA and WD for data-preprocessing techniques

2.1.1. Principal Component Analysis

The PCA is a statistical technique that allows resizing a dataset, taking into account uncorrelated and redundant information. The PCA technique finds the most significant variations of the variables. The covariance method is one of the approaches to implement the PCA [15]. Starting from a dataset $\mathcal{S}$ of dimensions Q × M where M is the observations number of Q variables, it is possible to obtain a new subset of L variables with L< Q.

Given the matrix B of dimension M × Q:

$$B = \mathcal{S} - h\, u^T \qquad (1)$$

where $u[j] = \frac{1}{M}\sum_{i=1}^{M}\mathcal{S}[i,j]$ and $h[i] = 1$ with j=1…Q and i=1…M, the Q × Q covariance matrix C of matrix B is defined as:

$$C = \frac{1}{M-1}\, B^T B \qquad (2)$$