Available online at www.sciencedirect.com

ScienceDirect

Solar Energy 105 (2014) 804–816

SOLAR ENERGY

www.elsevier.com/locate/solener

A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production

M. Zamo [a,1], O. Mestre [a,*], P. Arbogast [b], O. Pannekoucke [b]

[a] *Météo-France, Direction de la Production, 42 av. Coriolis, 31057 Toulouse cedex, France*
[b] *Météo France, CNRM/GMAP/RECYF, 42 av. Coriolis, 31057 Toulouse cedex, France*

Received 7 October 2013; received in revised form 14 March 2014; accepted 17 March 2014
Available online 28 April 2014

Communicated by: Associate Editor David Renne

Abstract

This pair of articles presents the results of a study about forecasting photovoltaic (PV) electricity production for some power plants in mainland France. Forecasts are built with statistical methods exploiting outputs from numerical weather prediction (NWP) models. Contrary to most other studies, forecasts are built without using technical information on the power plants. In each article, several statistical methods are used to build forecast models and their performance is compared by means of adequate scores. When a best forecast emerges, its characteristics are then further assessed in order to get a deeper insight of its merits and flaws. The robustness of the results are evaluated with an intense use of cross-validation.

This article deals with probabilistic forecasts of daily production 2 days ahead. By "probabilistic" we mean that our forecast models yield some quantiles of the expected production's probability distribution. Whereas probabilistic forecasts are quite usual in wind power forecasting, they are more unusual in the field of PV production forecasting. We show that our eight forecasts always perform better than a simple climatological reference forecast. Nevertheless, no forecast model clearly dominates the other, whatever the studied power plant. A post-processing technique aiming at improving forecasts may in fact decrease their performance. Finally, we use outputs from an ensemble NWP model in order to add information about the meteorological forecast uncertainty. Sometimes this ensemble improves the performance of forecasts, but this is not always true.

Keywords: Photovoltaic power forecasting; Benchmark; Statistical methods; Numerical weather prediction; Quantile regression; Ensemble numerical weather prediction

1. Introduction

In Zamo et al. (2014), we described how we forecast hourly electricity photovoltaic (PV) production one day ahead to answer needs of electricity grid managers, energy traders and producers. The forecast consisted in the mean hourly PV production at each day-time hour at some power plants in two French counties.

* Corresponding author. Permanent address: Météo France, DP/DPrévi/COMPAS, 42 av. Coriolis, 31057 Toulouse cedex 01, France. Tel.: +33 5 61 07 86 33.

E-mail addresses: michael.zamo@meteo.fr (M. Zamo), olivier.mestre@meteo.fr (O. Mestre), philippe.arbogast@meteo.fr (P. Arbogast), olivier.pannekoucke@meteo.fr (O. Pannekoucke).

[1] Principal corresponding author. Tel.: +33 5 61 07 86 36.

http://dx.doi.org/10.1016/j.solener.2014.03.026
0038-092X/

This current article deals with a farther lead time and a different purpose. We seek to forecast daily electricity PV production with an anticipation of two days. The purpose is to provide electricity producers with an accurate forecast to manage maintenance operations of their facilities in a cost-effective way. Indeed a skillful forecast should allow producers to efficiently plan those operations only when electricity production is expected to be low, in order to reduce the financial loss caused by stopping production during maintenance.

Since maintenance operations are likely to last the entire day, we forecast the expected daily production at chosen PV power plants. Furthermore, since meteorological predictions tend to be less precise with increasing lead times, the production forecast takes a probabilistic form. By "probabilistic", we mean the forecast is composed of several quantiles of the forecast production's probability distribution. One final element of our forecasting strategy is to use predictors from Météo France's ensemble numerical weather prediction (NWP) system, PEARP. PEARP is an ensemble NWP system composed of a set of 35 forecasts computed from different atmospheric initial states and with different physics. Ensemble forecasts aim at evaluating the forecast uncertainty originating from several error sources, such as initial conditions or uncertainty from unresolved physics in the case of PEARP. By taking predictors from PEARP, our goal is to assess the usefulness of incorporating such informations on meteorological uncertainty in daily production forecasts.

As stated in Zamo et al. (2014), since Statistics offers many methods to our purpose, we compare several statistical methods with appropriate scores and graphs. Two quantile regression methods are used (linear model in quantile regression and quantile regression forest) optionally taking into account PEARP information. Finally, we correct each forecast with a post-processing technique or let the forecast uncorrected. This makes a total of 8 different forecasting strategies to compare. A climatological forecast is also built, as a reference forecast. Lastly, three sets of quantiles are forecast, with a different number of quantiles. This aims at assessing the influence of the number of forecast quantiles on the predictive performance. During this benchmark, a systematic use of cross-validation is done in order to evaluate the robustness of our results.

The article proceeds as follow. Section 2 describes our main goal and the data. Section 3 details our modeling strategy. Relevant scores and assessment tools for probabilistic forecast are also described here. Section 4 presents the results of the comparative test. We conclude in Section 5 with a reminder of those main results along with some perspectives of possible improvements.

2. Purpose of the study and available data

We aim at building a forecast model of the daily production of photovoltaic electricity 2 days ahead at some power plants in mainland France. Most studies in the field of PV production forecasting rely on technical information about PV panels and/or modeling of direct and diffuse solar irradiation. This approach is presented in Heimo et al. (2012) and used in Bracale et al. (2013) for a probabilistic PV production forecast. Since we have no access to such information and models and in order to explore different solutions, we choose a direct modeling of the production with statistical methods. That is, we try to learn the statistical link between some carefully chosen meteorological predictors and the observed electricity production from a production archive. We briefly describe hereafter these two sets of data. This approach may be interesting when power plants are poorly documented and the usual approach cannot thus be employed. On the other hand, this requires a sufficient amount of past measured productions.

2.1. PV production data

The original PV production data already used in Zamo et al. (2014) are at a 10 min temporal resolution, for 28 power plants in two French counties covering about 7000 km^2 each. We will follow the same convention as in the first article by naming those two counties CountyA and CountyB. Individual power plants for CountyX are named CountyXxx with xx ranging from 1 to 18 for CountyA and from 1 to 10 for CountyB.

The forecasts we compare are numerous and require a lot of computation time. Furthermore the robustness of our results is assessed through cross-validation, which requires much more time. Thus only 5 power plants in each county are used for this part of the study. We give at the end of Section 3.1 some figures about the required computation times. The chosen power plants are evenly spread over each county. For CountyB, one power plant is specifically chosen because of its location by the seaside. For the quite hilly CountyA, some power plants are located on the reliefs, others in valleys.

For our purpose, the 10-min PV production were summed up on a daily basis between 6 and 18 UTC. This lapse roughly corresponds to day-time and working-time in mainland France. It is also compatible with the availability of predictors from PEARP (see the following subsection).

Further information about the pre-processing applied to the PV production data and the two counties can be found in Zamo et al. (2014).

2.2. Predictors

As stated above, most of our predictors come from Météo France's ensemble NWP system, named PEARP (Prévision d'Ensemble ARPege). An ensemble forecast is a set of several forecasts (or "members") built in such a way as to represent part of the uncertainty in meteorological predictions. Leutbecher and Palmer (2008) gives a good introduction to ensemble forecasting. PEARP is a