REPRESENTATION OF SOLAR COLLECTORS VIA NEURAL NETWORKS WITH REDUCTION OF THE TRAINING SET NUMBER

Elizabeth Marques Duarte Pereira
Energy Researches Group (GREEN)
Pontifical Catholic University of Minas Gerais,
500 Dom José Gaspar Ave., Coração Eucarístico
Belo Horizonte, MG, Brazil, 30535-610
green@pucminas.br

Luis E. Zárate
Applied Computational Intelligence Laboratory (LICAP)
Pontifical Catholic University of Minas Gerais
500 Dom José Gaspar Ave., Coração Eucarístico
Belo Horizonte, MG, Brazil, 30535-610
zarate@pucminas.br

Renato Vimieiro
Applied Computational Intelligence Laboratory (LICAP)
Pontifical Catholic University of Minas Gerais
500 Dom José Gaspar Ave., Coração Eucarístico
Belo Horizonte, MG, Brazil, 30535-610

Antônia Sônia Cardoso Diniz
Energy Company of Minas Gerais (CEMIG)

Abstract. *Normally efficiency measurements of solar collectors are made by means of experiments that use as operational parameters: solar irradiance, flow rate, and input and output water temperatures. The efficiency depends on some other aspects too, as of its structural aspects, material of its components, thermal insulation and position.*
For new operating conditions, new experiments are necessary to calculate efficiency. Linear regression has been proposed by several authors as a way of modeling solar collector. The linear regression may introduce significant errors when used with that purpose, due to its limitation of working better only with linear correlated values.
Due to their facility of solving nonlinear problems, ANN (Artificial Neural Networks) are presented here as an alternative to represent these solar collectors with several advantages on other techniques of modeling like linear regression. ANN correctly trained do not require new experiments or linear correlated values to obtain the output water temperature in new conditions and to, consequently, calculate the efficiency. This work tries to demonstrate that ANN are capable of estimating output temperatures with more precision. Techniques for selecting representative training sets are also discussed and presented in this paper.

Keywords*: solar energy, solar collector, neural networks*

1. Introduction

Due to the hydrographic basins situation around the world, associated with a constant population increase, a new reality can be observed, where alternative forms of energy production have essential importance. Solar energy systems are one of these forms.

Solar energy systems, specifically water heaters, have considerable importance as substitute of traditional electrical systems. In Fig. 1, an example of water heater, called thermosiphon systems, is schematically represented. The solar collector (collector plate) is the most important component in a thermosiphon system.

The performance of thermosiphon systems has been investigated, both analytically and experimentally, by numerous researches (Morrison and Ranatunga, 1980; Huang, 1984 and Kudish et. al. 2000). The equation used to calculate the solar collector efficiency is the following:

$$\eta = \frac{\dot{m} c_p (T_{out} - T_{in})}{G A_{extern}} \tag{1}$$

In the equation above, η is the thermal efficiency, $\dot{m}$, the mass flow rate, c_p, the specific heat of water, T_{out}, the output water temperature, T_{in}, the input water temperature, G, the solar irradiance and A_{extern}, the area of the collector.

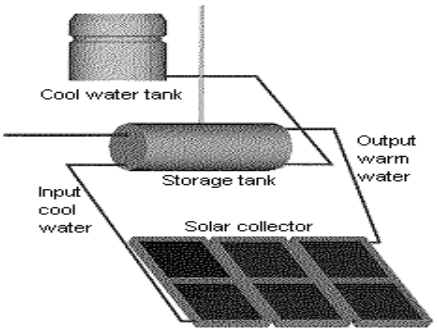

Figure 1. Schematic diagram of a thermosiphon system

Mathematical models (Morrison and Ranatunga, 1980; Huang, 1984; Kudish et. al. 2000; Kalogirou et. al. 1999 and Kalogirou 2000) have already been presented as a way of calculating the efficiency of solar collectors; however, the non-linearity nature of those models makes their application discouraging. Linear regression (Kudish et. al. 2000) has been proposed as a way of modeling solar collectors, instead of those complex mathematical models. But linear regression may introduce significant errors when used with that purpose, due to its limitation of working better only with linear correlated values.

In the last years, ANN (i.e. Artificial Neural Networks) have been proposed as a powerful computational tool. Some researches Kalogirou et. al. (1999) have already discussed their use in the representation of thermosiphon systems. ANN have several advantages on other techniques, including their performance when dealing with nonlinear problems, their capacity of learning and generalizing, the low time of processing that can be reached when trained nets are in operation etc.

This paper is organized in four sections. In the second one, solar collectors are physically described. In the third section, the use of ANN to represent solar collectors and statistical techniques to select training sets are discussed. And in the fourth section, conclusions are presented.

