Measurement 63 (2015) 346–363

Contents lists available at ScienceDirect

Measurement

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

Models for solar radiation prediction based on different measurement sites

Massimo Lazzaroni [a,b,*], Stefano Ferrari [c,*], Vincenzo Piuri [c], Ayşe Salman [d], Loredana Cristaldi [e], Marco Faifer [e]

[a] Department of Physics, Università degli Studi di Milano, Milan, Via Celoria, 16, 20133 Milano, Italy
[b] I.N.F.N. Milano, Via Celoria, 16, 20133 Milan, Italy
[c] Department of Computer Science, Università degli Studi di Milano, Via Bramante, 65, 26013 Crema, Italy
[d] Department of Computer Engineering, Doğuş University, Istanbul, Acibadem, Kadikoy, 34722 Istanbul, Turkey
[e] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci, 133, 20133 Milan, Italy

ARTICLE INFO

Article history:
Received 14 October 2013
Received in revised form 5 November 2014
Accepted 10 November 2014
Available online 17 December 2014

Keywords:
Solar radiation prediction
Autoregressive models
Extreme Learning Machines
Support Vectors Regression
Measurement

ABSTRACT

The modeling of solar radiation for forecasting its availability is a key tool for managing photovoltaic (PV) plants and, hence, is of primary importance for energy production in a smart grid scenario. However, the variability of the weather phenomena is an unavoidable obstacle in the prediction of the energy produced by the solar radiation conversion. The use of the data collected in the past can be useful to capture the daily and seasonal variability, while measurement of the recent past can be exploited to provide a short term prediction. It is well known that a good measurement of the solar radiation requires not only a high class radiometer, but also a correct management of the instrument. In order to reduce the cost related to the management of the monitoring apparatus, a solution could be to evaluate the PV plant performance using data collected by public weather station installed near the plant. In this paper, two experiments are conducted. In the first, the plausibility of the short term prediction of the solar radiation, based on data collected in the near past on the same site is investigated. In the second experiment, the same prediction is operated using data collected by a public weather station located at ten kilometers from the solar plant. Several prediction techniques belonging from both computational intelligence and statistical fields have been challenged in this task. In particular, Support Vector Machine for Regression, Extreme Learning Machine and Autoregressive models have been used and compared with the persistence and the *k*-NN predictors. The prediction accuracy achieved in the two experimental conditions are then compared and the results are discussed.

1. Introduction

The electric power system is mainly composed of units for energy production i.e. generators, loads and a power grid that connects them. Actual configuration principally includes large central generators which, through the transformers, inject electrical power in the transmission grid. The world energy infrastructure is nowadays subjected to an important transformation such as the growing number of distributed small generation units, based on different technologies, directly connected to the power grid. These small generation units put side by side to the large and traditional ones are defining a grid based on the so called distributed generation. This kind of network architecture

* Corresponding authors at: Department of Physics, Università degli Studi di Milano, Milan, Via Celoria, 16, 20133 Milano, Italy (M. Lazzaroni).
E-mail addresses: massimo.lazzaroni@unimi.it (M. Lazzaroni), stefano.ferrari@unimi.it (S. Ferrari), vincenzo.piuri@unimi.it (V. Piuri), asalman@dogus.edu.tr (A. Salman), loredana.cristaldi@polimi.it (L. Cristaldi), marco.faifer@polimi.it (M. Faifer).

http://dx.doi.org/10.1016/j.measurement.2014.11.037
0263-2241/

implies new problems concerning the management. In fact in traditional network the stability of the power system was achieved by means of the direct control of few large conventional power generators. By introducing distributed generation this approach cannot be followed, since the small generation units are basically not controllable by the network system operator. In particular this scenario is critical when units based on renewable energy resources are used, since they can only provide power as long as the source of energy is available. In many situation the energy production is mainly utilized directly by the producer or by nearby buildings. When energy production exceed the necessity the excess flows into the power grid of the utilities. In order to implement an electric grid allowing a large amount of distributed energy sources, a different approach to the problem of the network stability is required. It is clear that in this scenario the possibility to predict the plant's power production during the day greatly helps the management of such a power system. Besides, the prediction of energy production becomes extremely important for the trading in the free energy market.

