AN ANFIS-BASED PREDICTION FOR MONTHLY CLEARNESS INDEX AND DAILY SOLAR RADIATION: APPLICATION FOR SIZING OF A STAND-ALONE PHOTOVOLTAIC SYSTEM

A. Mellit[1,2], A. Hadj Arab[2,3] and S. Shaari[4*]

[1]Department of Electronics, Faculty of Sciences Engineering,
Jijel University of Médéa, 26000, Algeria
[2]Development Centre of Renewable Energy (CDER), P.O. Box 62, Bouzareah,
Algiers 16000, Algeria
[3]Departamento de Energias Renerables- CIEMAT, Arda Complutense, 22,
Madrid 28040, Spain
[4]Faculty of Applied Sciences, Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia

*Corresponding author: solarman@salam.uitm.edu.my

Abstract: *A suitable Neuro-Fuzzy model is presented for estimating sequences of monthly clearness index* $(\bar{K}_t)$ *in isolated sites based only on geographical coordinates. The clearness index* (K_t) *corresponds to the solar radiation data (H) divided by the corresponding extraterrestrial data* (H_0)*. Solar radiation data is the most important parameters for sizing photovoltaic (PV) system. The Adaptive Neuro-Fuzzy Inference System (ANFIS) model is trained by using the Multilayer Perceptron (MLP) based on the Fuzzy Logic (FL) rule. The inputs of the network are the latitude, longitude, and altitude, while the outputs are the 12-values of* $\bar{K}_t$*, where these data have been collected over 60 locations in Algeria. The* $\bar{K}_t$ *corresponding of 56 sites have been used for training the proposed ANFIS. However, the* $\bar{K}_t$ *relative to 4-sites have been selected randomly from the database in order to test and validate the proposed ANFIS model. The performance of the approach in the prediction* $\bar{K}_t$ *is favorably compared to the measured values, with a Root Mean Square Error (RMSE) between 0.0215 and 0.0235, and the Mean Relative Error (MRE) not exceeding 2.2%. In addition, a comparison between the results obtained by the ANFIS model and other Artificial Neural Networks (ANN) is presented in order to show the performance of the model. An example of sizing PV system is presented. Although this technique has been applied for Algerian locations, but can be generalized in any geographical location in the world.*

Keywords: clearness index K_t, solar radiation, sizing PV system, ANFIS, ANN

1. INTRODUCTION

The clearness index (K_t) is defined as the ratio between total H and the H_0. The amount of global solar radiation and its temporal distribution are the primary variables for designing solar energy systems. Knowledge of these parameters is required for the prediction of system efficiency of a possible solar energy system at a particular location. It is the most important parameter for sizing of stand-alone PV systems.[1–4] The application of PV system can be used for electrification of villages in rural areas, telecommunications, refrigeration, water pumping (particularly in agricultural irrigation), water heating and, etc. Several studies in literature have been developed in order to estimate these data (H) based on statistical approach and the ANN techniques.[5–9] The application of the wavelet analysis with ANNs has been proposed in order to predict the total H in the missing period,[10,11] good accurate results have been obtained with a correlation coefficient of 97%. Therefore, these techniques are not adequate for isolated locations, but it is a very good proposition in the missing data period case. The proposed method[12] can solve this problem but it needs the availability of mean temperature and sunshine duration. A critical study of the prediction global H from sunshine duration is proposed in Yorukoglu and Celik.[13] The authors[14] have proposed the use of Radial Basis Function Network (RBFN) in order to estimate the monthly H for 41 Saudi Arabia sites, the results for testing obtained were within 16% (MRE), and the same principle is applied for Spain and Turkey locations based on the MLP for developing the solar radiation map.[15,16] A more recent study has been presented.[17] In this study, a hybrid model based on ANN (MLP) and Matrices Transition Markov (MTM), has been developed in order to estimate the total H in isolated sites for Algeria locations. The model is called the MLP-MTM approach and the correlation coefficient obtained ranges between 90% to 92%.

The major objective of this paper is to investigate the potential of an ANFIS system in the modeling and prediction of the $\overline{K}_t$, in isolated sites and to assess its performance relative to ANNs, and then for improving the results obtained in an earlier work.[17] In this work, we have used the 12-values of $\overline{K}_t$ in the output of the model instead of the average H as often used. The data used in this study was collected from meteorological stations of Algeria.

2. DATA SET

The database used in this study consists of 60 × 12 monthly solar radiation values collected from the National Office of Meteorology (NOM) in Algeria. Each site contains 12-values corresponding to monthly radiation data. The database has been normalized by dividing each monthly H to the H_0, to obtain a database of 60 × 12 $\overline{K}_t$. Figures 1(a) and (b) show the monthly total solar radiation and $\overline{K}_t$ data for some sites. Table 1 presents the database of average total H and K_t used in the simulation.

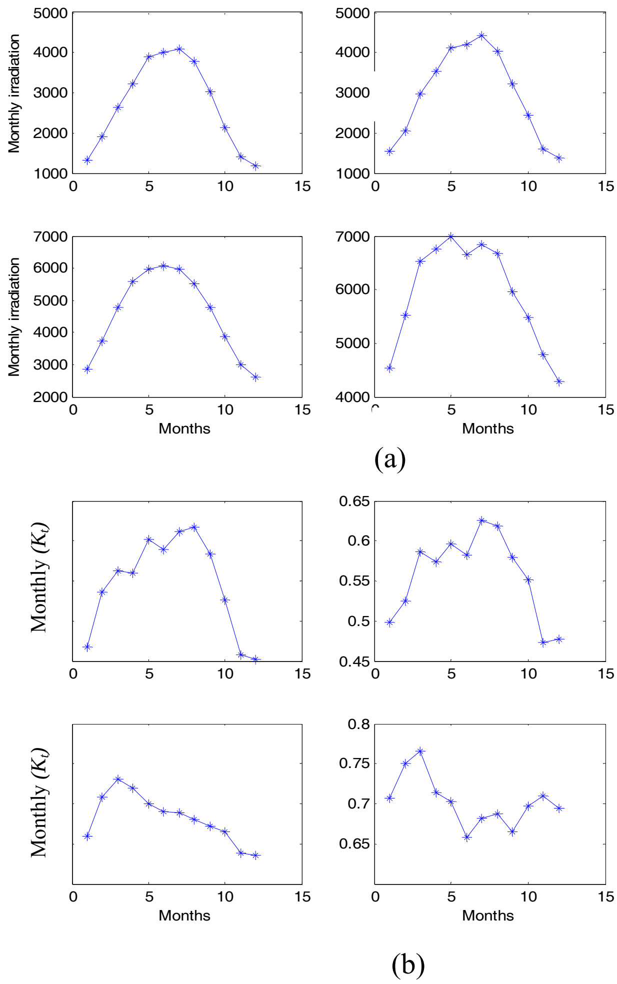


Figure 1 (a): Monthly H (Wh/m²/day), and (b) Monthly K_t for four sites.

Table 1: Database of average total H and K_t.