2. Physical description of the solar collector

The working principles of thermosiphon systems are based on thermodynamic laws (Duffie and Beckman 1999). In those systems, water circulates through the solar collector due to the natural density difference between cooler water in the storage tank and warmer water in the collector. Although they demand larger cares in their installation, thermosiphon systems are of extreme reliability and lower maintenance. Their application is restricted to residential, small commercial and industrial installations.

Solar irradiance reaches the collectors, which heat up the water inside them, decreasing the density of heated up water. Thus, the cooler and denser water forces the warmer water to the storage tank. The constant water flow that happens between the storage tank and the collector results in a natural circulation called “thermosiphon effect”.

In a very similar way to the one that happens during the day, thermosiphon tends to accomplish the opposite process during the night, cooling water. In order to avoid this opposite effect, a minimum height between the collector and the storage tank is expected to be established. This height is typically 30cm.

3. Neural representation of the solar collector

Multi-layer ANN have been used in this work. The values of entries are presented to the hidden layer and satisfactory answers are expected to be obtained from the output layer. The most suitable number of neurons in the hidden layer is still a non-solved problem, although researches discuss some approaches. In Kóvacs (1996), the suggested number of hidden neurons is $2n+1$, where n is the number of entries. On the other hand, the number of output neurons equals the number of expected answers from the net.

Input water temperature (T_{in}), solar irradiance (G) and ambient temperature (T_{amb}) are variables used as entries to the ANN. The output water temperature (T_{out}) is the wanted output from the net. In this work, ANN represent the thermosiphon system according to the following equation

$$f(T_{in}, T_{amb}, G) \xrightarrow{ANN} T_{out} \tag{2}$$

The structure of the ANN in this work is schematically represented as shown in the Fig. 2. The net contains seven hidden neurons (i.e. $2n+1$) and one neuron in the output layer, from which the output water temperature is obtained.

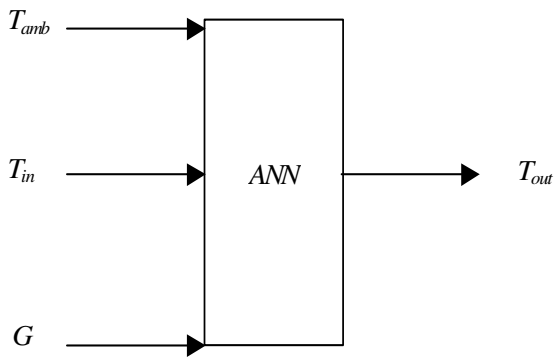

Figure 2. Schematic diagram of the net used in this work

Supervised learning has been adopted to train the net. As training algorithm, the backpropagation algorithm was used. Nonlinear sigmoid function has been chosen in this work as the axon transfer function:

$$f = \frac{1}{1 + \exp^{-\sum Entries \text{ x Weigths}}} \tag{3}$$

Backpropagation algorithm uses a training set, from where inputs and wanted outputs may be extracted. For each parameter of the chosen training set, the weights of the net are adjusted in order to minimize errors obtained in the output values.

3.1. Collecting data from the solar collector

Data have been collected to a typical solar collector by means of experiments under specific standards ASHRAE (1986), during three days in different ambient situations. Fig. 3 shows the relation between the output water temperature (T_{out}) and the hours of collecting (*hours*). It can be observed that collected data cover several situations.

In order to verify the non-linearity of the data, Fig. 4, Fig. 5 and Fig. 6 have been built. However those graphics demonstrate that there is linearity in the collected data. Although linear regression gives a valid approach, this work tries to demonstrate that ANN are capable of estimating output temperatures with more precision.

The total number of collected data equals 633 (a sample of those data is shown in Tab. I.1, in the appendix). A subset containing 30 data has been removed from the initial set with the purpose of being used in the validation process of the net, as shown in the next sections. 603 data still remain in the training set.

