Energy Conversion and Management 108 (2016) 539–548

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Smart baseline models for solar irradiation forecasting

R. Urraca *, J. Antonanzas, M. Alia-Martinez, F.J. Martinez-de-Pison, F. Antonanzas-Torres

EDMANS Group, Department of Mechanical Engineering, University of La Rioja, Logroño, Spain

ARTICLE INFO

Article history:
Received 12 September 2015
Accepted 17 November 2015
Available online 8 December 2015

Keywords:
Solar forecasting
Intra-daily solar irradiance
Random forests
Soft computing
Persistence models

ABSTRACT

This work presents a kind of smart baseline models for solar irradiation forecasting, as models are only fed with meteorological records and solar-computed values, easy-to-obtain inputs that facilitate their implementation worldwide. Global horizontal irradiation (GHI) is predicted for horizons of 1 h in a site of Southeast Spain. Two types of approaches are undertaken: fixed models, trained just once with a global database, and moving models, where the training database is updated based on the features of the testing sample. The approaches are implemented with two machine learning algorithms, support vector regression (SVR) and random forest (RFs), along with the classic linear regression and kNN. Besides, genetic algorithms (GAs) are used to automate the training process of fixed models, a task traditionally performed based on the experience or the researcher.

Significant improvements were obtained over the basic persistence methods with both approaches. In the case of moving models, results proved that the best approach to update the calibration set was by computing the Euclidean distance in the principal components space. Results of both approaches were comparable in terms of MAE and forecast skill (s), though slightly superior predictions were obtained with the moving SVR, with a forecast skill ranging from 8% to 23% and a testing MAE ranging from 49 to 64 W/m^2 for the different states of cloudiness. Anyway, both approaches are valid baselines to compare new forecasting models fed with more difficult-to-obtain features, supplementing the classic but naive persistence models.

1. Introduction

The rising share of on-grid photovoltaic technology within national electric markets introduces a new issue in electricity grids control. Other traditional technologies such as hydropower and gas and fuel-fired power, denoted manageable energy sources, are able to adapt generation to consumption in small periods of time. However, solar photovoltaic energy fluctuates depending on available global solar irradiance. Therefore, manageability of this energy source from the controller point of view is approached from irradiance forecasting. Thus, photovoltaic power can be forecasted and fluctuations can be compensated with other technologies. From the independent power producer side, irradiance forecasting leads to a better generation planning, and therefore, to more competitive bids and lower penalizations in case of deviations respect to energy offered [1].

In the last years, a wide range of predictive methods have been proposed to forecast global horizontal irradiation (GHI). The type of technique used mainly depends on the forecasting horizon, which has been broadly categorized in the state-of-the-art into short-term or intra-hour, medium-term or intra-day, and day ahead [2]. The differences between the various models proposed lie in the information available, i.e., the inputs for the model, and on the modeling techniques selected.

For short-term forecasting and in regard to the input variables, models can be classified in satellite and sky imagers based models and purely statistical models. The first try to model cloud motion and from there forecast solar irradiation. They required specific equipment for gathering the images and a subsequent image analysis. For the short-term, they are based on images taken by on-ground sky cameras. Some examples are the works proposed West et al. [3], Pedro and Coimbra [4], and Dambreville et al. [5]. The second group agglutinates all kind of numerical methods that estimate solar irradiation based on historical solar data plus some exogenous variables, mainly meteorological records [6]. Occasionally, other exogenous variables used are topographical values [7] and even the forecasts from other meteorological models [8]. In regard to the technique used, forecasting models are broadly classified into linear and nonlinear. The most implemented linear models are the autoregressive techniques. They are specific techniques for time series based on the importance of temporal

* Corresponding author.
E-mail address: ruben.urraca@unirioja.es (R. Urraca).