N°	°	Sites °	m	H Wh/m2/day	K_t
01	36.43N	3.15E	25	4.6884	0.5718
02	35.38 N	3.70E	99	4.6468	0.5599
03	31.38N	2.1E	806	5.8516	0.6768
04	22.47N	5.31E	1378	6.4221	0.6913
05	36.50N	7.49E	4	4.377	0.534
06	24.33N	9.28E	1054	6.6096	0.7207
07	32.23N	3.49E	450	5.7866	0.6748
08	30.34N	2.54E	398	6.1417	0.7034
09	32.45N	0.90E	1072	5.5766	0.6517
10	33.07N	6.04E	69	5.6812	0.6681
11	34.48N	5.44E	81	5.3009	0.5708
12	34.41N	3.15E	1144	5.2566	0.6273
13	33.46N	2.56E	767	5.5118	0.6572
14	31.57N	5.24E	141	5.7116	0.6743
15	27.53N	1.70E	264	6.3545	0.7363
16	27.40N	8.08E	420	6.3151	0.7058
17	27.12N	2.28E	243	6.1814	0.6901
18	30.08N	2.10E	498	6.0168	0.6702
19	26.30N	8.26E	559	5.8433	0.7125
20	36.52N	6.57E	9	4.5705	0.5058
21	36.45N	5.05E	92	4.0668	0.4963
22	36.17N	6.37E	687	4.7917	0.5843
23	36.11N	5.25E	1081	5.207	0.6329
24	35.26N	8.08E	816	4.8053	0.5837
25	35.11N	1.80E	486	4.7022	0.5659
26	33.22N	6.53E	70	5.4549	0.6554
27	28.38N	9.38E	562	5.8828	0.6929
28	34.56N	1.19E	810	4.9304	0.555
29	26.58N	1.05E	290	5.7767	0.6899
30	24.36N	1.14E	347	6.3082	0.6996
31	33.41N	1.01E	1305	5.5333	0.6032

(continue to next page)

Table 1: *(continued)*

N°	°	Sites °	m	H Wh/m2/day	K_t
32	31.40N	6.09E	143	5.7445	0.6779
33	29.15N	1.40E	284	6.1266	0.7087
34	35.33N	6.11E	1040	5.1917	0.5883
35	36.19N	2.14E	750	4.8864	0.5885
36	36.10N	1.21E	112	4.6563	0.5661
37	28.06N	6.49E	381	5.9614	0.7241
38	36.42N	4.30E	9	4.2838	0.4809
39	36.30N	8.23E	4	4.309	0.5257
40	35.17N	1.20E	99	4.6084	0.5553
41	20.10N	4.10E	1351	6.2394	0.7584
42	24.21N	10.0E	1134	6.2365	0.7142
43	23.21N	2.12E	704	6.4563	0.7611
44	29.25N	3.00E	561	6.9742	0.7854
45	29.38N	7.00E	490	6.4531	0.7652
46	28.17N	2.12E	346	6.5369	0.7584
47	26.12N	1.00E	275	5.8356	0.6568
48	30.20N	6.14E	561	6.2565	0.7281
49	31.25N	8.21E	418	5.8365	0.6251
50	30.45N	2.01E	561	6.0662	0.6982
51	33.24N	1.02E	490	6.2254	0.7014
52	32.10N	2.00E	471	5.7848	0.5984
53	32.25N	2.57E	1062	4.9996	0.5114
54	35.42N	7.00E	991	5.1155	0.6025
55	36.74N	3.01E	49	4.5921	0.5145
56	27.41N	2.14E	120	5.9851	0.6251
57	28.35N	5.00E	458	6.3521	0.6351
58	35.00N	1.20E	994	5.9461	0.6581
59	28.70N	1.58E	350	5.8284	0.5896
60	21.50N	3.50E	1151	7.1454	0.7584

3. AN ANFIS

The Neuro-fuzzy modeling[18] involves a way of applying various learning techniques developed in the neural network literature to fuzzy modeling or to a fuzzy inference system (FIS). The basic structure of a FIS consists of three conceptual components: a rule base, which contains a selection of fuzzy rules; a database, which defines the membership functions (MF) used in the fuzzy rules; and a reasoning mechanism, which performs the inference procedure upon the rules to derive an output (Fig. 2). In a situation in which both data and knowledge of the underlying system are available, a neuro-fuzzy approach is able to exploit sources based on network and FL models. The neuro-fuzzy system used here is the ANFIS. The system is an adaptive network functionally equivalent to a first-order *Sugeno* FIS. The ANFIS uses a hybrid-learning rule combining back-propagation, gradient-descent, and a least-squares algorithm to identify and optimize the *Sugeno* system's parameters. The equivalent ANFIS architecture of a first-order *Sugeno* fuzzy model with two rules is shown in Figure 3. The model has five layers and every node in a given layer has a similar function. The fuzzy IF-THEN rule set, in which the outputs are linear combinations of their inputs, is

Rule 1: *if* x *is* A_1 *and* y *is* B_1 *then* f_1: $= p_1x+q_1x+r_1$

Rule 2: *if* x *is* A_2 *and* y *is* B_2 *then* f_2: $= p_2x+q_2x+r_2$

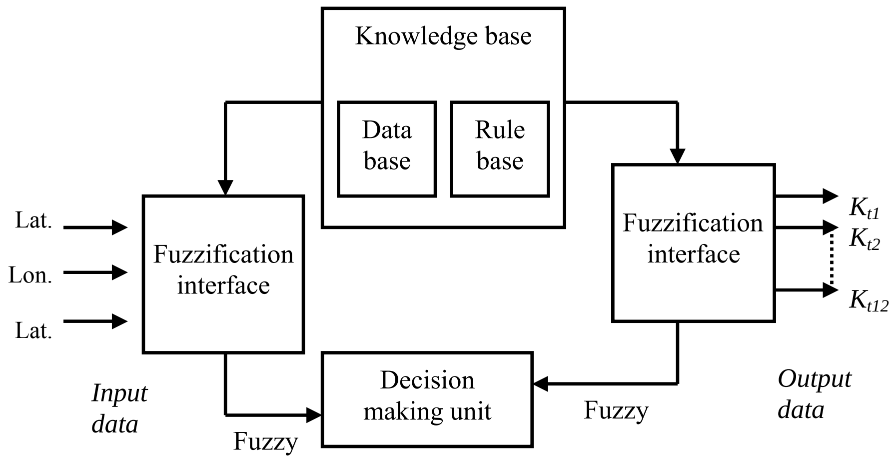


Figure 2: Fuzzy inference system.

