International Journal of Forecasting 34 (2018) 529–547

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data

Mathieu David [a,*], Mazorra Aguiar Luis [b], Philippe Lauret [a]

[a] University of La Réunion - PIMENT Saint-Pierre, Réunion
[b] University Institute for Intelligent Systems and Numerical Applications in Engineering, Las Palmas de Gran Canaria, Spain

ARTICLE INFO

Keywords:
Solar forecasting
Time series
Very short term horizons
Probabilistic forecasting
Prediction intervals

ABSTRACT

Accurate solar forecasts are necessary to improve the integration of solar renewables into the energy grid. In recent years, numerous methods have been developed for predicting the solar irradiance or the output of solar renewables. By definition, a forecast is uncertain. Thus, the models developed predict the mean and the associated uncertainty. Comparisons are therefore necessary and useful for assessing the skill and accuracy of these new methods in the field of solar energy.

The aim of this paper is to present a comparison of various models that provide probabilistic forecasts of the solar irradiance within a very strict framework. Indeed, we consider focusing on intraday forecasts, with lead times ranging from 1 to 6 h. The models selected use only endogenous inputs for generating the forecasts. In other words, the only inputs of the models are the past solar irradiance data. In this context, the most common way of generating the forecasts is to combine point forecasting methods with probabilistic approaches in order to provide prediction intervals for the solar irradiance forecasts. For this task, we selected from the literature three point forecasting models (recursive autoregressive and moving average (ARMA), coupled autoregressive and dynamical system (CARDS), and neural network (NN)), and seven methods for assessing the distribution of their error (linear model in quantile regression (LMQR), weighted quantile regression (WQR), quantile regression neural network (QRNN), recursive generalized autoregressive conditional heteroskedasticity (GARCHrls), sieve bootstrap (SB), quantile regression forest (QRF), and gradient boosting decision trees (GBDT)), leading to a comparison of 20 combinations of models.

None of the model combinations clearly outperform the others; nevertheless, some trends emerge from the comparison. First, the use of the clear sky index ensures the accuracy of the forecasts. This derived parameter permits time series to be deseasonalized with missing data, and is also a good explanatory variable of the distribution of the forecasting errors. Second, regardless of the point forecasting method used, linear models in quantile regression, weighted quantile regression and gradient boosting decision trees are able to forecast the prediction intervals accurately.

1. Introduction

Forecasts of the power output of solar renewables are required in order to increase their penetration rate into electricity grids, and also to ensure the security of the supply–demand balance. Indeed, accurate forecasts allow a

* Corresponding author.

E-mail address: mathieu.david@univ-reunion.fr (M. David).

https://doi.org/10.1016/j.ijforecast.2018.02.003

better scheduling of the energy resources and a better operation of the units' commitment. The recent development of grid-connected storage associated with intermittent renewables (solar, wind and wave) also requires power forecasts in order to optimize their operational management (Haessig, Multon, Ben Ahmed, Lascaud, & Bondon, 2015; Hanna, Kleissl, Nottrott, & Ferry, 2014; Hernández-Torres, Bridier, David, Lauret, & Ardiale, 2015). In the case of solar renewables, the power output is related directly to the level of solar irradiance received. Thus, forecasting the solar irradiance is the key to PV power prediction.

In the context of the kind of decision-making that grid operators must face, a point forecast plus a prediction interval adds real value. Indeed, it is a way of judging the reliability of the forecasts. In the field of solar energy, there are numerous studies (for example Kleissl, 2013; Lauret, Voyant, Soubdhan, David, & Poggi, 2015; Lorenz, Hurka, Heinemann, & Beyer, 2009; Perez et al., 2013) dedicated to models of point forecasts, also called deterministic forecasts. By definition, a forecast is doubtful, and therefore, it is also necessary to know the level of uncertainty associated with a forecast. The development of models that generate probabilistic information about forecasts of solar energy is relatively recent. The existing models can be classified according to their lead-times.

