Energy 165 (2018) 620–629

Contents lists available at ScienceDirect

Energy

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

Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability

Alexis Fouilloy [a,*], Cyril Voyant [b,c], Gilles Notton [a], Fabrice Motte [a], Christophe Paoli [a], Marie-Laure Nivet [a], Emmanuel Guillot [d], Jean-Laurent Duchaud [a]

[a] University of Corsica, CNRS UMR SPE 6134, Centre Georges Péri, Route des Sanguinaires, 20000 Ajaccio, France
[b] Castelluccio Hospital, Radiotherapy Unit, BP 85, 20177 Ajaccio, France
[c] University of La Réunion - PIMENT Saint-Pierre, Réunion, France
[d] PROMES-CNRS Laboratory, 7 Rue du Four solaire, 66120 Font-Romeu-Odeillo-Via, France

ARTICLE INFO

Article history:
Received 29 August 2017
Received in revised form
3 September 2018
Accepted 17 September 2018
Available online 19 September 2018

Keywords:
Time series forecasting
Machine learning
Variability
ARMA
ANN
Regression tree
Gaussian process
SVR

ABSTRACT

Eleven statistical and machine learning tools are analyzed and applied to hourly solar irradiation forecasting for time horizon from 1 to 6 h. A methodology is presented to select the best and most reliable forecasting model according to the meteorological variability of the site. To make the conclusions more universal, solar data collected in three sites with low, medium and high meteorological variabilities are used: Ajaccio, Tilos and Odeillo. The datasets variability is evaluated using the mean absolute log return value. The models were compared in term of normalized root mean square error, mean absolute error and skill score. The most efficient models are selected for each variability and temporal horizon: for the weak variability, auto-regressive moving average and multi-layer perceptron are the most efficient, for a medium variability, auto-regressive moving average and bagged regression tree are the best predictors and for a high one, only more complex methods can be used efficiently, bagged regression tree and the random forest approach.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The management of an electrical network is a very complex task, particularly on small grids like islands which are generally not interconnected to the mainland grid [1]. The equality between production and consumption is the most difficult challenge; moreover, the increased rate of intermittent and stochastic renewable energy sources into the energy mix adds a layer of difficulty. For an efficient management of the energy mix and a safest electricity supply, production and consumption must be planned before. There are many forecasting methods for solar radiation, all the methods have advantages and inconvenient and their accuracy depends on the forecasting horizon and on the geographical situation. The three main categories of models are sky imaging, physical models (or numerical weather prediction, e.g. NWP) and machine learning models. The sky imaging models are mainly used for short term forecasting (1–60 min), the physical models concern numerical weather prediction models often used for long term forecasting (several days). The medium-term forecasting (from 1 to few hours) is often made by machine learning methods, a large state of the art was made concerning the solar power forecasting [2]. This study is related to the comparison of eleven machine learning models in order to forecast solar global irradiation and to compare their performances in three different sites to make our results more universal. The main goal is to propose a prioritization methodology to highlight the best forecasting method between the tested ones according to the level of irradiation variability considering the complexity of the method. To distinguish the different weather conditions and to characterize the global behavior of the solar time series in the three sites, a property of time series named variability is calculated, the characterization of this time series property can be made by various statistical parameters [3]. The solar forecasting is realized with an hourly time granularity for 1 h to 6 h' time horizons. The comparison between the models is

* Corresponding author.
E-mail addresses: fouilloy_a@univ-corse.fr (A. Fouilloy), cyrilvoyant@gmail.com (C. Voyant), notton@univ-corse.fr (G. Notton), motte_f@univ-corse.fr (F. Motte), cpaoli@univ-corse.fr (C. Paoli), nivet@univ-corse.fr (M.-L. Nivet), emmanuel.guillot@promes.cnrs.fr (E. Guillot), duchaud_jl@univ-corse.fr (J.-L. Duchaud).

https://doi.org/10.1016/j.energy.2018.09.116
0360-5442/© 2018 Elsevier Ltd. All rights reserved.

evaluated by statistical indexes and classical error metrics, the eleven tested models are:

- Persistence (P) and smart persistence (SP) can be considered as naïve models, they are often used in intraday solar radiation forecasting and are usually taken as reference to compare the performances with other models [4].
- Auto regressive moving average (ARMA) and artificial neural network (ANN) are two types of models allowing to realize respectively linear and non-linear regressions. ARMA is certainly the most studied tool especially in econometry, but also in global irradiation forecasting [5]. Artificial neural network (ANN) is taken as well-known global radiation forecasting method [6].
- Regression tree based models family is composed by five models with different levels of complexity and optimization. Classical regression tree (RT) and pruned regression tree (RT-pruned) are sometimes used in very short term forecasting and allow good results compared with other machine learning models [7]. Boosted regression tree (RT-boosted) consists in combine the responses of a collection of weak classifiers which once averaged form a strong classifier to make prediction [8]. Then the aggregated versions of regression trees denoted bagged regression tree (RT-bagged) [9] are developed to build an aggregated predictor in order to improve accuracy of the forecast. Finally, we propose also the random forest method (RF) which is another improvement of regression tree methods with more robustness [10].
- Models based on Kernel estimation like Gaussian processes (GP) and support vector regression (SVR). The first one is a nonparametric kernel-based probabilistic model giving very good results in time series prediction [11], The second one (SVR) can be considered as a generalization of linear classifiers and can be used to forecast solar irradiation and can be improved by several methods of optimization [12].

