Solar Energy 134 (2016) 132–146

Contents lists available at ScienceDirect

Solar Energy

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

Multi-Model Ensemble for day ahead prediction of photovoltaic power generation

Marco Pierro [a,e], Francesco Bucci [a], Matteo De Felice [d], Enrico Maggioni [c], David Moser [e], Alessandro Perotto [c], Francesco Spada [c], Cristina Cornaro [a,b,*]

[a] Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy
[b] CHOSE, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy
[c] Ideam Srl, via Frova 34, Cinisello Balsamo, Italy
[d] Casaccia R.C., ENEA Climate Modelling Laboratory, Rome, Italy
[e] Institute for Renewable Energy, EURAC Research, Viale Druso, 1, 39100 Bolzano, Italy

ARTICLE INFO

Article history:
Received 18 November 2015
Received in revised form 14 March 2016
Accepted 26 April 2016
Available online 9 May 2016

Keywords:
Photovoltaics power forecasting
Day-ahead PV forecasts
Ensemble prediction
Probabilistic forecast
Error metrics
Data-driven models
NWP models

ABSTRACT

The aim of the paper is to compare several data-driven models using different Numerical Weather Prediction (NWP) input data and then to build up an outperforming Multi-Model Ensemble (MME) and its prediction intervals. Statistic, stochastic and hybrid machine-learning algorithms were developed and the NWP data from IFS and WRF models were used as input. It was found that the same machine learning algorithm differs in performance using as input NWP data with comparable accuracy. This apparent inconsistency depends on the capability of the machine learning model to correct the bias error of the input data. The stochastic and the hybrid model using the same WRF input, as well as the stochastic and the non-linear statistic models using the same IFS input, produce very similar results. The MME resulting from the averaging of the best data-driven forecasts, improves the accuracy of the outperforming member of the ensemble, bringing the skill score from 42% to 46%. To reach this performance, the ensemble should include forecasts with similar accuracy but generated with the higher variety of different data-driven technology and NWP input. The new performance metrics defined in the paper help to explain the reasons behind the different models performance.

1. Introduction

Large share of PV power introduces new challenges for the stability of the electrical grid, both at the local and national level.

In particular, it introduces into the electric load a stochastic variability dependent on meteorological conditions (EU-PVTP Working Group on grid integration, 2015). Indeed the residual load that should be covered by the fossil or non variable renewable energy generation results from the difference between the electric consumption and the distributed PV production. The residual load, in case of high PV power injection, varies on the daily time scale between two extreme limits. During very overcast days, it is almost equal to the electric consumption that slowly increases until the night peak. During very clear sky days it is minimum at noon and it rapidly increases in the evening when the reduction of PV generation is added to the growing demand. Thus, in case of high PV generation higher secondary reserves and ready supply are needed to overcome the unpredictability and variability of the residual load. Moreover, according to the Italian market rules, this large amount of energy needed to ensure electrical balancing should be purchased at higher price on Real-Time Energy Market.

Mid-term (intra-day and day-ahead) PV generation forecasts could mitigate the effects of high PV power injection into the electricity grid, both on grid management and on the energy market. On one hand it could be used for load following to control voltage and frequency instability and for transmission scheduling to reduce the secondary reserve. On the other hand, it could be employed for a better match between the intra-day and day-ahead market commitment and the real PV production, reducing the energy unbalancing costs.

In 2014, PV penetration was estimated to exceed the 1% mark for 19 countries with Italy leading at 7.9% followed by Greece and Germany at 7.6% and 7%. Different IEA and EPIA scenarios

* Corresponding author at: Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy.

E-mail addresses: marco.pierro@gmail.com (M. Pierro), frabucci@gmail.com (F. Bucci), matteo.defelice@enea.it (M. De Felice), enrico.maggioni@ideamweb.com (E. Maggioni), david.moser@eurac.edu (D. Moser), alessandro.perotto@ideamweb.com (A. Perotto), francesco.spada@ideamweb.com (F. Spada), cornaro@uniroma2.it (C. Cornaro).

http://dx.doi.org/10.1016/j.solener.2016.04.040
0038-092X/

predict for 2030 a PV generation of 1.5–5% of the world electric consumption and of 10–25% of the UE27 electric demand (IEA, 2014a,b). For this reason, the site and regional day ahead forecast of PV generation is now mandatory in many European and non-European countries (Italy, Germany, Spain, Romania, USA, Japan, etc.).

For day-ahead forecasts (24/72 h horizon), Numerical Weather Prediction (NWP) data should be employed to obtain an acceptable accuracy level, as for intra-hour and intra-day forecast, the use of full sky images and satellite data is essential. The NWP data are generated by global or mesoscale simulation models able to provide the numerical integration of the coupled differential equations describing the dynamics of the atmosphere and radiation transport mechanisms (Perez et al., 2013). Moreover, these data are usually corrected by post-processing algorithms called Model Output Statistics (MOS) that use past ground measurements to partially remove systematic errors (Perez et al., 2007; Lorenz et al., 2009a; Pierro et al., 2015a).