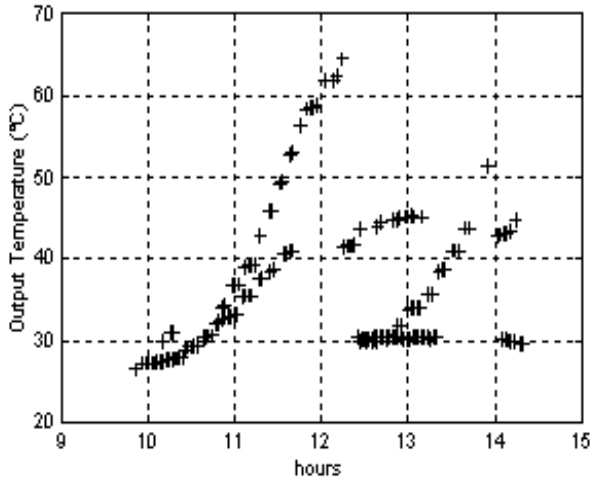

Figure 3. Collected water output temperatures

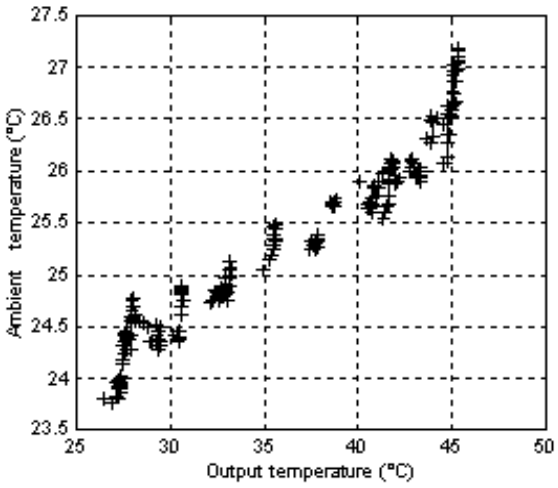

Figure 4. *Tamb x Tout*

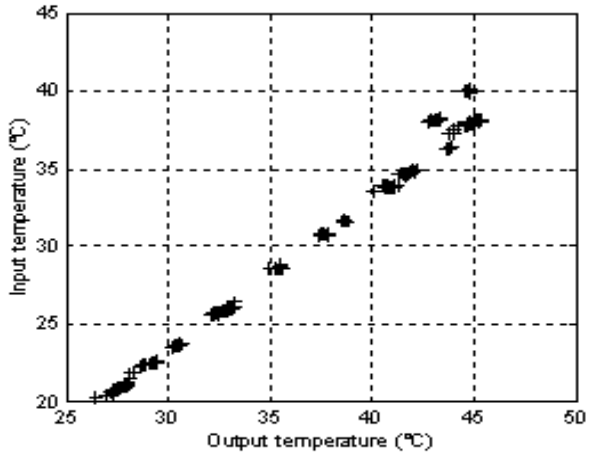

Figure 5. *Tin x Tout*

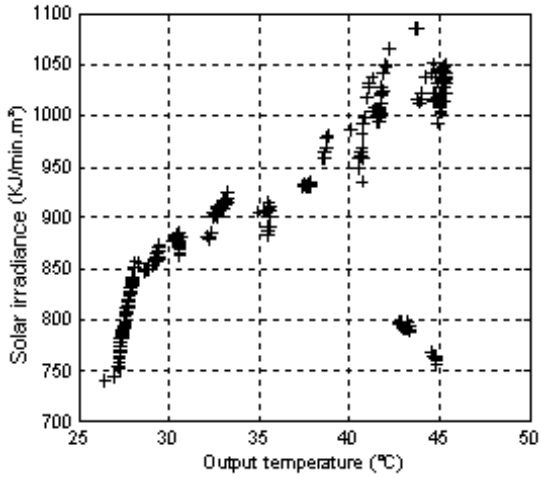

Figure 6. *Solar irradiance X Tout*

3.2. Preparing data for training

Pre-processing input data is a process of considerable importance for the performance of ANN. In this work, the following procedure has been applied to collected data, before the use of them in the net structure:

1. Data values have been normalized in order to be within the interval [0.2, 0.8].

2. The following equations have been used in the normalization process:

$$f^{a}(\mathrm{L}o) = \mathrm{L}n = (\mathrm{L}o - \mathrm{L}mín) / (\mathrm{L}\max - \mathrm{L}\min) \tag{4}$$

$$f^{b}(\mathrm{L}n) = \mathrm{L}o = \mathrm{L}n * \mathrm{L}\max + (1 - \mathrm{L}n) * \mathrm{L}mín \tag{5}$$

The equations above must be applied to each variable of the training set (e.g. T_{amb}, T_{in}, G), normalizing all their values.

3. *Lmin* and *Lmax* have been computed as follows

$$L_{\min} = L_{\sup} - (N_s / (N_i - N_s)) * (L_{\inf} - L_{\sup}) \tag{6}$$

$$L_{\max} = ((L_{\inf} - L_{\sup}) / (N_i - N_s)) + L_{\min} \tag{7}$$

where L_{sup} is the maximum value of that variable, L_{inf} is its minimum value, N_i and N_s are the limits for the normalization (in this case, $N_i = 0.2$ and $N_s = 0.8$).

3.3. Selecting data for training

The actual training set contains 603 data, but a reduction in its size may be beneficial, decreasing the time spent in the training process and also maintaining the capacity of generalization of the trained net. The literature (Bittencout and Zárate, 2002) suggests some techniques to build small better-defined training sets. Even if smaller, trainings sets must cover all the possible situations that may happen in the problem or, at least, a major part of them.

