Available online at www.sciencedirect.com

ScienceDirect

Solar Energy 111 (2015) 320–331

SOLAR ENERGY

www.elsevier.com/locate/solener

Tailored vs black-box models for forecasting hourly average solar irradiance

Marek Brabec [a], Marius Paulescu [b,*], Viorel Badescu [c,d]

[a] Department of Nonlinear Modeling, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic

[b] Physics Department, West University of Timisoara, V. Parvan 4, 300223 Timisoara, Romania

[c] Candida Oancea Institute, Polytechnic University of Bucharest, Spl. Independentei 313, 060042 Bucharest, Romania

[d] Romanian Academy, Calea Victoriei 125, Bucharest, Romania

Received 28 May 2014; received in revised form 1 November 2014; accepted 3 November 2014
Available online 26 November 2014

Communicated by: Associate Editor David Renne

Abstract

Accurate prediction of solar radiation is of high importance for proper operation of the electrical grid. Over short horizons, forecasting solar irradiance is often performed by extrapolation of field measurements. Four *tailored* statistical models for forecasting hourly average solar irradiance are proposed and assessed in this paper. These follow from the well-known regression and ARIMA class of models, but bring into the model formulation various physically motivated additional features. These capture the distribution of solar radiation more effectively. Their performance is compared with the performance of a standard model used in the strictly black-box style often encountered in practice. Overall results demonstrate that the proposed models are significantly more accurate than the standard model, under conditions of mostly cloudy skies.

Keywords: Solar irradiance; Forecasting; Tailored statistical models

1. Introduction

Recently, the importance of solar electricity in the energy mix increased dramatically and this trend is generally expected to continue. The photovoltaic (PV) cumulative installed capacity in Europe was enlarged from less than 1 GW in 2003 to over 30 GW in 2010 and 70 GW in 2012 (EPIA, 2013). Sometimes, aided by favorable policies, the growth exceeds the most optimistic predictions. An amazing example is Romania, where on December 2011 the installed PV capacity was less than 2 MW and the Governmental PV Systems Strategy was targeting 260 MW by 2020 (Iacobescu and Badescu, 2012); in May 2014 the PV installed capacity in Romania was very close to 1.1 GW (Transelectrica, 2014).

A challenge standing against the growing importance of PV electricity in the energy mix stems from the stochastic nature of the solar energy, fluctuating in time due to weather pattern irregularities. Therefore, in order to reduce the costs of integrating the solar plants into the existing power grid, forecasting the energy generated by these plants becomes a key issue. A high-quality forecast will enable grid operators to schedule non-renewable classical capacities (like coal and gas power plants) to compensate for the PV power output variations. Several research projects in Europe, such as the COST Action ES1002 *Weather Intelligence for Renewable Energies,* are testing procedures for obtaining accurate forecasts of PV output power (WIRE, 2014).

* Corresponding author. Fax: +40 256 592383.
E-mail address: marius.paulescu@e-uvt.ro (M. Paulescu).

http://dx.doi.org/10.1016/j.solener.2014.11.003
0038-092X/

Inherently, forecasting the output power of a solar system involves two interrelated issues: (1) predicting the solar resource availability and (2) realistic modeling of the photovoltaic converter response. Nonetheless, the accuracy of forecasting PV plant output power is mainly related to the accuracy of forecasting solar energy.

Different methods for solar energy forecasting are considered, depending on the time horizon. For nowcasting (horizons shorter than 3 h), the forecasts are based on extrapolation of real-time measurements. A good review of these methods can be found in Diagne et al. (2012). The most popular methods used in this area are: Autoregressive Integrated Moving Average (ARIMA) (Reikard, 2009) with their seasonal version sARIMA (Brockwell and Davis, 1991) and Artificial Neural Network (Mellit and Pavan, 2010; Ibeh et al., 2012; Kumar et al., 2013). Beyond the traditional methods, some new type models have been proposed recently (Paulescu et al., 2014). Arguably, neural networks are typical *black-box*-type models. Namely, they show statistical connections between input and output quantities, with no or very little insight into the underlying physical phenomenology. Models like ARIMA are of different kind: they provide insight into the nature of the dynamics through the careful analysis of their coefficients. Nevertheless, their typical usage in practical forecasting of solar radiation is very close to the black-box approach (with automatic model selection, almost no interpretation of the coefficient structure and no critical appraisal of subtle features of the modeled distributions). Note that the common selection among black-box models by using statistical indicators such as the root mean square error or the mean absolute percentage error and/or other global measures leaves only relatively small space for improvements.