For forecasting horizons of longer than six hours and up to several days ahead, models based on numerical weather predictions (NWP) are the most suitable. Several works have proposed some form of post-treatment of the deterministic forecasts provided by NWP models in order to assess the prediction intervals. Lorenz et al. (2009) were the first to use this approach. They assumed a Gaussian distribution of the forecast error of the global horizontal solar irradiance (GHI) provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), and fitted a simple linear regression in order to derive the standard deviation from the predicted sky conditions. More recently, the GEFCom 2014 (Hong et al., 2016) required the competitors to generate probabilistic forecasts of three solar farms in Australia. The aim of the competition was to compute 99 quantiles of the power output of the PV fields on a rolling basis for 24 h ahead. Twelve weather variables from the ECMWF were made available to the participants and were used as inputs to their models. The main approaches used during the competition were multiple quantile regression (Juban, Ohlsson, Maasoumy, Poirier, & Kolter, 2016), the gradient boosting technique, K-nearest neighbors (Huang & Perry, 2016), and the quantile regression forest combined with gradient boosting decision trees (Nagy, Barta, Kazi, Borbély, & Simon, 2016). Another original method was developed by Alessandrini, Delle Monache, Sperati, and Cervone (2015) for post-processing the deterministic forecasts of the Regional Atmospheric Modeling System (RAMS; see also Pielke et al., 1992). They used an analog ensemble approach applied to a set of forecasted weather variables (GHI, cloud cover, air temperature, etc.) to estimate the quantiles of the error on the solar power of three PV systems in Italy.

Some NWP models also provide ensemble forecasts; i.e., a perturbed set of forecasts is computed by changing the initial conditions of the control run slightly and modeling any unresolved phenomena (Leutbecher & Palmer, 2008). An ensemble forecast system allows the uncertainties of the prediction scheme to be represented. Sperati, Alessandrini, and Delle Monache (2016) proposed using the ensemble prediction system (EPS) provided by the ECMWF in order to compute probabilistic forecasts for three PV farms in Italy. Their method involves two steps. First, a neural network (NN) is used to derive the power of the PV farms from the 51 members of the EPS: one forecast from the control run and 50 perturbed forecasts. Second, a post treatment of the 51 forecasts is performed for generating the quantiles. In their article, Sperati et al. (2016) proposed two methods for this last step. The first, the variance deficit (VD) method, adjusts the spread of the forecasts. The second, the ensemble model output statistics (EMOS), which was initially designed by Gneiting, Raftery, Westveld, and Goldman (2005), is used to minimize the continuous ranked probability score (CRPS; see Section 3) of the probabilistic forecasts. Following the same approach, Zamo, Mestre, Arbogast, and Pannekoucke (2014) proposed a comparison of two different models of quantile regression applied to Météo France's ensemble NWP system, named PEARP. The PEARP's ensemble forecasts consist of one control run and 34 perturbed members. Two probabilistic forecasts are computed from the quantile regression models: the first using only the control run and the second by averaging the quantiles obtained from the 34 perturbed members. In addition, the authors also detailed a method of calibrating the spread of the probabilistic forecasts.

For lead-times in the range of two to six hours, models that use satellite images as inputs are the most suitable. To the best of our knowledge, no work has yet been published regarding probabilistic solar forecasts derived from satellite images.

The last category concerns forecasts with horizons ranging from several seconds to several hours. These very short-term forecasts are useful for the real-time operation of the grid or of storage systems. Two methodologies have been developed for forecasting at these very short-term horizons: sky imagery from the ground and time series models (also called statistical models). Regarding sky imagery, only one work has been published very recently regarding their use for assessing the uncertainty of solar forecasts (Chu & Coimbra, 2017). However, several articles have dealt with the use of time series models for producing probabilistic forecasts. Iversen, Morales, Møller, and Madsen (2014) propose a time series model that integrates deterministic and probabilistic forecasts simultaneously. The methodology is based on stochastic differential equations (SDE), and was originally designed for forecasting wind power, but was later tested by using it to forecast the GHI of a weather station in Denmark. Similarly to the post-processing of the NWP models, most of the studies of time series models start by using a point forecast as the input, then assess the prediction intervals in a second step. Reproducing this scheme in order to forecast the output of a PV system, Bacher, Madsen, and Nielsen (2009) used a weighted quantile regression based on a Gaussian kernel to post-process deterministic forecasts produced using an autoregressive model. Unfortunately, though, they provided no evaluation of the quantiles generated. Following a similar approach,