The study is developed as follow: section 2 describes the materials and methods, the data and the models. Section 3 presents the results and the discussion about the forecasting reliability, the variability of solar data and the link between variability and efficiency of the forecasting models.

2. Materials and methods

This section presents information related to the preliminary stage before the forecasting phase. Fig. 1 shows the different steps (detailed afterwards) of this methodology.

It consists in:

1. A quality control of the solar data: often, mistakes appear in the temporal series of solar data due to problems with the acquisition system; an automatic quality check used in the frame of GEOSS project (Group on Earth Observation System of System) (http://www.earthobservations.org/geoss.php) has been applied to the data;
2. A data preprocessing: the night hours are removed, due to problems occurring at the sunset and the sunrise (mask effects and bad response of pyranometers) [13,14], a filter is applied on the datasets which remove all the data that correspond to a solar elevation angle up to 10°.
3. Stationarity: the solar irradiation time series contains some properties like seasonality and periodicity. Most of the machine learning forecasting methods are applicable only to stationary time series [15] describe the necessary steps in order to use data in forecasting with machine learning. Thus, the solar data time series must be made stationary using a solar radiation model by clear sky and computing the clear sky index (paragraph 2.1), this method is described in Ref. [16].
4. At last, the choice of the number of input data for each forecasting tool is realized by auto mutual information method [17] and described in paragraph 2.2.

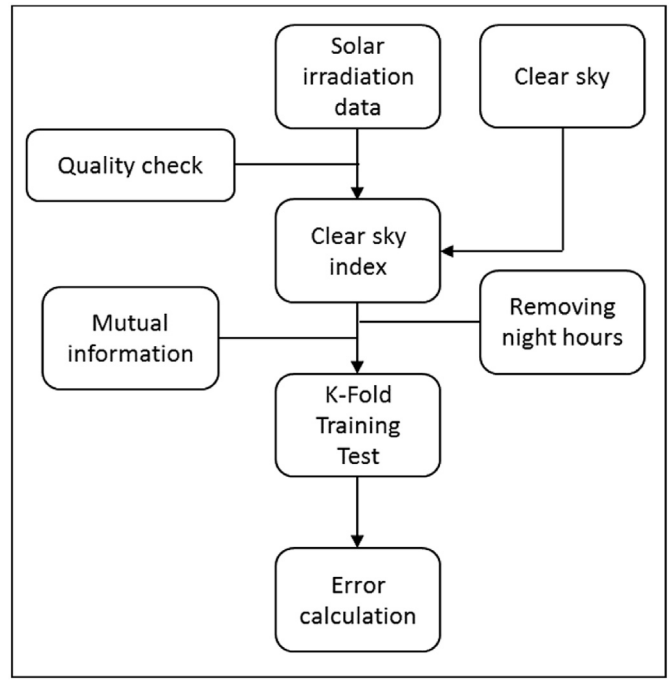

Fig. 1. Flowchart of the model development before forecasting.

2.1. The clear sky model (CS) and clear sky index (CSI)

The solar radiation by clear sky computing is the third step of the preprocessing method, it consists in calculate the maximum global horizontal irradiation (GHI) at the ground level in cloudiness conditions (taking into account scattering and absorption by atmosphere). Almost all the machine learning approaches are based on a stationary hypothesis of the time series. they are usually based on the assumption that the data generation mechanism does not change over time. For the global irradiation forecasting in dividing each measured solar irradiation by its corresponding clear sky value, the periodicity of the solar irradiation time series is then removed; thus, a new time series is created with values between 0 and 1 and the forecasting method is then applied to predict only the cloud occurrences and not the fact that GHI is more important at 12 h than at 6 h, or in summer than in winter. The CS model chosen is the SOLIS model developed by Mueller [18] and based on the studies of Ineichen et al. [19,20]; it gives excellent results when it was compared with solar measurements in Europe. It is based on radiative transfer models, Beer Lambert function, and their integration on the solar spectra uses the information's of news satellites ERS-2/ENVISAT. The calculation of the clear sky irradiation CS(t) is given by:

$$CS(t) = H_0.e^{\frac{-\tau}{\sin^b(h(t))}}.\sin(h(t)) \quad (1)$$

h(t) is the solar height in degrees and H_0 is the extraterrestrial irradiation (irradiation on the top of the atmosphere before being scattered and absorbed by the atmosphere [13]); τ is the global

Download English Version:

https://daneshyari.com/en/article/11026431

Download Persian Version:

https://daneshyari.com/article/11026431

Daneshyari.com