Solar Energy 150 (2017) 408–422

Contents lists available at ScienceDirect

Solar Energy

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

Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric

Loïc Vallance [a,b,*], Bruno Charbonnier [b], Nicolas Paul [b], Stéphanie Dubost [b], Philippe Blanc [a]

[a] MINES ParisTech, PSL Research University, Centre Observation, Impacts, Energy (O.I.E.), CS 10207, F-06904 Sophia-Antipolis, France
[b] EDF R&D, 6 quai Watier, BP 49, 78401 Chatou, France

ARTICLE INFO

Article history:
Received 10 March 2017
Received in revised form 11 April 2017
Accepted 25 April 2017

Keywords:
Solar irradiance
Forecast
Metric
Performance
Classification

ABSTRACT

The temporal variability of the solar resource – which occurs at different time scales – is a major concern regarding its impact on the grid and photovoltaic (PV) power plants. To handle the increasing penetration rate of solar energy, grid managers require accurate forecasts of incident solar irradiance. This work discusses standard and new procedures to assess the quality of solar forecasting models and highlights the limitations of some widely used metrics in concrete situations. The paper recommends practices for characterizing the quality of a forecast, i.e. a quality control to reject any suspicious data; classifications to relate each performance criterion to the nature of the solar irradiance variability; and a wide selection of established and new metrics. In this context, two new metrics are notably proposed in order to obtain more complete information about the performance of a forecasting method. This article aims to make a step towards standardizing metrics for solar forecasting by taking into account some of the new metrics recently presented in the literature.

1. Introduction

The emergence of renewable energy resources raises many questions regarding their operational implementation. The rising share of wind and solar energy in global electricity production requires adapting technical means to cope with their uncontrollable nature. Conventional energy sources like fossil and nuclear have a stable and controllable output power which make them easily adaptable to electrical grid requirements (Lorenz and Heinemann, 2012). On the contrary, the weather dependence of wind and solar energies makes them highly variable in both time and space. This "intermittent" nature of solar energy is one of the main issues that hinder the expansion of its penetration rate (Rodriguez, 2010). The solution that is generally considered – and currently under study – is to forecast its production at different time scales in order to rebalance global production to match global consumption. Solar forecasting can also serve load-shifting methods where electricity storage aims to reduce the mismatch between peak demand and renewable power supply (Denholm and Margolis, 2007; International Energy Agency (IEA), 2014; Kaur et al., 2016; Lorenz and Heinemann, 2012). For these reasons, in recent years the focus has been on improving the quality of existing forecasts and developing new ways of extending the range of forecast horizons. An accurate assessment of forecast performance is thereby required to demonstrate improvements of the resulting forecast.

Some conventional statistical metrics are widely used by the solar forecasting community because they give an overview of the global performance of a given forecast. Researchers mostly use the root mean square error (RMSE) and mean bias error (MBE) to characterize and validate forecasting methods. However, interpretations of the scores of these metrics should always be considered in context. Indeed, these scores depend on various factors that make inter-site comparison difficult if not impossible, e.g. spatial and temporal resolution of the data being compared, the months used in the comparison, the percentage of clear-sky days in the set of data, etc. In this article, we identify some limitations of using these metrics and suggest a contextualization of forecast validation results. In order to complete the partial information provided by statistical metrics, we describe some recent metrics and suggest an alternative way of characterizing the quality of a forecasting method.

Section 2 describes the methods used in this article. Then, Section 3 establishes a list of standard metrics and gives concrete examples of days on which these kinds of metrics fail to fully describe the performance of a forecasting method. Section 3 is also dedicated to presenting two new emerging kinds of metrics, i.e. the

* Corresponding author.
E-mail address: loic.vallance@mines-paristech.fr (L. Vallance).

http://dx.doi.org/10.1016/j.solener.2017.04.064

Temporal Distortion Index (TDI) and ramp metrics. Two metrics are then proposed with their utilization framework in Section 4, describing behaviors that are particularly interesting for local short-term forecasts. The first metric quantifies the temporal alignment of the forecast and the second focuses on ramp tracking. Section 5 is devoted to the proposed procedure that aims to assess the quality of a forecasting method, including a quality check, some classifications and the use of suggested metrics. Finally the analysis and comments on the results of three reference methods can be found in Section 6.