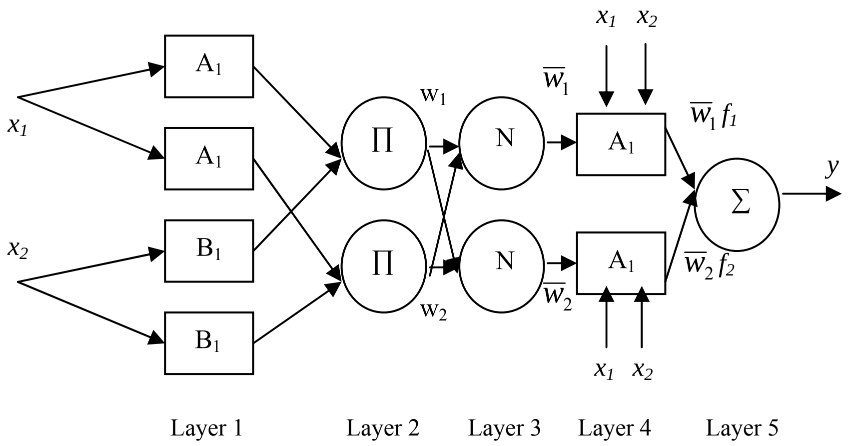


Figure 3: Architecture of an ANFIS equivalent to a first-order *Sugeno* fuzzy model with two inputs and two rules.

Layer 1, consists of adaptive nodes that generate membership grades of linguistic labels based upon premise parameters, using any appropriate parameterized MF such as the generalized bell function:

$$O_{1,i} = \mu_{Ai}(x) = \frac{1}{1 + \left| \dfrac{x - c_i}{a_i} \right|^{2b_i}} \tag{1}$$

where output $O_{1,i}$ is the output of the i^{th} node in the first layer, is the input to i, Ai node , is a linguistic label ("small," "large," etc.) from fuzzy set $A = (A_1, A_2, B_1, B_2)$ associated with the node, and $\{a_i, b_i, c_i\}$ is the premise parameter set used to adjust the shape of the MF. The nodes in layer 2 are fixed nodes designated $\prod$, which represent the firing strength of each rule. The output of each node is the fuzzy AND (product, or MIN) of all the input signals.

$$O_{2,i} = w_i = \mu A_i(x) \mu B_i(y) \tag{2}$$

The outputs of layer 3 are the normalized firing strengths. Each node is a fixed rule labelled N. The output of the i^{th} node is the ratio of the i^{th} rule's firing strength to the sum of all the rules firing strengths:

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2} \tag{3}$$

The adaptive nodes in layer 4 calculate the rule outputs based upon consequent parameters using the function:

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i \ (p_i x + q_i y + r_i) \tag{4}$$

where w_i is a normalized firing strength from layer 3, and (p_i, q_i ,r_i) is the consequent parameter set of the node. The single node in layer 5, labelled ∑, calculates the overall ANFIS output from the sum of the node inputs:

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{5}$$

Training the ANFIS is a two-pass process over a number of epochs. During each epoch, the node outputs are calculated up to layer 4. At layer 5, the consequent parameters are calculated using a least-squares regression method. The output of the ANFIS is calculated and the errors propagated back through the layers in order to determine the premise parameter (layer 1) updates.

4. MODEL DEVELOPMENT AND TESTING

The described ANFIS model is adopted and used for predicting the $\overline{K}_t$ in isolated sites. The block diagram of the proposed model is presented in Figures 4(a) and (b). The inputs of the model are the geographical coordinates of the site (altitude, longitude and latitude), while the outputs of the model are the 12-values corresponding to the $\overline{K}_t$. The input and the output of the model are fuzzified before used.

Figure 5 shows the initial MF for each input data of the ANFIS. When the data are fuzzified into class, a total of 56-patterns have been used for training the model and 4-patterns have been used for testing the model. Therefore the testing sites are selected randomly.

Figure 6 illustrates the evolution of the RMSE for the different networks [MLPN, RBFN, Recurrent Neural Network (RNN)] and the proposed ANFIS. In order to test the performance of the model, we have plotted a cumulative function between measured $\overline{K}_t$ and predicted monthly clearness index ($\hat{\overline{K}}_t$) as presented in Figure 7. From observation of these curves, we note that there is no important difference between measured and predicted $\overline{K}_t$ for each site. Table 2 summarizes

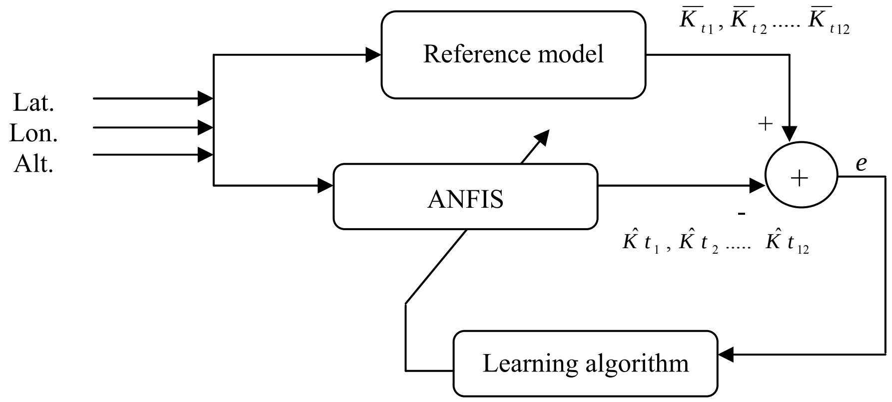


Figure 4(a): Block diagram of the developed model.

