Available online at www.sciencedirect.com

SciVerse ScienceDirect

Solar Energy 86 (2012) 2017–2028

SOLAR ENERGY

www.elsevier.com/locate/solener

Assessment of forecasting techniques for solar power production with no exogenous inputs

Hugo T.C. Pedro, Carlos F.M. Coimbra *

Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0411, USA

Received 2 October 2011; received in revised form 14 March 2012; accepted 6 April 2012
Available online 4 May 2012

Communicated by: Associate Editor David Renne

Abstract

We evaluate and compare several forecasting techniques using no exogenous inputs for predicting the solar power output of a 1 MWp, single-axis tracking, photovoltaic power plant operating in Merced, California. The production data used in this work corresponds to hourly averaged power collected from November 2009 to August 2011. Data prior to January 2011 is used to train the several forecasting models for the 1 and 2 h-ahead hourly averaged power output. The methods studied in this work are: Persistent model, Auto-Regressive Integrated Moving Average (ARIMA), k-Nearest-Neighbors (kNNs), Artificial Neural Networks (ANNs), and ANNs optimized by Genetic Algorithms (GAs/ANN). The accuracy of the models is determined by computing error statistics such as mean absolute error (MAE), mean bias error (MBE), and the coefficient of correlation (R^2) for the differences between the forecasted values and the measured values for the period from January to August of 2011. This work also addresses the accuracy of the different methods as a function of the variability of the power output, which depends strongly on seasonal conditions. The findings show that the ANN-based forecasting models perform better than the other forecasting techniques, that substantial improvements can be achieved with a GA optimization of the ANN parameters, and that the accuracy of all models depends strongly on seasonal characteristics of solar variability.

Keywords: Solar forecasting; Solar energy; Regression analysis; Stochastic learning

1. Introduction

One of the critical challenges in transitioning to an energy economy based on renewable resources is to overcome issues of variability, capacity and reliability of non-dispatchable energy resources such as solar, wind or tidal. The variable, and sometimes intermittent, nature of these resources implies substantial challenges for the current modus operandi of power producers, utility companies and independent service operators (ISOs); especially when high market penetration rates (such as the ones now mandated by law in California and other US states) are considered. As of January of 2012, the US Department of Energy lists 29 states with Renewable Portfolio Standards (RPSs) varying from 10% to 33% renewable penetration for 2020 to 40% in 2030. Other US states have similar goals. Many European countries have more aggressive RPS goals. Most of the future renewable generation needed for satisfying these aggressive RPS goals will likely come from variable resources such as solar and wind power.

Although solar energy is clearly the most abundant power resource available to modern societies, the implementation of widespread solar power utilization is so far impeded by its sensitivity to local weather conditions, intra-hour variability, and dawn and dusk ramping rates.

* Corresponding author.
E-mail address: ccoimbra@ucsd.edu (C.F.M. Coimbra).

0038-092X/$ - see front matter
http://dx.doi.org/10.1016/j.solener.2012.04.004

The variability directly affects both capital and operational costs, also contributing to lower capacity factors. Solar forecasting, i.e., the ability to forecast the amount of power produced by solar farms and rooftop installation feeding power substations, has the ability to optimize decision making at the Independent System Operator (ISO) level by allowing corrections to unit commitments and out-of-region trade. Short-term, intra-hour forecasts are relevant for dispatching, regulatory and load following purposes, but the intra-day (especially in 1–6 h ahead horizon) forecasts are critical for power system operators that handle multiple load zones, and trade outside of their territory. In particular, the direct sunlight beam, which is critical for concentrating solar technologies, is much less predictable than the global irradiance, which includes the diffuse component from the sky hemisphere (see e.g., Marquez and Coimbra, 2011), and is also more susceptible to truly intermittent availability due to cloud cover. So in order to characterize the viability of solar production in a given location, detailed variability and forecastability studies become imperative. Fig. 1 depicts the variations of 100 kW to 600 kW (i.e. variations from 10% to 60% of the nominal PV plant peak output) during the diurnal period of 9:00–14:00 in a monthly basis. As expected for California's Central Valley, larger fluctuations occur in the Winter, late Fall and early Spring. The Months of July and August show much smaller variability. However, even during some periods within the sunshine months, sudden changes on the power output can be observed as exemplified by an event in August 28, 2010, in which there was an integrated 60% ramp between 10:00 am and 11:00 am (this event is depicted in Fig. 2).

