Renewable Energy 90 (2016) 513–519

Contents lists available at ScienceDirect

Renewable Energy

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

Comparison of Photovoltaic plant power production prediction methods using a large measured dataset

G. Graditi*, S. Ferlito, G. Adinolfi

ENEA Portici Research Center, Italy

ARTICLE INFO

Article history:
Received 24 August 2015
Received in revised form
14 November 2015
Accepted 5 January 2016
Available online 18 January 2016

Keywords:
Artificial neural network
Estimation
Genetic algorithms
MLP
Photovoltaic power production
Regression analysis

ABSTRACT

Nowadays the estimation of power production yield by stand-alone and grid-connected Photovoltaic (PV) plants is crucial for technical and economic feasibility design analyses. The main goal is to overcome renewables unpredictability by properly estimating the power production and by suitably balancing generation and consumption. In this context, many methods can be applied to forecast renewables energy production. The scope of this paper is a comparative analysis of three different methods to estimate the power production of a preexisting PV plant. It is installed at ENEA Research Centre located in Portici (South Italy) and it is integrated in a Micro Grid (MG) configuration. In detail a phenomenological model proposed by Sandia National Laboratories and two statistical learning models, a Multi-Layer Perceptron (MLP) Neural Network and a Regression approach, are compared. These models are deeply different also in terms of required input data and parameters. In detail, phenomenological model application requires the availability of design parameters and technical devices specifications. Statistical machine learning models need, however, input variable previously acquired datasets. The a-Si/μc-Si PV plant, installed at Portici, represents an adequate case study for the three models comparison, as both design and acquired data are available. In fact, the plant was designed at the ENEA Research Centre so this makes possible the knowledge of the design parameters and, being a part of the MG, its data are continuously acquired and transmitted to other network devices. Obtained results demonstrate more accurate power predictions can be reached by statistical machine learning approaches. The main novelty of the paper consists in the optimization of the considered models by the appropriate identification of the minimum and more representative training dataset. Authors underline the unnecessary use of thousands samples by suitably selecting the dataset size and samples by means of a Genetic Algorithm. The optimization strategy effectiveness is verified comparing the prediction performances obtained employing the optimal dataset with those obtained with a randomly chosen dataset. In this scenario, Genetic Algorithm strategy represents a successful approach to the suitable identification of statistical models datasets.

1. Introduction

1.1. Motivation and approach

In the last years, global demand for electric energy is considerably increased, so requiring the identification of new energetic solutions. In this scenario, integration of renewable energies, in preexisting energy systems represents an important task to be considered [1]. At present, PV and wind technologies have reached high levels performances in terms of both efficiency and reliability, so representing promising solutions to the Zero Energy Balance application and realization. The price reduction and the provided user-friendly cable solutions have contributed to these limitless sources diffusion in civil and industrial contexts [2]. These systems are constituted by large amount of devices: PV generators [3], Distributed Maximum Power Point Tracking (DMPPT) converters [4,5], inverters, storage systems, grid interface devices, etc. Proper and efficient operating modes are assured only by conveniently sized and matched components. To achieve this aim, a previous accurate model development is an essential task.

An aspect influencing renewable energy sources widespread is their intrinsic unpredictability. In fact, their power production

* Corresponding author.
E-mail address: giorgio.graditi@enea.it (G. Graditi).

http://dx.doi.org/10.1016/j.renene.2016.01.027
0960-1481/

Nomenclature

ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial neural network
DAS	Data Acquisition System
DMPPT	Distributed Maximum Power Point Tracking
GAs	Genetic Algorithms
GI_{poa} [W/m^2]	Global Irradiation on plane of array
GI_{csk} [W/m^2]	Global Clear Sky Irradiation
LS	Least Square
MG	Micro Grid
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression
PV	Photovoltaic
RES	Renewable Energy Sources
SAM	System Advisor Model
SVM	Support Vector Machine
T_{amb} [°C]	Ambient temperature
T_m [°C]	PV module temperature
W_{speed} [m/s]	Wind speed

depends on continuously changing and not predictable variables, such as irradiance, temperature, wind speed and shadows. As a consequence, the introduction of not programmable energy sources in energy supply systems presume the availability of suitable techniques and tools to accurately predict renewables productions in order to correctly schedule them with other energy sources for generation-load balancer requirements. In this scenario, the identification of methods for a meticulous estimation of alternatives productions under realistic operation conditions [6,7] constitutes a valid solution to increase these sources integration, thereby allowing an accurate economic and technical assessment.