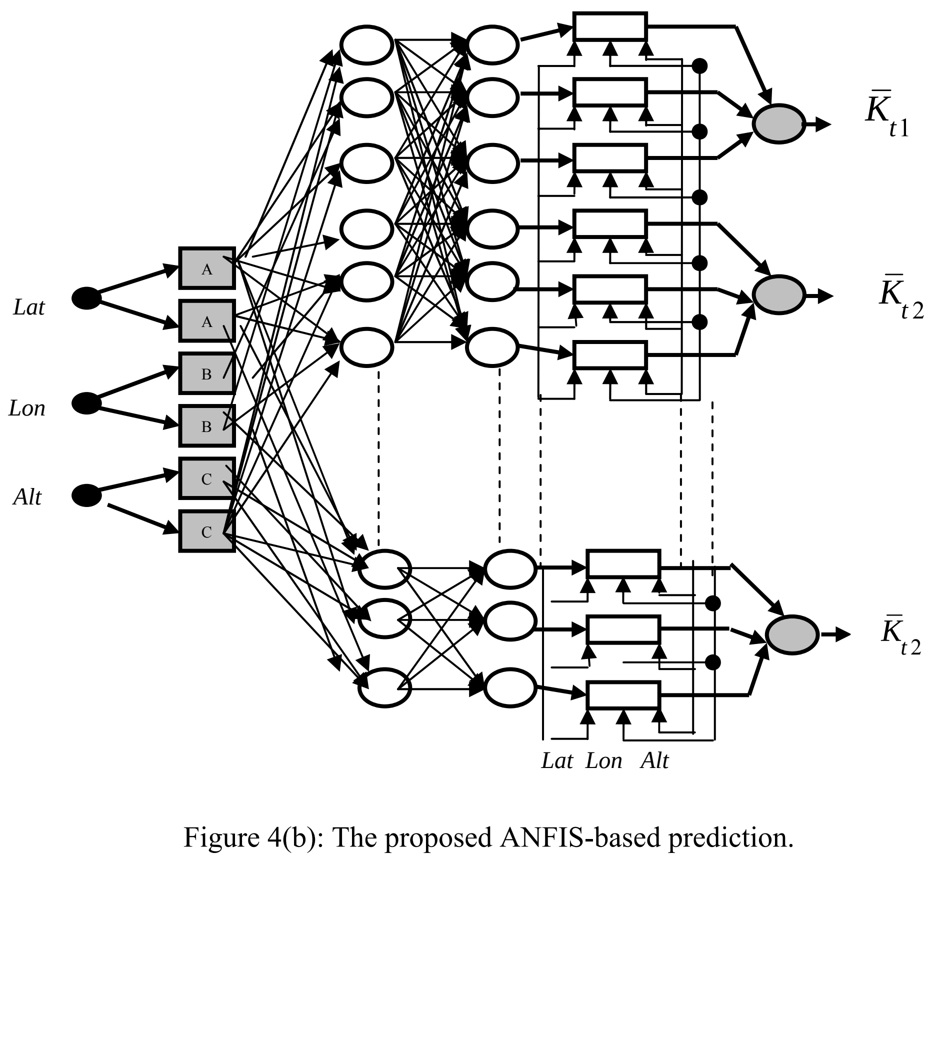


Figure 4(b): The proposed ANFIS-based prediction.

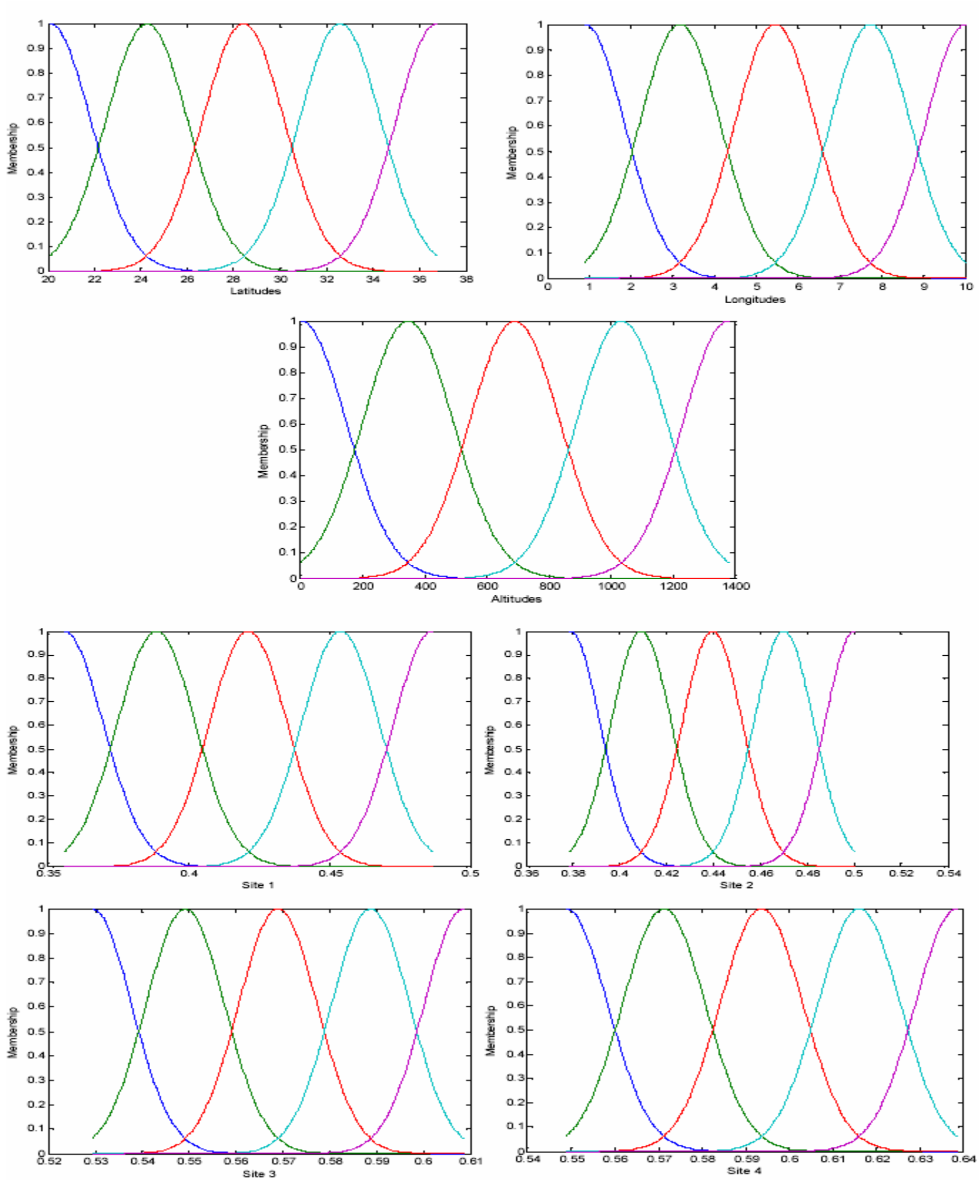


Figure 5: The initial division of input and output spaces into five fuzzy regions and their corresponding Gaussian MF.

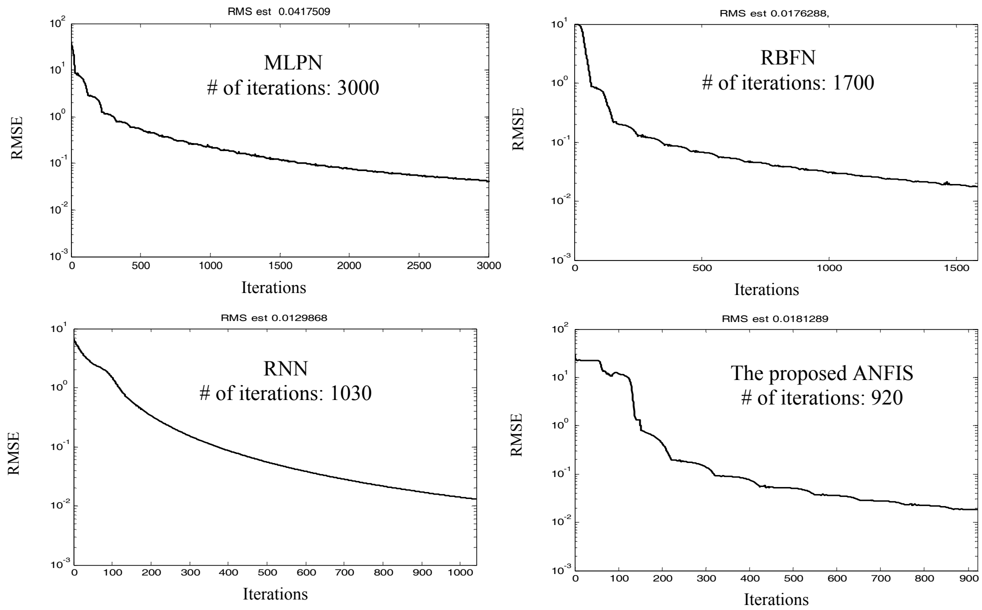


Figure 6: RMSE for the different ANNs used in this simulation and the proposed ANFIS.

