Renewable Energy 96 (2016) 469–478

Contents lists available at ScienceDirect

Renewable Energy

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

One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models

Guochang Wang [a], Yan Su [b], Lianjie Shu [c,*]

[a] College of Economics, Jinan University, GuangZhou 510632, China
[b] Department of Electromechanical Engineering, University of Macau, Taipa, Macau
[c] Faculty of Business Administration, University of Macau, Taipa, Macau

ARTICLE INFO

Article history:
Received 18 December 2015
Received in revised form
16 March 2016
Accepted 29 April 2016

Keywords:
Partial functional linear regression
Solar irradiance
Photovoltaic system
Efficiency

ABSTRACT

The intra-day time-varying pattern of solar data is more informative than the aggregated mean daily data. However, most of the traditional forecasting models often construct the 1-day ahead daily power forecast based on its historical daily averages but ignore the information from its intra-day dynamic pattern. Intuitively, the use of aggregated data could cause certain loss of information in forecasting, which in turn adversely affects forecasting accuracy. In order to make use of the valuable trajectory information of the power output within a day, this paper suggests a partial functional linear regression model (PFLRM) for forecasting the daily power output of PV systems. The PFLRM is a generalization of the traditional multiple linear regression model but enables to model nonlinearity structure. Compared to the neural network models that are often criticized by the requirements of past experience and reliable knowledge in the design of network architecture, the suggested method only involves a few parameter estimates. A regularized algorithm was used to estimate the PFLRM parameters. It is shown that the regularized PFLRM improves the forecast accuracy of power output over the traditional multiple linear regression and neural network models. The results were validated based on a 2.1 kW grid connected PV system.

1. Introduction

In recent years, introduction of an alternative energy resource has been expected. Due to the advantages of being clean, abundant, and inexhaustible, solar energy has received gradually increasing attention as one of the best solutions for the alternative energy resources, and photovoltaic (PV) technology has been rapidly developed recently. PV systems become more and more popular in grid-connected applications rather than being established in remote areas [1]. However, the power output of a PV system is not constant but depends on solar irradiation and weather conditions. For a PV system, there are many factors that can influence the power output such as solar irradiance, temperature, insolation, and installation angle. Due to the variability of solar irradiance and environmental factors, the power output of a PV system is dynamically changing with time.

The variability of power output not only adversely affects the stability of the electrical system being connected but also may increase operating costs for the electricity system by increasing requirements of primary reserves [2]. For this reason, accurate prediction of power output is an important task to improve the integration stability of output of a solar PV system into electric grid and to help producers to implement operations strategy in an efficient way. The forecast of power output can be made at different time scales for different purposes. The short-term like intra-hour forecasts are often relevant for dispatching, regulatory and load following purpose [3]. However, the 1-day ahead forecast of the PV system is often required by the end-users such as energy traders for the purpose of trading the electricity market and transmission system operators for operational planning. This paper limits discussion to the accurate 1-day ahead forecast of PV power output.

A lot of research has been devoted to the development of forecasting tools for predicting PV power output with good accuracy. These forecasting tools generally fall into two categories, indirect and direct approaches. The indirect approach involves a two-step procedure for forecasting PV power output. In the first step, the solar irradiance is forecasted based on various approaches, including time series models, artificial neural networks, support

* Corresponding author.
E-mail address: ljshu@umac.mo (L. Shu).

http://dx.doi.org/10.1016/j.renene.2016.04.089

Nomenclature

s or t	time period
y	response
$x(s)$ or $x(t)$	functional predictor
$\mu(t)$	mean of $x(t)$
$\mathbf{z}$	a vector of covariates
α	the intercept term for the PFLRM
$\beta(t)$	regression coefficient function
$\boldsymbol{\gamma}$	regression coefficient vector
T	the time set
ε	random error
ξ_i	functional principal component (FPC) score
K	number of leading components selected
$G(s,t)$	covariance function
λ_k	the kth eigenvalue
$\phi_k(s)$	the kth eigenfunction
c_k	the kth basis coefficient
$\widehat{\xi}_i$	(or $\widehat{\lambda}_k, \widehat{\phi}_k(s), \widehat{\beta}(t)$) estimate of ξ_i (or λ_k, $\phi_k(s)$, $\beta(t)$)
$\widehat{G}(s,t)$	estimate of $G(s,t)$
$\widehat{y}_i$	fitted value of y_i
n	sample size
t_{ij}	the observation time of the jth observation in day i
U_{ij}	the recorded power output at time t_{ij}
e_{ij}	measurement error
$x_i(t_{ij})$	the ith realization of $x(t)$
L_i	number of observations in day i
$\widehat{\mu}(t)$	estimate of $\mu(t)$
$\widehat{c}_k$	estimate of c_k
$\widehat{\boldsymbol{\gamma}}$	estimate of $\boldsymbol{\gamma}$

