Electrical Power and Energy Systems 82 (2016) 233–241

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

Classification of hourly solar radiation using fuzzy *c*-means algorithm for optimal stand-alone PV system sizing

Khalil Benmouiza, Mohammed Tadj, Ali Cheknane *

Laboratoire des Semiconducteurs et Matériaux Fonctionnels, Université Amar Telidji de Laghouat, Bd des Martyrs, BP 37G, Laghouat (03000). Algeria

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form 11 March 2016
Accepted 12 March 2016
Available online 25 March 2016

Keywords:
Hourly solar radiation
Sizing
Clustering
Stand-alone PV systems
Loss of load probability

ABSTRACT

Recent applications in remote areas need a continuous source of power. A photovoltaic system is an arrangement of components; generally PV panels, converters and storage components designed to supply usable electric power for a variety of purposes. Stand-alone photovoltaic systems could assure this power if the optimal sizing is reached especially in small scales. They are ideal for remote rural areas and applications where other power sources are either impractical or are unavailable. This paper presents a method for techno-economic optimization for the optimal stand-alone PV system sizing in hourly scale basing on hourly solar radiation classification and the loss of load probability (*LLP*) calculation. For this task, the fuzzy *c*-means clustering algorithm was applied to extract useful information from hourly solar radiation time series. To this end, an inclined solar radiation data from the site of Ghardaia, Algeria with an inclination angle equals to 32° for 2012 is used. Secondly, the lowest solar radiation was chosen for the reliability sizing based on the numerical method for a desired *LLP* equal to 1%. Finally, the genetic algorithm optimization method was used for choosing the best configuration with the lowest cost. The simulation results show the simplicity, rapidity and the accuracy of the proposed method. The sizing in hourly solar radiation scales ($C_A = 0.91; C_S = 3.2$) gives better results than daily solar radiation scales ($C_A = 1.09; C_S = 4.4$).

Introduction

The growing demand of the energy in our industrialized modern life requires the search of alternative sources of energy outside the limited and polluted current sources such as fossil fuels. Renewable energies, such as the wind and solar energies are an environment-friendly solution. On the other hand, solar energy was used widely in two large categories; grid-connected and stand-alone PV (SAPV) systems. The SAPV systems are used in remote areas and space applications, where the load demand is assured only by the PV system (panels with batteries) [1–4]. Moreover, other types of stand-alone systems are based on a hybrid methodology that improves the efficacy of the installed system [5–7]. However, the sizing of SAPV systems are known to be a complicated task. It consists of finding the best compromise between the reliability and the cost that the system (PV panels and batteries) can feed the load at any time considering the desired loss of load probability (*LLP*) [1–3]. The *LLP* is a parameter used to characterize the system design. It is defined as the relationship between the energy deficit and the energy demand, as referred to the load. In statistical terms, the *LLP* value refers to the probability of the system to be unable to meet the demand [1]. The main reason for this failure is the stochastic characteristics of the solar radiation that effect the sizing process. Kaplani and Kapanis [8], Chen [9] and Cabral et al. [10] proved that the fluctuation observed in the daily solar radiation highly affect the reliability of the PV system sizing. Hence, it is an important task to study the dynamic behaviour of the solar radiation time series before any sizing processes. In addition, the failures can occur due to losses in the cables, in the battery storage, the ageing of the components, the degradation of the PV panel's performance and the charge/discharge effects of the batteries.

Several methods are shown in the literature in how to calculate the optimal sizing parameters for a constant *LLP* [2]. Among them, intuitive methods [11,12], where the size of the system is taken in such way to ensure the load demand without gives a relation between the number of panels, batteries and the *LLP*. Analytical methods [13–15], based on the graphical information obtained from of the iso-probability curves, and the numerical methods [16–18], which are based on a detailed simulation of the PV system

* Corresponding author. Tel./fax: +213 (0)29 14 58 47.
E-mail addresses: benkhalil3@gmail.com (K. Benmouiza), mohtej@gmail.com (M. Tadj), a.cheknane@mail.lagh-univ.dz, cheknanali@yahoo.com (A. Cheknane).