2. Reference methods

For the sake of illustration, some solar irradiance forecasting methods are used here for comparison:

- The persistence method, which is defined for solar forecasting by:

$$\hat{I}(t+\Delta t) = k_c(t) I_{clr}(t+\Delta t) \tag{1}$$

where $\hat{I}$ corresponds to the forecasted irradiance value, Δt is the forecast horizon of the persistence, and k_c is the clear-sky index, defined by:

$$k_c(t) = \frac{I(t)}{I_{clr}(t)} \tag{2}$$

where I denotes a measured value and I_{clr} is a clear sky model. In this paper the persistence method uses in-situ 15-min average irradiance measurements and the corresponding ESRA (European Solar Radiation Atlas) clear-sky model. The Linke turbidity required by this model comes from the worldwide climatological database of monthly means of Linke turbidity provided by Remund et al. (2003). A wide variety of clear-sky models are proposed in the literature and different articles compare their skills (Badescu et al., 2013; Engerer and Mills, 2015; Gueymard, 2012; Ineichen, 2006; Inman et al., 2013). For this study, we chose the ESRA model because it is one of the most commonly used and robust methods according to these references.

- A Numerical Weather Prediction model from the European Centre for Medium-Range Weather Forecasts (ECMWF) called Integrated Forecasting System (IFS) (Morcrette et al., 2008). The ECMWF forecast is computed daily, at midnight and 12 h UT, and gives the forecasted values for the next day at a 3-h time step. In this paper we only use the midnight run providing a daily forecast. These values are then interpolated at the same 15-min time step as the two other forecasting methods for comparison purposes. This interpolation step is a linear one, but applied to the clearness index (cf. Appendix A).
- A Cloud Motion Vector (CMV) method using satellite-based 15-min average Surface Solar Irradiance (SSI) estimations. In this case, the source of satellite-based SSI estimation is the database HelioClim-3, which provides SSI estimations in near real-time, every 15 min at 3 km nadir resolution. This database is based on the Heliosat-2 method applied to images from the visible bands of the SEVIRI sensor onboard the geostationary meteorological satellite Meteosat Second Generation (MSG). More details and references can be found in the article by Blanc et al. (2011).

These methods – providing a Global Horizontal Irradiance (GHI) forecast – are applied to two sites in northern France where SPN1 pyranometers (Badosa et al., 2014) are installed to acquire 1-min irradiance measurements. They are referred to this paper as site no. 1 and site no. 2.

The three methods presented above exhibit completely different *a priori* behavior in their resulting forecasts:

- The persistence follows the fluctuations of the measurements but with a systematic delay equal to the forecast horizon.
- The ECMWF forecast is not able to capture the temporal variability of irradiance with a time scale under 3 h.
- The CMV method can detect changes due to advection from cloud movements, but does not take into account cloud diffusion, convections and complex movements due to orography. Various sources of error result from use of this method (uncertainties on SSI estimation of the Heliosat-2 method, errors in CMV, non-advective movements, etc.), but are more likely to correspond with time than the persistence model and capture more rapid changes than ECMWF.

Our aim is not to compare these forecasting methods but to use them as examples to introduce, illustrate and discuss new criteria to better characterize forecast results. Notably, the ECMWF forecast, even resampled at 15-min, is not comparable to a 15-min-ahead CMV forecast, since its forecast horizon varies depending on the time of day. Nevertheless, the difference in representativeness of these two methods on time -and space-scales is of interest in this paper to clearly illustrate our discussions.

3. Overview of existing metrics for solar forecasting

3.1. Standard metrics

Jolliffe and Stephenson (2012, 2003) proposed three categories of metrics depending on their use: administrative, scientific and economic. Although the boundaries of these categories are not clearly defined and can overlap, the authors point out that each metric should be related to a specific use. In this section, we focus on the metrics commonly used in the scientific community to assess forecast accuracy (Beyer et al., 2009).

In order to assess the quality of forecasting methods, quantitative metrics compare the forecast data with reference data – usually irradiance or electrical power in-situ measurements. It should be noted that these “ground truth data” are subject to uncertainties.

The following metrics are computed *a posteriori* to provide statistical information on the past performance of the assessed method. First, the most commonly used metric is the Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{I}_i - I_i\right)^2} \tag{3}$$

where N is the number of data pairs.

Nighttime data obviously offer no relevant information on the solar forecast and so only daytime data are employed to compute this metric. The impact of including nighttime values is described by Hoff et al. (2013). In addition, in-situ measurements are more subject to error for low solar elevation. As a consequence, a threshold on the solar elevation angle is generally imposed. In this article, we have chosen only data in which the solar elevation is greater than 7°, as suggested by Ruiz-Arias et al. (2010).

Relative RMSE corresponds to a normalization by the mean irradiance or the installed capacity of the power plant. We recommend exclusively using the mean irradiance in order to obtain a metric that is as independent as possible from the technical characteristics of the site under study. Depending on the use of the relative RMSE, the mean value can be calculated on the whole set of data