In this paper we show that *tailored* statistical models for hourly average solar irradiance forecasting can be used as an attractive alternative to the wide-spread practice of black-box-style modeling and forecasting. Quite substantial forecast quality improvements can be achieved when the statistical models underlying the forecasts are derived in semi-empirical style. There, a mixture of empirical (data-based) and physical approach is used when deriving the structure of a predictor, the distributional form of the statistical model and other features. Several new models tailored in this style are proposed in this paper. The starting point comes from the fact that the PV output power depends in quite complex and time-varying way on the geometry of Sun's apparent movement on the celestial vault and on the tilt and orientation of PV panel surface. Some of the relationships involved here are deterministic while others are related to the statistical characteristics of the irradiance distribution. Both kinds of relationships may be used to build statistical models for forecasting solar irradiance and/or PV plant output power.

The paper is organized as follows. Section 2 presents the data used and gives details on the measurement procedures. New forecasting models are proposed in Section 3. The accuracy of these models is discussed in Section 4. Sensitivity analysis is performed in Sections 4.3 to 4.6, showing the effects of various model parameters upon the prediction quality. The main conclusions and some perspectives are presented in Section 5.

2. Data

In this study we use radiometric and meteorological data recorded on the Solar Platform of the West University of Timisoara (Solar Platform, 2014). The town of Timisoara (latitude 45°46′N, longitude 21°25′E and 85 m altitude) features a temperate continental climate, with warm summer, typical for the whole Pannonian Basin. Its climate classification is *Cfb*, according to the digital Köppen–Geiger world climate classification map based on data from the second half of the 20th century (Kottek et al., 2006).

Global and diffuse solar irradiance on horizontal surface and total solar irradiance on a 45° tilted, South oriented surface are recorded with LP PYRA 02 pyranometers (DeltaOHM, 2014). LP PYRA 02 are first class pyranometers, fully complying with ISO 9060 standards. They also meet all the requirements defined by the World Meteorological Organization. Meteorological data (air temperature, atmospheric pressure, air relative humidity) are recorded by the DeltaOHM HD2001.1 station. All measurements are performed simultaneously, 24 h a day, at equidistant time intervals of 15 s. The sensors are integrated into an acquisition data system based on the National Instruments PXI platform including a PXI-6259 data acquisition board. The LabVIEW software is used for the management and control of the system.

The quality control of solar radiation data is performed in two steps: (1) Passive data quality assurance during data collection by regular maintenance of the instruments (verification of the instruments every morning and cleaning the external dome if necessary, adjusting the shadow ring of the pyranometer for diffuse radiation according to the manufacturer prescription and changing periodically the silica gel cartridge); (2) Active data quality assurance through a routine retrospective data analysis. Collected data are post processed to check in a final quality assessment whether a value is physically feasible. Regarding the global and diffuse irradiance, two tests are performed currently: (1) checking whether global solar irradiance is greater than diffuse irradiance and (2) checking whether clearness index is in the usual range, for instance $(0, 0.8)$. If data are identified as outlier, they are marked for further investigation.

For this paper, we used the measurements performed from January 1, 2009 to December 31, 2010. While the tilted surface irradiance measurements were used for statistical modeling, the horizontal surface measurements were used to calculate state of the sky indicators.

3. Models

The variable of interest here is the hourly average total solar irradiance on tilted surface. Note that, the hourly