Available online at www.sciencedirect.com

ScienceDirect

Solar Energy 105 (2014) 792–803

SOLAR ENERGY

www.elsevier.com/locate/solener

A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly 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 01, France*

Received 22 February 2013; received in revised form 10 October 2013; accepted 4 December 2013
Available online 3 May 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.

The companion article Zamo et al. (2014) will deal 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.

This article deals with forecasting hourly PV production for the next day in a deterministic way, which means the mean expectable hourly PV production is forecast for each day-time hour. In this part of our study, predictors comes from ARPEGE, Météo France’s deterministic NWP model. Our best model is very reliable and performs well, even compared to best expectable performances computed while using observations as predictors. It also points at the interest of using predictors based on human forecasters’ experience.

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

1. Introduction

Whereas solar irradiation is an inexhaustible source of energy, it is also intrinsically fluctuating in a way human beings cannot control. This intermittent nature poses a great challenge to electricity producers and electric grid operators. A challenging task is to provide accurate photovoltaic (PV) energy forecasts to actors of PV production and transport. Furthermore, in order to meet these actors’

* 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. Permanent address: Météo France, DP/DPrévi/COMPAS, 42 av. Coriolis, 31057 Toulouse Cedex 01, France. Tel.: +33 5 61 07 86 36.

http://dx.doi.org/10.1016/j.solener.2013.12.006
0038-092X/

needs, one's forecast must take into account the lead times and temporal and spatial granularities required by the end-users. Following Kostylev and Pavlovski (2011), we can categorize these needs as:

- intra-hour: a lead-time from 15 min to 2 h with a time step of about 1 min, to forecast ramps and high-frequency variability in electricity production;
- hour ahead: hourly granularity with a maximum look-ahead time of 6 h, to follow power load;
- day ahead: one to three days ahead for hourly production, to meet needs in unit commitment, transmission management and trading;
- medium-term: from 1 week to 2 months lead-time and daily production, for hedging, planning and asset optimization;
- long-term: an anticipation of one to several years for monthly or annual production, to select potentially interesting sites and assess resources.

Since electricity production from solar irradiation is rapidly increasing worldwide, these needs and the associated economical, technical and operational concerns grow in a similar way. This motivates several researches such as Action COST Wire presented in Heimo et al. (2010), and in the field of solar electricity production forecasting through statistical models, such as Bacher et al. (2009), Brabec et al. (2011), Lorenz et al. (2012, 2011), Mellit and Pavan (2010), Panagopoulos et al. (2012), Pelland et al. (2011), to cite just a few (more references can be found in Espinar et al. (2010)). We are interested in this study in two aspects of this forecast challenge: (1) building a statistical model to forecast hourly electricity production today for tomorrow, and (2) forecasting daily electricity production for the day after. To do so our predictors consist in or derive from outputs from respectively the deterministic NWP model ARPEGE (Courtier et al., 1991) in this article and the ensemble forecast model PEARP in Zamo et al. (2014). Whatever the lead time, we do not forecast solar irradiation as in many other studies but we try to directly forecast PV production. This last choice results from a lack of technical information about PV panels and irradiation model, that may not be always available.

In this first article, we describe how we forecast hourly PV production one day ahead to answer needs of electricity grid managers, energy traders and producers. The forecast consists in the value of the expected hourly PV production at each day-time hour. Energy traders can benefit from these forecasts on the energy markets. Also some producers may be submitted to financial incentives or penalties for providing an accurate estimate of their expected PV production for the next day. Indeed, fluctuations of solar power can result in an imbalance between electricity production and demand and thus can threaten the stability of the electricity grid. As a consequence, electricity grid operators, in charge of this stability, are also interested in 1-day ahead production forecasts. Electric grid operators are also interested in forecasts of the total production over an extended area. Consequently we compare two possible forecasting strategies of such integrated production: forecasting the production for each individual power plant and summing up over the target area, or forecasting directly the total production of the area.

Zamo et al. (2014) will deal with a farther lead time: we will seek to forecast daily electricity production with an anticipation of two days. Our forecast is then probabilistic, *i.e.* composed of several quantiles of the forecast production's distribution.

In each of those two parts of our study we proceed with a benchmark of several statistical forecast methods to choose the most skillful, according to some objective criterion. Then if a best forecast clearly dominates, it is more deeply studied. We also make an intensive use of cross-validation to assess the uncertainty and robustness of our results. This is rarely done in other studies that very often draw conclusions with only one test sample.

This article presents our results in forecasting hourly electricity production for a lead time between 28 and 45 h. Section 2 describes our main goal and the data. Section 3 is about our modelisation strategy. Section 4 explores features concerning the best model we retained at the issue of our benchmarking procedure. We compare in Section 5 two strategies to forecast the total production of some area. Section 6 concludes with a reminder of the main results and some perspectives of possible improvements.

2. Purpose of the study and available data

We aim at building a forecast model of the hourly production of photovoltaic electricity for the next day. Other studies rely on the temporal auto-correlation of electricity production to optimize a forecast model based on time-series analysis, as in Bacher et al. (2009). Others include preliminary steps to model the PV panel temperature and/or diffuse *vs.* direct components of solar irradiation as in Pelland et al. (2011). The latter solution requires technical information about the technology used in the PV systems and models of solar irradiation. When such information or irradiation models are not available as it is the case for us, we have to rely on different solutions. Consequently we choose a direct modelisation of the production with statistical methods. In order to do so, several regression methods are used to learn the statistical link between some carefully chosen meteorological predictors and the observed electricity production. The observed electricity production comes from a production archive and predictors consist in outputs of a NWP model. 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.