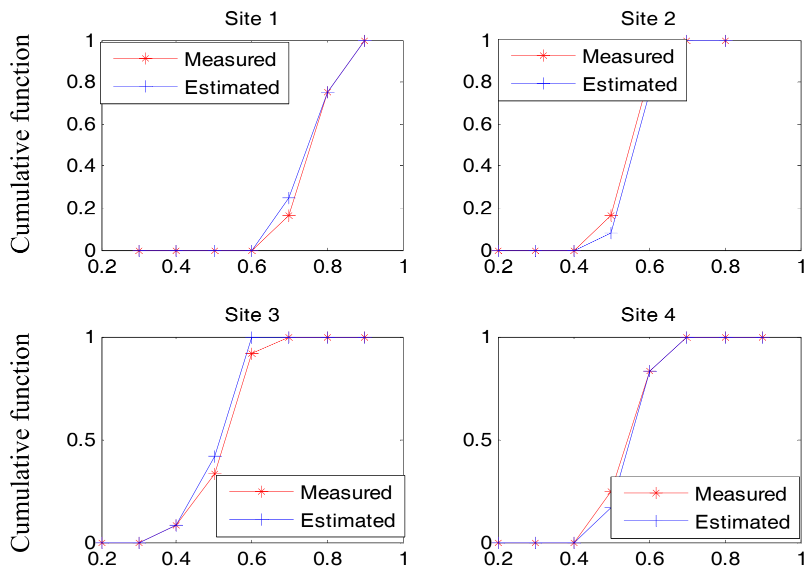


Figure 7: Cumulative function for four tested sites.

the mean, variance, ANFIS Kolmogorov-Test (*KS*) and RMSE between measured $\hat{\bar{K}}_t$ and $\hat{\bar{K}}_t$. Generally from the statistical point of view, the results are very satisfactory.

In order to assess its performance relative to different ANN architectures (MLPN, RBFN and RNN) we have plotted the estimated $\hat{\bar{K}}_t$ by the different ANN and the proposed ANFIS (Fig. 8) for one selected site. According to this curve, we remark that the ANFIS and the RNN gave good results compared to those obtained by MLPN and RBFN.

Table 2: Statistical tests.

Sites (geographical coordinates) (°,')	(°,')	m	Measured Mean $\bar{K}_t$	Predicted Mean $\hat{\bar{K}}_t$	Variance σ	*KS*	RMSE
27,12 N	2.28E	243	0.758	0.721	0.0391	0.068	0.0214
36,17 N	6.37E	687	0.463	0.484	0.0418	0.062	0.0221
35,17 N	1.20E	99	0.509	0.524	0.0387	0.065	0.0245
35,33 N	6.11E	1040	0.557	0.548	0.0374	0.059	0.0235

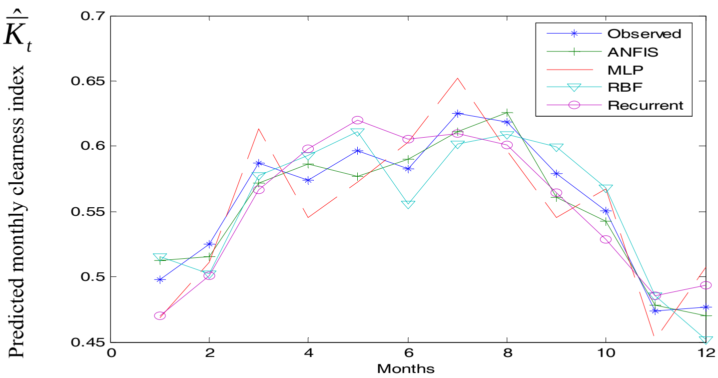


Figure 8: Comparison between different ANN architectures and the proposed ANFIS predicting $\hat{\bar{K}}_t$.

Table 3 presents a comparison for the MRE, RMSE and number of iteration, between different ANNs structures and the ANFIS-model developed in this work. From the comparison, it is clear that the ANFIS-model developed in this work has the best convergence time and of the number iteration of 920 and the MRE of 2.2%.

Table 3: MRE between the different ANNs and the proposed ANFIS.

Neural network architecture	Predicted $\hat{\bar{K}}_t$	Number of iterations	MRE (%)
MLPN	0.5526	3000	8.1
RBFN	0.5542	1700	6.3
RNN	0.5556	1030	3.2
ANFIS-model developed in this work	0.5561	920	2.2
Measured K_t	0.5571		

5. APPLICATION FOR SIZING PV SYSTEMS

In this section, we present an example for sizing PV system based on the predicted data proposed by our ANFIS model. Firstly $\hat{\bar{K}}_t$ corresponding to 4-locations have been used for generating sequences of daily total H, based on the MTM method proposed by Aguiar et al.[5] (Appendix 1[17]). Several models have been developed in the literature in order to find the optimal sizing of PV system based on numerical (Appendix 2[17]), analytical and hybrid approaches.[18–23] The construction of a sizing curve based on the Loss Load Probability (LLP) requires the modeling of PV system operation over substantial periods of time. Time series of solar radiation then cannot come directly from observation but need to be reproduced ''synthetically'' based on an algorithm which is faithful to the solar radiation statistics. The relationship between the LLP values and the perceived reliability requirements of the user are then indirect, although generally accepted correspondence exist for most standard applications.[3,19] Secondly, based on the numerical method[19] and the proposed hybrid approach (ANN-GA),[24] we can determine the optimal sizing surface of PV-generator (A_{PV}) and storage batteries (C_U) in order to satisfy a given (L) consumption, for each 4-locations used in this simulation. A 10-year daily H has been generated based on the ANFIS-model proposed as shown in Figure 9.