http://dx.doi.org/10.1016/j.enconman.2015.11.033
0196-8904/

Nomenclature

AERONET	AErosol RObotic NEtwork	p	atmospheric pressure (Pa)
AI	artificial intelligence	p_0	atmospheric pressure at sea level (Pa)
AM	air mass (–)	PC	principal component
AOD	aerosol optical depth (–)	PCA	principal components analysis
ANN	artificial neural network	PMS	parsimonious model selection
ARIMA	autoregressive integrative moving average model	PT	parameter tuning
ARMA	autoregressive moving average model	PV	photovoltaic
DHI	diffuse horizontal irradiance (W/m^2)	QRF	quantile regression forest
DNI	direct normal irradiance (W/m^2)	R	rainfall (l)
ELM	extreme learning machines	RFs	random forests
ESRA	European Solar Radiation Atlas	RH	relative humidity (%)
F_d	diffuse angular function (–)	RMSE	root mean squared error (W/m^2)
FS	feature selection	RMSE/Avg	normalized RMSE by the average irradiance (–)
GAs	genetic algorithms	s	forecast skill (%)
GDF	generalized degrees of freedom	SIAR	Spanish Agency for Irrigation in Agriculture
GHI	global horizontal irradiance (W/m^2)	SOM	self-organized maps
h	solar hour angle (°)	SVR	support vector regression
I_0	solar constant (W/m^2)	T	temperature (°)
I^{cs}	clear-sky irradiance, DNI or GHI (W/m^2)	T_d	diffuse transmission function (–)
I^{ex}	extraterrestrial irradiance, DNI or GHI (W/m^2)	T_L	linke turbidity factor (–)
k_t	clearness Index (–)	w	water vapor column (cm)
k_{cs}	clear-sky index (–)	WS	wind speed (m/s)
kNN	k-nearest neighbors algorithm	α	solar elevation angle (°)
M	boolean variable of rain (–)	δ_R	rayleigh optical thickness (–)
m	relative optical mass (–)	ϵ	excentricity factor due to the variation in the Sun–Earth distance (–)
MAE	mean absolute error (W/m^2)	θ	solar zenith angle (°)
MAE/Avg	normalized MAE by the average irradiance (–)	ϕ	solar azimuth angle (°)
MBE	mean bias error (W/m^2)		

persistence in solar irradiance, and the most common ones are the Autoregressive Moving Average (ARMA) and the Autoregressive Integrated Moving Average (ARIMA) [9,10]. On the other hand, non-linear models take algorithms from machine learning and artificial intelligence (AI) areas for predicting irradiance. The most used learning technique are artificial neural networks (ANN) [11–13]. Some variations of the basic ANN algorithm have been also tested such as Elman neural networks [14], self-organized maps (SOM) [15], recurrent network [16], bootstrap-ANN [17] and extreme learning machines (ELM) [18]. Other implemented non-linear techniques are the support vector regression (SVR), which have been used for short-term solar irradiance prediction [19,20] and also for PV forecasting [21–23], Ranfom Forest, used by Kratzenberg et al. [24] for correcting NWP estimations, and Quantile regression forests (QRF), used by Almeida et al. [8] for the forecasting of 1 day ahead photovoltaic power and by Brabec et al. [6] for GHI prediction. Moreover, all these predictive techniques have been complemented with different metaheuristics with optimization purposes. Here, Genetic Algorithms (GAs) have been most widely use in the solar forecasting field [25–28].

This study focuses on forecasting 1-h ahead GHI, which falls within the boundary of intra-hour and intra-day techniques. Models were fed with irradiance and meteorological records and some computed solar variables based on the classical analytical equations. These are widely retrievable inputs that ease the replication of the models in most locations, compared to more difficult-to-obtain information such as the aforementioned sky images or meteorological forecasts. Hence, our aim is to deliver a model trained with easy-to-access data as a kind of smart baseline to compare more complex techniques, a task traditionally performed only with the naive persistence models. Within this goal, two types of models are compared. Fixed models, trained just once with the whole available time series, and moving models, where the training set is permanently updated. Two machine learning techniques, Random Forests (RFs) and SVR, and the classical linear regression are used. Besides, an optimization methodology is proposed for selecting relevant features and tuning model parameters based on GAs towards the automation of the modeling process. All these techniques are benchmarked against classic persistence models and evaluated in a site of Southeast Spain with special interest due to the great PV installed capacity in the surroundings.

The remainder of this paper is organized as follows. Data used for training and validating the predictive models is described in Section 2. The methodology used is explained in Section 3. In this section, the different techniques implemented and evaluation parameters are introduced. Besides, the steps for training and validating the predictive models are described in detail. Results obtained are presented in Section 4 and the conclusions drawn are shown in Section 5.

2. Data

The set of meteorological variables used was recorded in Algemesí (39°11′38″N, 0°26′13″W), a station from the Spanish Agency for Irrigation in Agriculture (SIAR) [29] located in the eastern coast of Spain. Variables were obtained on a 30 min resolution from 01/01/2013 to 31/12/2013. Variables recorded were GHI, rainfall (R), wind speed (WS), temperature (T) and relative humidity (RH) (Table 1). Temperatures were recorded with *Vaisala HMP45-Pt1000 IEC 751 1/3 Class B* sensors with tolerance of 0.3 °C. Relative humidity was measured with *Vaisala HMP45-Humicap 180* sensor with tolerance of ±2% for 0–90% and ±3% for 90–100%. Rainfall was recorded with *Campbell Scientific ARG100* sensor with ±2%. Wind speed was measured with *Young 05103* sensors with