In the last five years data-driven approach was extensively tested for PV power generation forecast on the horizon of 24/72 h. This approach involves a wide range of black-box techniques (e.g. machine learning, time-series methods) that can be built using past measurements. They are statistic or stochastic models that try to reconstruct relationships between input and output data sets. They do not require knowledge of the physical laws describing the phenomenon and none or very little plant information should be provided. Once trained, these algorithms can use a variety of NWP products to directly provide the PV generation forecast. Hybrid models could be obtained using different models in series. On the contrary, the combination of different predictions obtained through various models leads to a Multi-Model Ensemble (see Fig. 1).

A variety of machine-learning models using different techniques were developed by various authors. For site day-ahead power forecasts, Yona et al. (2008) implemented models based on a Multilayer Perceptron (MLPNN), Radial Basis function (RBFNN) and Recurrent Neural Networks (RNN). Chen et al. (2011) and Tao et al. (2010) used RBFNN and Non linear autoregressive exogenous NN (NARX) while Wang et al. (2011) coupled MLPNN with the Gray Model (GM). Mellit et al. (2014) introduced adaptive feed-forward back-propagation network (AFFNN) for Short-term forecasting. For site and regional forecast, Zamo et al. (2014a) compared the accuracy of many different techniques: Bagging, Random forest, Boosting, Support Vector Machine (SVM) and Generalized Additive Model (GAM), concluding that the Random forest was the outperforming model. For regional forecast, da Silva Fonseca Junior et al. (2014) explored an interesting technique based on Support Vector Regression (SVR) coupled with a Principal Component Analysis (PCA) pre-processing.

Although literature and methodology is established for the separate models, a Multi-Model Ensemble approach is not fully explored and it is one of the main goals of the paper.

More recently, a probabilistic approach has been adopted not only for wind forecasting but also for PV power predictions. This method is focused on informing about the distribution of potential events through a set of conditional probability density functions or ensemble of a statistically relevant number of alternative forecast obtained through one or more models. The probabilistic approach provides at the same time the best forecast and its prediction intervals. This approach based on different statistical regression methods was explored by Zamo et al. (2014b) and Almeida et al. (2015).

An overview on PV power forecast techniques can be found in Paulescu et al. (2012), Kleissl (2013) and “Photovoltaic and Solar Forecasting: State of the Art” (IEA, 2013).

The aim of the paper is to compare several data-driven models using different NWP input and then to build an outperforming Multi-Model Ensemble. In particular, relevant questions are investigated: using different irradiance predictions (NWP) as input for the same model, does the outperforming NWP produce also the best power forecast? On the contrary, using the same NWP input for different models how much could the accuracy change? How can an outperforming Multi-Model Ensemble be constructed and which is the improvement with respect to the best forecast of the ensemble? How can prediction intervals be estimated?

Here several models are developed and applied: a time-series seasonal model, a Support Vector Machine and two different models both based on ensemble of Artificial Neural Networks. Two different Numerical Weather Prediction data (NWP) were used as model input. The first NWP data were obtained by the mesoscale Weather Research and Forecasting model (WRF) and the second comes from the global Integrated Forecasting System model (IFS) used by the ECMWF. The WRF irradiance forecast was refined with a Model Output Statistic post-processing algorithm (MOSRH). The models performance were analyzed and new metrics for accuracy evaluation were introduced to better understand the impact of the different NWP on the models performance. Four years of monitored weather and production data from a 662 kWp Cadmium Telluride PV plant, located in Bolzano (Italy), were employed to train and test the models. Finally a Multi-Model Ensemble was constructed and its prediction intervals were calculated.

In Section 2, the experimental and NWP data are presented. In Section 3 the Clear Sky Performance index definition, introduced

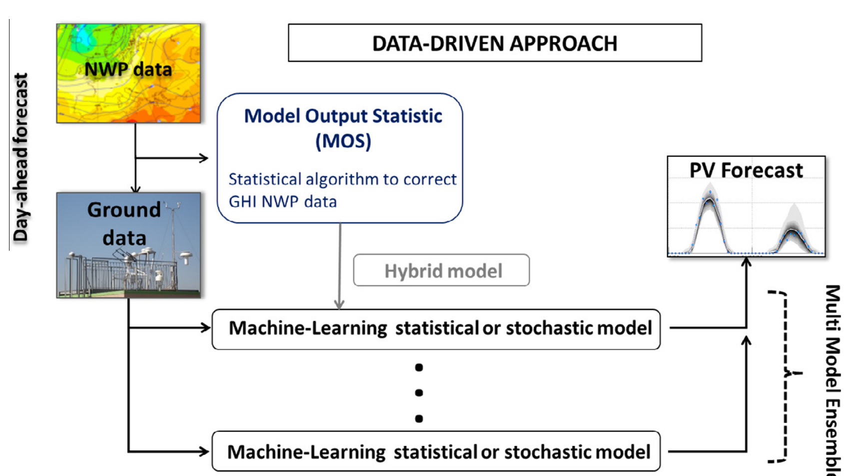

Fig. 1. Schematics of different data-driven approaches for day ahead PV power forecast.