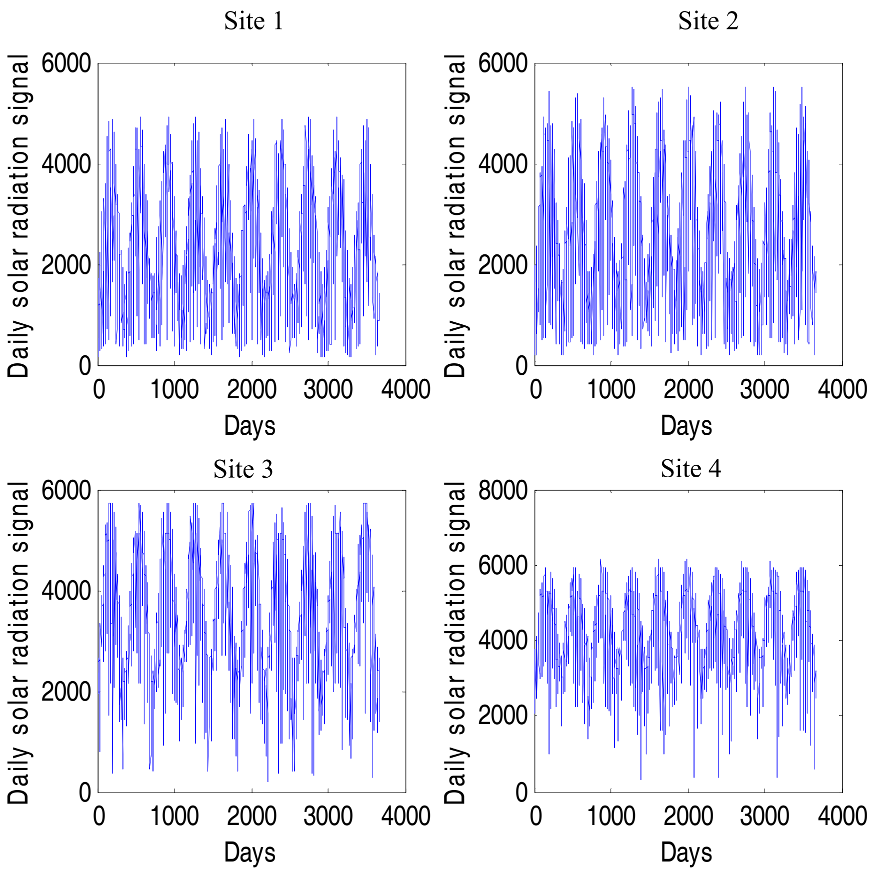


Figure 9: Sequences of daily H obtained from the $\hat{\bar{K}}_t$ based on the proposed ANFIS and MTM approach corresponding to 10-years.

Figures 10(a), (b), (c) and (d) summarize the histogram and the MRE of the sizing parameters based on measured daily *H* and estimated by the different ANN architectures and the proposed ANFIS.

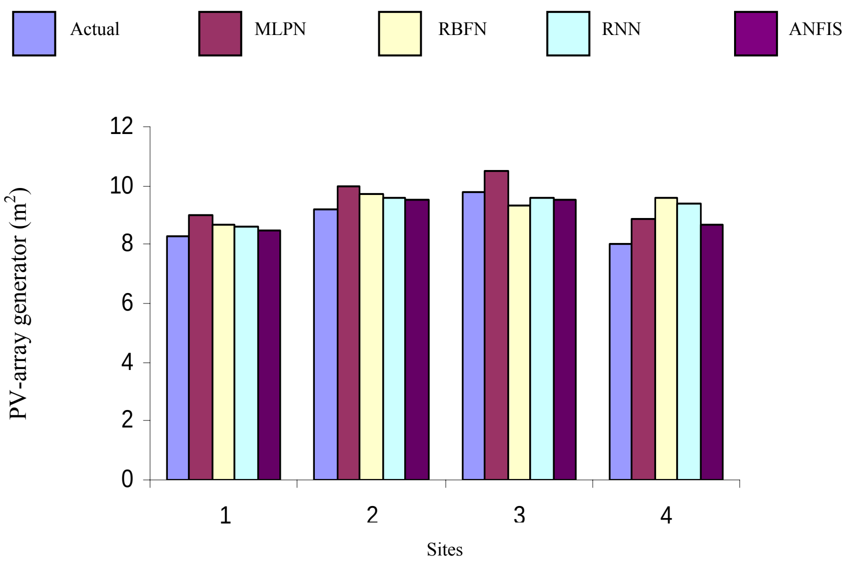


Figure 10(a): Comparison between actual PV-array measured and estimated by the different ANN and the proposed ANFIS.

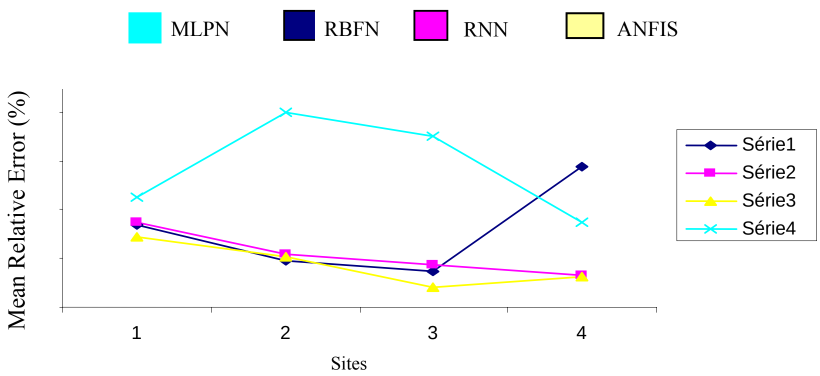


Figure 10(b): MRE.

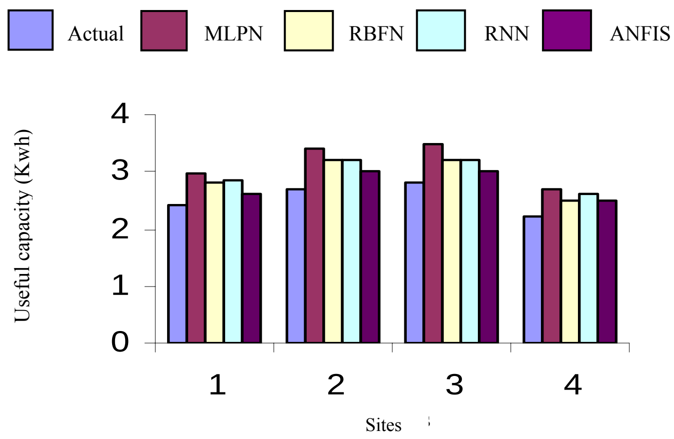


Figure 10(c): Comparison between actual useful capacity measured and estimated by the different ANN and the proposed ANFIS.

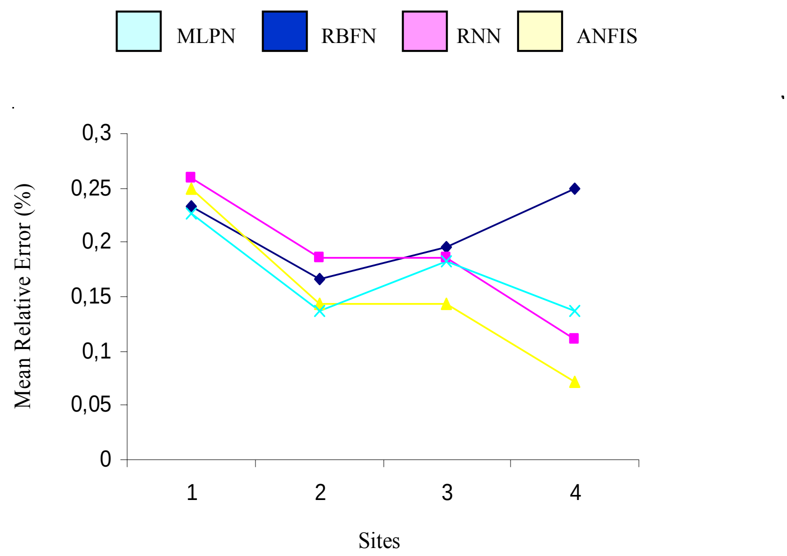


Figure 10(d): MRE

According to these curve, we observe that there is a good correlation obtained by all ANN models used. However, the proposed method present more satisfactory results compared to the reported ANN.[17] In addition, the MRE does not exceed 0.2%.