vector machines, and others. For example, Lopez and Cardona [4] and Martín et al. [5] discussed the use of time series models for predicting global irradiance while Reikard [6] made a comprehensive comparison of time series models for predicting solar radiation at high resolution. Another frequently used method to forecast solar irradiance is based on artificial neural network. A sample of research in this respect includes [7–11]. Chen et al. [12] and Wu and Liu [13] employed support vector machine to estimate monthly solar radiation from measured temperatures. Cao and Lin [14] proposed a new model for forecasting global solar irradiance based on diagonal recurrent wavelet neural network. In the second step of the indirect approach, the forecasted solar irradiance and temperature data are often used as inputs in the commercial simulation softwares such as TRNSYSM [15], PVFORM [16] and HOMER [17].

The second category aims at directly predicting the power output based on some prior information or readily accessed data. Due to the similarity of forecasting solar irradiance and power output, it is straightforward to extend the methods for forecasting solar irradiance to the forecast of power output of PV systems. A list of research on the use of time series models for direct forecasting power output includes [3,18–20]. A sample of research on the use of neural network models for this purpose can be referred to [21–25].

With rapid development and advancement in sensor and data acquisition technology, the solar data are often recorded at high frequency such on the basic of 1 min, which can be also seen from the underlying data set collected from a 2.1 kW grid connected PV system. The high frequency solar data depicts the time-varying pattern (or trajectory) of the power output generated within a day. Clearly, the intra-day time-varying pattern of the power output is more informative than its daily average. However, most of the traditional methods construct the 1-day ahead forecast of power output based on the aggregated data, i.e., the historical daily power output, but ignore the valuable intra-day information. Intuitively, the intra-day trajectory of power output could be used to improve the forecast accuracy. Motivated by this, this paper suggests a new model based on the partial functional linear regression model (PFLRM) for generating the 1-day ahead forecast of the daily power output of PV systems. Although the PFLRM has been widely discussed in the statistical literature and successfully employed in other applications [27,28], it has been less studied for forecasting power output.

The PFLRM is a generalization of both the classical multiple linear regression model and functional linear regression model, which has been proved to be a powerful tool in forecasting [26]. It can include function-valued random variable and some real-valued variables as predictor variables. Therefore, it is natural to apply the PFLRM to forecast the power output of PV systems by treating the intra-day trajectory of power output of the PV system in the previous day as a functional predictor and some real-valued climatic variables as covariates in the model. The PFLRM maintains simplicity as the traditional multiple linear regression model but provides more flexibility than the latter to model the nonlinear structure in the data. As compared to the completely data-driven forecasting models such as neural network models, the PFLRM achieves certain simplicity. The PFLRM often involves only a few parameter estimates that are entirely driven by the regression data. Instead, the design of neural network models often involves the design of network architecture and the selection of a good learning algorithm. There are many parameters that need to be prespecified. This heavily relies on past experience and is subject to trial and error processes since the optimal configuration is not known a priori. This is a limitation of neural network models often criticized in the literature.

The remainder of this paper is organized as follows. In Section 2, we briefly review the PFLRM and discuss the parameter estimation based on a regularized algorithm. In Section 3, the regularized PFLRM was identified and validated based on a data set collected from a 2.1 kW grid connected PV system during Jan. 1 2011 to June 30, 2012. In Section 4, the prediction performance between the regularized PFLRM and some neural network models is compared. Some concluding remarks are given in Section 5.

2. The partial functional linear regression model

2.1. Classical functional linear model via functional principal components

Functional principal component (FPC) analysis is a statistical method for investigating the dominant variation of functional data. Denote $x(t)$ as the square-integrable random function defined on a closed interval T of the real line. Let $\mu(t) = E(x(t))$ and $G(s,t) = cov(x(s),x(t))$ be the mean and covariance of $x(t)$, respectively. The classic functional linear model is defined as