http://dx.doi.org/10.1016/j.ijepes.2016.03.019
0142-0615/

Nomenclature

A_f	actualization factor
A_G	used photovoltaic array area, m^2
A_r	actualization rate
B	temperature coefficient
c	number of clusters
C_A	capacity of the photovoltaic panel array
C_B	nominal battery capacity, Ah
C_b	price of the battery, $
C_c	total components cost, $
C_m	maintenance cost, $
$C_{m(1)}$	maintenance cost of one year, $
C_p	price of one panel, $
C_r	total replacement cost, $
$C_{r'}$	system cost, $
C_S	capacity of the storage system
C_t	actualized total cost, $
C_U	maximum battery useful capacity, W h
DOD	maximum depth of discharge
$d^2(x_k, v_i)$	distance measure between data and cluster centre
E^{Aux}	auxiliary generator
E_f	annual expansion factor
Err	error
FCM	fuzzy c-means
GA	genetic algorithm
$H_{g\beta_i}$	hourly global solar radiation received on inclined surface, Wh/m^2
$\overline{H_{g\beta_i}}$	mean hourly global solar radiation received on inclined surface, Wh/m^2
$I(x(t), x(t-\tau))$	mutual information
J_{FCM}	objective function
L_F	system life time, year
LLP	load of loss probability
LLP_s	desired LLP
L_t	load at time t, W
M	number of embedded points in the m-dimensional space
m	embedding dimension
N	total number of points of the time series
n	length of the data
N_b	number of batteries
nb	battery life time, year
N_p	number of panels
$p(x(t), x(t-\tau))$	joint probability mass function
PV	photovoltaic
q	weighting exponent on each fuzzy membership
R_f	replacement factor
SAPV	stand-alone photovoltaic systems
SOC	state of charge
T_c	cell temperature, °C
T_r	reference temperature of the panel, °C
$U^{(t)}$	membership matrix
V_B	voltage for the unit of storage, V
v_i	centre of the ith cluster
$V^{(t)}$	cluster centre matrix
$X(t_i)$	embedded time series into an m-dimensional space
$x(t_i)$	scalar time series
ε	small error
η_p	efficiency of a solar cell at a referenced solar radiation
η_t	total panels efficiency
u_{ik}	degree of membership
τ	delay time

in small scales (daily, hourly... etc.). In the numerical methods, the produced energy from the PV generator and the state of the charge of the batteries are calculated at each time t. The advantages of the numerical methods are the precision and the simplicity of choosing different elements of the system [17–20]. Several algorithms have been proposed in the literature in how to calculate LLP values based on numerical methods. They are based on the measured monthly or daily solar radiation data over a long time period, taking into account the worst month of the year [17,19,20]. However, they present major drawbacks as the length and non-occurrence of the used solar radiation data (only in daily or monthly basis), where most of solar radiation applications need hourly solar radiation data [1,21,22]. In addition, choosing the worst month in the year does not give all the information on the dynamic characteristics of the measured solar radiation (bad weather may exists outside the worst month) that limited the importance of the sizing in the daily basis. We propose in this paper, a modified numerical algorithm based only on the classified hourly solar radiation using a clustering algorithm to obtain the lowest solar radiation data over one year period that ensures the optimal sizing in hourly basis. The novelty of the proposed method, compared to other methods presented in literature [1,2,7,12–17,19,20], is to first assure, the full identification of the dynamic characteristics of hourly solar radiation time series, and second, guaranteed optimal sizing by using only one year hourly solar radiation data.

At the first stage, time series data mining was applied to hourly solar radiation data. It consists of grouping similar elements into clusters that have the same characteristics [23,24]. For this purpose, several methods were proposed in the literature based on unsupervised clustering methods such as k-means and fuzzy c-means (FCM) algorithms [25–28]. We have chosen FCM algorithm, which are based on fuzzy methods that gives good results. Before that, phase space reconstruction is needed to overcome the nonlinearity of the solar radiation data. It presents the data in high dimensional space based on Takens [29] theorem. The motivation of using the FCM algorithm because it is a powerful tool, which presents more precise results comparing with k-means algorithm [30]. Nevertheless, it strongly depends on the initialization parameters (initial number of clusters), Hence, the sub-clustering method [31] was used in this paper to decide the right number of clusters.

In the second phase, the low solar radiation data obtained from the clustering processes is used to determine the iso-probability curves using the numerical method. For each hour, the LLP is calculated and compared to the desired LLP (1% in our example). After that, an economic study of the system is achieved by minimizing the total cost of the system to get satisfactory results [32–34]. We have used the genetic algorithm (GA) method as optimizing technique for the search of the optimal solution. Finally, a general decision is selected that ensure the minimization of an objective function, which guaranteed the minimum cost of the panels and batteries pairs that give an LLP equals to 1% for a desired hourly load.

The remaining part of this paper is organized as follows; Section 'Methodology' present the methodology used in both technical and economical PV sizing, a background of space phase reconstruction, fuzzy c-means clustering algorithm, reliability and economical sizing method are also considered. In Section 'Results and discussion', we present the simulation results of the proposed method comparing the results of different configurations. The last section gives the conclusion of the proposed work.