6. CONCLUSION

This paper reports a proposal on an ANFIS for predicting $\bar{K}_t$ in isolated locations. The proposed model has been applied and tested in Algerian locations. The results obtained allow us to conclude that the ANFIS is effective compared to the reported ANN architectures (MLPN, RBFN and RNN). The advantage of the model is that it can estimate K_t from only the geographical coordinates of the site, without having to resort the traditional ambient parameters such as: mean temperature, sunshine duration, wind speed, and etc. In addition the convergence time and the MRE are improved. Thus, having obtained the $\bar{K}_t$ based on the MTM method, the ANFIS-model can generate sequences of daily solar radiation over an extended period.

The number of sites used together with their geographical range allow us to conclude that the proposed ANFIS-model is generally valid for estimating sequences of daily total H in latitudes ranging from 21° 0'N to 36° 5'N and the longitudes ranging from 1° 0' to 9° 5'. These data is required for sizing of the PV system. The application of sizing PV systems shows clearly the advantage of the proposed model to the alternative ANN architectures.

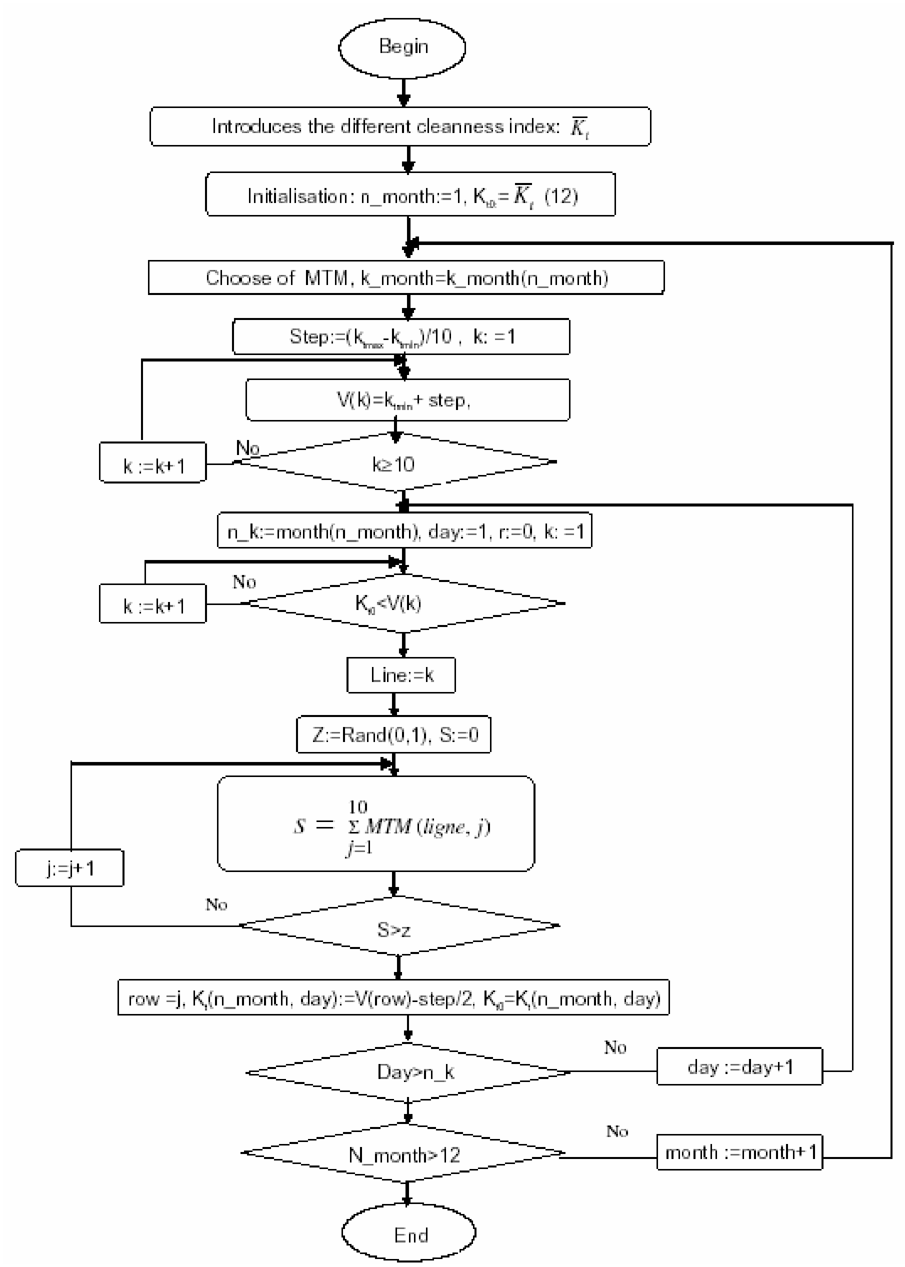
The results have been obtained for the Algerian locations, but the methodology can be generalized for use in other parts of the world. In addition, the proposed technique can be extended to any meteorological data, e.g. wind, humidity, temperature, and etc.

7. ACKNOWLEDGEMENT

The authors would like to thank the Director of the ONM (Office of National Meteorology of Algiers), for making available the database of solar radiation data for different sites, and Prof. A. Guessoum (Head of Signal Processing Laboratory of Blida University) for his remarks.