To understand and predict the variability of the solar resource, many attempts to forecast solar irradiance (the resource) have been presented (Mellit, 2008; Mellit and Pavan, 2010; Marquez and Coimbra, 2011; Elizondo et al., 1994; Mohandes et al., 1998; Hammer et al., 1999; Sfetsos and Coonick, 2000; Paoli et al., 2010; Lara-Fanego et al., 2011), while other researchers have extended their models to power output from PV plants (Picault et al., 2010; Bacher et al., 2009; Chen et al., 2011; Chow et al., 2011; Martín et al., 2010). Artificial Neural Networks (ANNs), Fuzzy Logic (FL) and hybrid systems (GA/ANN, ANN-FL) are well suited to model the stochastic nature of the underlying physical processes that determine solar irradiance at the ground level (and thus the power output of PV installations). Other regression methods often employed to describe complex non-linear atmospheric phenomena include the Auto-Regressive Moving Averages (ARMAs) method, as well as its variations, such as the Auto-Regressive Integrated Moving Averages (ARIMAs) method (Gordon, 2009).

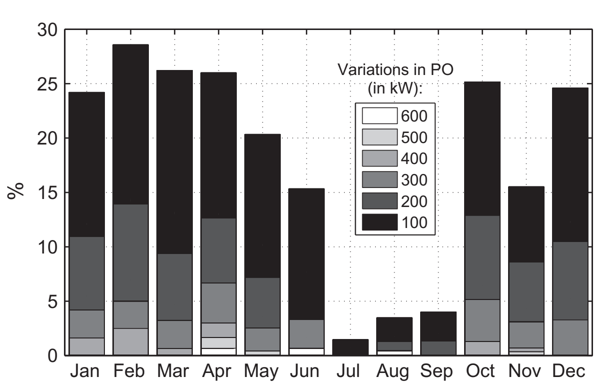

Fig. 1. Percentage of variations in the P of more than: 100 kW, 200 kW, ..., 600 kW, in a monthly basis. Only data in the peak production period between 09:00 and 14:00 was used to generate this figure.

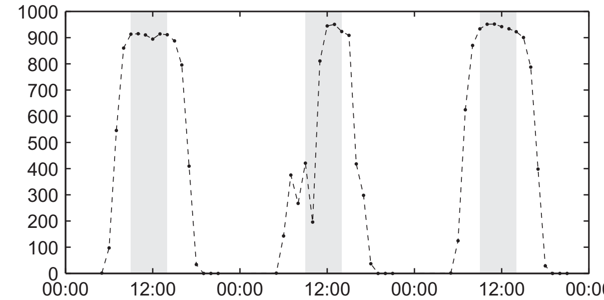

Fig. 2. Power output for the period between 08/27/2010 and 08/29/2010. August 28 exemplifies a day in which there is a sudden ramp up in the power output of more that 600 kW (60% of the nominal peak P) within the period of expected maximum Power output.

In this work, the 1-h averaged data for the 1 MWp PV farm power output (P) collected from November 2009 to August 2011 is used to develop and train several forecasting models for predicting the power output 1 and 2 h-ahead of time, using only the single-axis panels as network of ground sensors. The goal is to study several of the most popular forecasting methodologies and assess their accuracy in order to determine a minimum performance level for which comparison with more complex forecasting methods (using a variety of radiometric and meteorological inputs) should be carried out. In essence, we apply different forecasting engines to a "zero-telemetry" operational data set to evaluate the performance of non-exogenous methodologies, thus creating a baseline for developing more sophisticated.

2. Data

This work uses data collected from a single-axis tracking, polycrystalline photovoltaic, 1 MW peak solar power plant located in Central California (Merced). This solar farm provides between 17% and 20% of the power consumed yearly by the University of California, Merced campus, and is used as a test-bed for solar forecasting and fast demand response studies by our research group. The time period analyzed spans from November 3, 2009 (the first full day of operation of the solar farm) to August 15, 2011. The data points collected from the power plant site correspond to the hourly average of power output (P). Although available to us, additional solar irradiance and weather variables, such as global horizontal irradiance, cloud cover,