Among the renewable energy sources, the photovoltaic (PV) technology, which allows to obtain electric energy from solar radiation [1], has the benefit of the low environmental impact. On the other end, its main weakness is that its availability cannot be fully controlled. Many aspects need to be considered such as geographic position, local climate, weather and global efficiency of the panel [2–12]. Among these, the position and the climate influence on the solar radiation can be easily obtained from astronomical and statistical data, but the weather is characterized by a high variability and depends on many physical factors. According to [13], the forecasts required by the activity related to the grid management can be divided in two categories. The first is related to grid stability problem (intra-hour, hour ahead, and day ahead), while the second concerns planning and assets optimization on medium and long-term (monthly and yearly forecasts, respectively). Since the main factor for solar radiation availability is the local weather, approaches based on weather forecast have been widely used in literature. These are based on data obtained from satellite observations and ground stations. The geographic and time availability of data are the main aspects that have to be taken into account. Besides, the sampling rate of the measurement (both in time and in space) have to be related to the granularity of the forecast.

The solar radiation prediction can be based on data obtained by several data sources, characterized by the type of data they produce, as well the space–time granularity they provide. These data source are, for example: Numerical Weather Prediction models, Satellite-base forecast, All-sky imagers, Ground measurements.

In this paper we consider the problem of the solar radiation prediction from data captured by ground weather stations, in two different operational conditions: the local (when the data are captured at the plant site) and the remote prediction (when the weather station is distant from the plant site).

Several forecasting approaches have been used in literature. Among these, the most effective in producing hour-ahead predictions are based on empirical regression, neural networks [14,15] and time-series models (e.g., ARMA, ARIMA, Non-Linear autoregressive models) [16–18]. In particular, in [14] several prediction techniques have been challenges to forecast the daily solar radiation in Algeria using the historical data in the period from 1981 to 2001, and a wavelet-neural network resulted the best. In [15] a good review of the literature on the use neural networks for solar radiation prediction is presented. It highlight how the different approaches differ for the type of solar radiation predicted, for the input parameters and for the prediction error used. In [16] the historical data of the period 2003–2009 for the solar radiation in Bogot (Colombia) have been used to determine an autoregressive model in order to provide a long period forecast useful for assessing the productivity of solar plants. A similar study have been conducted in [17], where the autoregressive models of the solar radiation of several Nigerian cities have been built using the 1986–1990 data. Instead in [18] a Non-linear Autoregressive model is used to model the solar radiation in Edinburgh (UK) using 30 s averages of radiation measured for 8 h per day for one month.

For large plants, the investment to provide the plant with the instrumentation (and the management) for on-site weather measurement is appropriate, since it can provide data for both a reliable production forecast and an effective maintenance scheduling [19–22,11,12]. For small plants, instead, this cost is questionable and the exploitation of data coming from a remote public weather station can be appealing. In fact, this plants are mainly intended for fulfilling the local needs, and only the excess of production is sent to the grid. Hence some requirements on the prediction accuracy can be relaxed, since only the integral of the contribution of the small plants is important for the grid management. On the other hand, there is a clear interest for a reliable scheduling of the maintenance that guarantees the efficiency of the plant as far as a system with an high level of Dependability features, i.e. Reliability, Availability, Maintainability, Safety (RAMS requirements) [23–28].

In our previous works [29–31], several models (namely, Autoregressive models, Support Vector Machine, and Extreme Learning Machine) have been challenged in the task of predicting the global horizontal illuminance. In the present work, instead, another physical characteristics – the global horizontal radiation – will be considered and two experiments, already introduced in [32,33], will be described. The first experiment consists in a short term prediction of the solar radiation using only data measured in the near past. For this scope, a two-year hourly dataset of the global horizontal radiation will be used to feed some autoregressive models in order to obtain a one-hour forecast. In particular, the dataset has been collected in two years by the MeteoLab [34,35], Milan, Italy. In the second experiment, the same prediction will be performed using data of a different set-up. Data from two public weather stations [36] in the neighborhood of Milano, separated by ten kilometers in a straight line, are used for assessing the effectiveness of short term prediction of the production of a plant using measurements taken in a nearby measurement station. In particular, a 3-year hourly dataset will be used to model the time series of the global horizontal