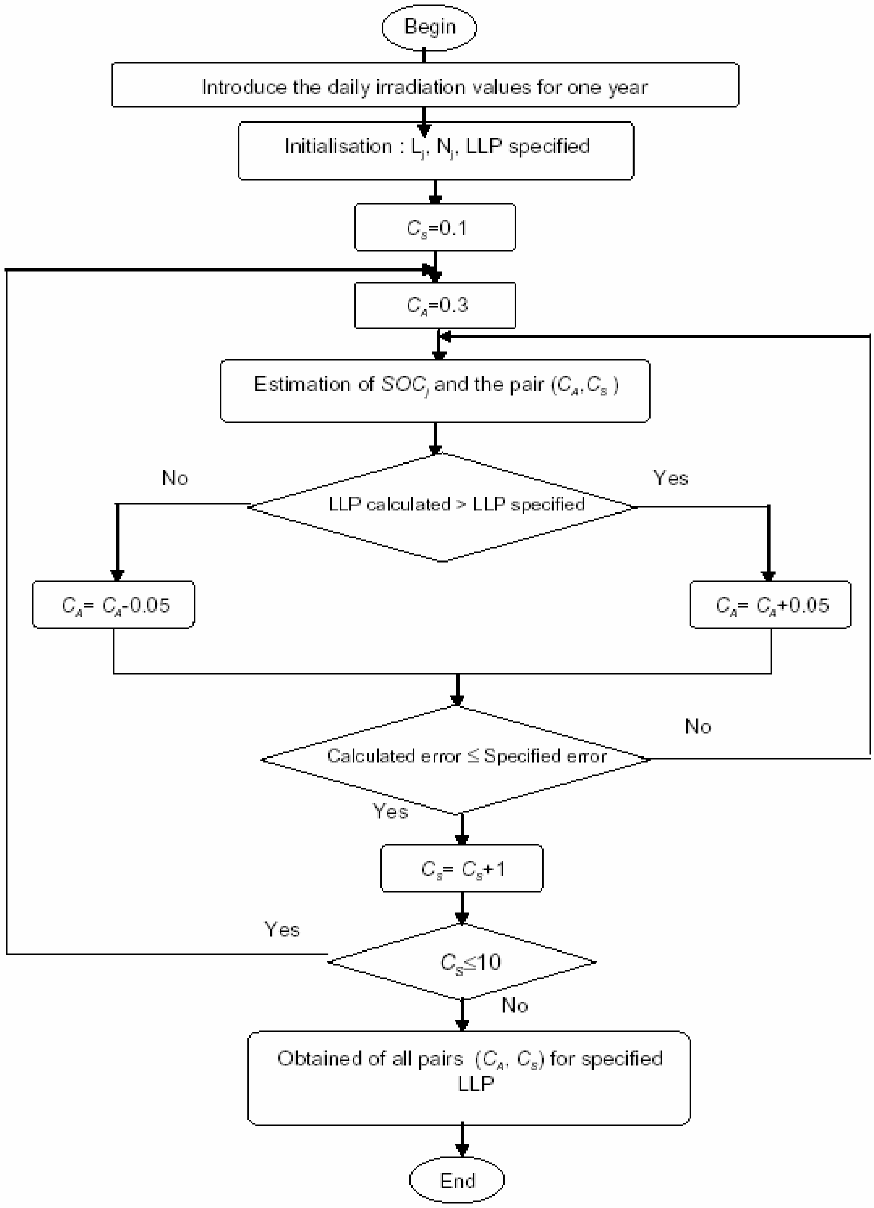
Appendix 1

MTM procedure for generating sequences of daily clearness index



Appendix 2

Numerical procedure for construction LLP-curve



8. REFERENCES

1. Chapman, R.N. (1990). The synthesis solar radiation data for sizing stand-alone PV system. *In Photovoltaic Specialist Conference 1990, Conference Record of the Twenty First IEEE,* 21–25 May 1990, Kissimmee, FL, USA.
2. Mellit, A., Benghanem, M. & Bendekhis, M. (2005). Artificial neural network model for prediction solar radiation data: Application for sizing stand-alone photovoltaic power system. *In Power Engineering Society, General Meeting*, *IEEE*, June 12–16, 2005, USA.
3. Markvart, T., Fragaki, A. & Ross, J.N. (2006). PV system sizing using observed time series of solar radiation. *Solar Energy*, 80, 46–50.
4. Mellit, A. (2006). *Artificial intelligence techniques for sizing and simulation of photovoltaic system.* PhD Thesis, University of Sciences and Technology Houari Boumediene, Algiers, Algeria.
5. Aguiar, R.J., Collares-Perrira, M. & Conde, J.P. (1988). Simple procedure for generating sequences of daily radiation values using library of Markov transition matrices. *Solar Energy*, 40, 269–279.
6. Aguiar, R.J. & Collares-Pereira, M. (1992). TAG: A time-dependent autoregressive: Gaussian model for generating synthetic hourly radiation. *Solar Energy*, 49, 167–174.
7. Mora Lopez, L. & Sidrach-de-Cardona, M. (1998). Multiplicative ARMA models to generate hourly series of global irradiation. *Solar Energy*, 63, 283–291.
8. Santos, J.M., Pinazo, J.M. & Canada, J. (2003). Methodology for generating daily clearness index values K_t, starting from the monthly average daily value K_t: Determining the daily sequence using stochastic models. *Renewable Energy*, 28, 1523–1544.
9. Guessoum, A., Boubkeur, S. & Maafi, A. (1998). A global-irradiation model using radial basis function neural network. *In Proc. of the 5th WREC*, Florence, Italy, September 20–25 Part IV. Oxford (UK): Elsevier Sciences Ltd., 2533–2536.
10. Shuanghua, C. & Jiacong, C. (2005). Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. *Applied Thermal Engineering*, 25, 161–172.
11. Mellit, A., Benghanem, M. & Kalogirou, S.A (2006). An adaptive wavelet-network model for forecasting daily total solar radiation, *Applied Energy*, 83, 705–722.
12. Mellit, A. (2007). An ANFIS-based forecasting for solar radiation data from sunshine duration and ambient temperature. *Accepted paper in IEEE Power Engineering Society General Meeting.*

13. Yorukoglu, M. & Celik, A.N. (2006). A critical review on the estimation of daily global solar radiation from sunshine duration. *Energy Conversion and Management*, 47, 2441–2450.
14. Mohandes, M., Balghonaim Kassas, A., Rehman, M.S. & Halawani, T.O. (2000). Use of radial basis functions for estimating monthly mean daily solar radiation, *Solar Energy*, 68, 161–168.
15. Hontoria, L., Aguilera, J. & Zufiria, P. (2005). An application of the multilayer perceptron: Solar radiation maps in Spain. *Solar Energy,* 79, 523–530.
16. Sozena, A., Arcaklyogolub, E., Zalpa, M.O. & Agolarc, N.C. (2005). Forecasting based on neural network approach of solar potential in Turkey. *Renewable Energy*, 30, 1075–1090.
17. Mellit, A., Benghanem, M., Arab, A.H. & Guessoum, A. (2005). A simplified model for generating sequences of global radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach. *Solar Energy*, 79(5), 468–482.
18. Jang, J.S.R. (1993). ANFIS: Adaptive network-based fuzzy inference system. *IEEE Trans. Syst., Man, Cybern*., 23, 665–685.
19. Groumpos, P.P. & Papageorgiou, G. (1987). An optimal sizing method for PV power system. *Solar Energy*, 38, 314–351.
20. Egido, M. & Lorenzo, E. (1992). The sizing of sand-alone PV systems: A review and a proposed new method. *Solar Energy Materials and Solar Cells*, 26, 51–69.
21. Shrestha, G.B. & Goel, L. (1998). A study on optimal sizing of stand-alone photovoltaic stations. *IEEE Transactions on Energy Conversion*, 13(4), 373–378.
22. Sidrach-de-Cardona, M. & Mora López, L. (1998). A simple model for sizing stand alone photovoltaic systems. *Solar Energy Materials and Solar Cells*, 55, 199–214.
23. Agha, K.R. & Sbita, M.N. (2000). On the sizing parameters for stand-alone solar-energy systems. *Applied. Energy*, 65(1–4), 73–84.
24. Mellit, A., Benghanem, M., Arab, A.H & Guessoum, A. (2005). An adaptive artificial neural network model for sizing stand-alone photovoltaic systems: Application for isolated sites in Algeria. *Renewable Energy*, 80, 1501–1524.