1.2. Literature review

In literature, many studies on forecasting and estimation methods are available. In detail, the specific literature on PV plant power production estimation presents three different types of models: phenomenological, stochastic/statistical learning and hybrid ones.

Deterministic approaches based on physical phenomena try to predict PV plant output by considering the electrical model of the PV devices constituting the plant. Such kind of models are usually implemented in commercial software such as PVSyst, System Advisor Model (SAM) and so on. For example in Refs. [8], a deterministic approach was used to model electrical, thermal, and optical characteristics of PV modules. By using hourly solar resource and meteorological data, the model has been validated for different modules types. Other approaches are the statistical and machine learning ones, such as: ANN, Support Vector Machine (SVM), Multiple Linear Regression (MLR), Adaptive Neuro-Fuzzy Inference System (ANFIS). These methods operate without any a priori knowledge of the system under consideration. They try to “understand” the relation between inputs and outputs by adequately analyzing a dataset containing acquired input and output variables collection. Statistical learning algorithms have many advantages. In fact, starting from examples, they are able to learn from them, and they can work also in presence of incomplete data. Once trained, they are able to generalize and to provide predictions. Their features make them suitable to be used in different contexts. For example, in Refs. [9,10], an adaptive ANN was used to model and size a stand alone PV plant, using a minimum input dataset. An ANFIS was applied in Ref. [11] to model the different devices constituting a PV power system and its output signals. In Refs. [12,13], a linear regression model and an ANN were applied to estimate daily global solar radiation. A hybrid model can represent a further approach [14]. It combines different models to overcome limitations characterizing one single technique. In addition, “ensemble” methods [15] build predictive models by integrating multiple strategies in order to improve the overall prediction performance. In literature, terms “forecast” and “estimate” are often interchangeably used. More correctly, the term “forecast” should be adopted to provide an expectation of future values of considered variables based on previously known/acquired data. On the other hand, the term “estimate” means to represent the relation between output and inputs, at the same timestamp. In this paper, the aim is to provide an “estimate” for an output variable. The models provide a single output value without a prediction horizon knowing inputs at the same timestamp.

1.3. Innovative contribution

In this paper, the PV power production of a 1 kW$_p$ experimental micro-morph silicon PV plant is estimated. It is an experimental facility installed at the ENEA - Italian National agency for new technologies, Energy and sustainable economic development - Research Center of Portici located in the South Italy. Authors main attention is focused on the comparative analysis between a phenomenological and two statistical learning techniques taking advantage of meteorological, thermal and electrical in situ acquired data. The innovative aspect consists in the optimization of ANN and Regression models training datasets. These models need to analyze previously acquired data to “learn” the input/output relationship. Such approach is advantageous in many cases, especially when data are incomplete or unknown or, as in MG, PV plants are continuously monitored. In this paper, the attention is also focused on statistical learning models optimization. The optimization target not only consists in the training dataset minimization, but also in the identification of the most representative and meaningful set of data to train the ANN and the Regression model.

1.4. Paper organization

The paper is organized as follows. In section II, the experimental PV plant is described. In section III, the different adopted approaches to estimate PV power production are presented. In section IV, metrics to evaluate models performances are illustrated. Simulation results are reported and discussed in section V. Finally, conclusions and proposal for future works are reported.

2. Experimental PV plant description

In this paper, a preexisting PV plant installed at the ENEA Research Center is considered. It consists in 28 amorphous and μ-crystalline silicon (a-Si/μc-Si) modules [16,17], divided in 14 strings of two series connected generators. They are 37.5 W$_p$ micromorph silicon commercial PV generators connected to a commercial 1 kW$_p$ output nominal power inverter. The PV modules are South oriented and installed with a 20° tilt angle on a roof structure. PV plant and DAS details are reported in Fig. 1. This PV plant represents a suitable case study for the models comparative analysis since both its design and acquired data are available. In detail, the carried out analysis uses a dataset consisting of meteorological, thermal and electrical data acquired from 2006 to 2012 and sampled every ten minutes. This dataset is constituted by the global radiation on plane of array GI_{poa}, the ambient temperature T_{amb}, the wind speed