The following equation has been borrowed from statistics area and has been used in this work to calculate suitable sizes of training sets

$$n = \left(\frac{z}{e}\right)^2 * (f * (1 - f)) \tag{8}$$

where n is the size of the set, z is the confidence level, e is the error around the average and f is the population proportion.

Fixing z in 90%(z = 1.645), f in 0.5 and varying e from 0.04°C to 0.1°C, several sizes of training sets have been calculated (Tab. 1). It is extremely important to emphasize that the value of errors considered here are not the maximum errors of the net.

Table 1. Calculated set sizes

Value e	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Set size	423	271	188	139	106	84	68

The following procedure has been adopted to build training sets with each one of the calculated sizes, selecting the necessary quantities of data among the 603 data:

1. For each variable of the net, the maximum and the minimum values found in the set have been selected.

2. The operation point of the problem (i.e. the most representative parameter in the set) has been chosen and added to the new training set.

3. The remaining elements have been randomly selected, with the purpose of reaching the calculated size for each training set.

Eight training sets of different sizes (seven presented in Tab. 1 and the original one of 603 data) have been built. The most representative of them must be chosen.

3.4. Training process

Several nets have been trained with each one of the different training sets. The net has been configured with 7 neurons in the hidden layer, learning rate equivalent to 0.08, maximum error of 0.05 and all the initial weights on zero. The necessary number of iterations to train each of the nets is presented in Fig. 7.

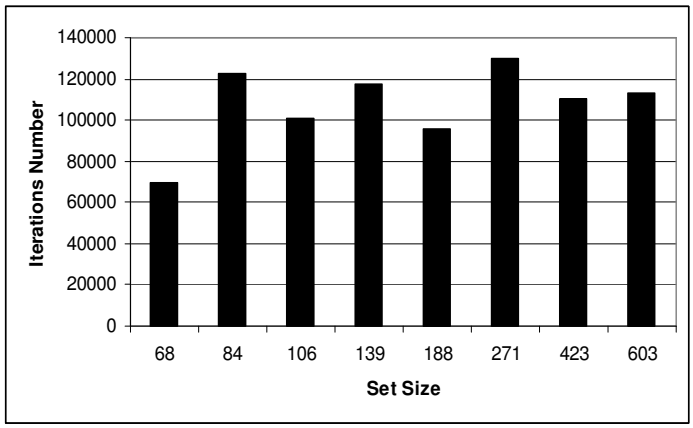

Figure 7. Number of iterations in the training of each net

Table 2 presents the average errors obtained in the validation process of each trained net. For the validation process, the 30 sets previously separated data have been used.

Table 2. Average error of the validation process of each trained net

Size set	68	84	106	139	188	271	423	603
Error	0.8544	0.6005	0.7367	0.5616	0.7328	0.5831	0.6674	0.6213

The training set composed by 84 data has been chosen, because the net trained with it maintains a satisfactory capacity of generalization (better than the one trained with 603 data), despite of the fact of spending a larger number of iterations.

In order to verify the performance of the chosen net, another net has been trained with 84 training sets and with the same initial configuration. However, the maximum tolerable error has been changed and now it is equivalent to 1°C (0.016 in normalized values), and the weights have been initialized with random values instead of a constant zero value. Fig. 8 shows the results of that new training.

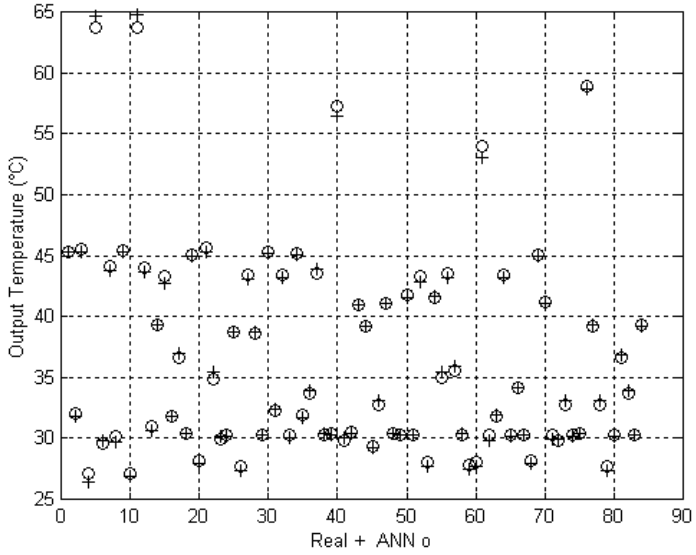

Figure 8. *Real (o)* and *ANN (+)* output temperatures

3.5. Validation process

Table I.3 (in the appendix) shows the data set used to validate the trained net and also shows the output of the net trained with 84 data and its errors. A comparison between validation processes involving nets trained with 84 and 603 data is shown in Tab. 3. In Fig. 9, the results of the validation process are graphically shown.

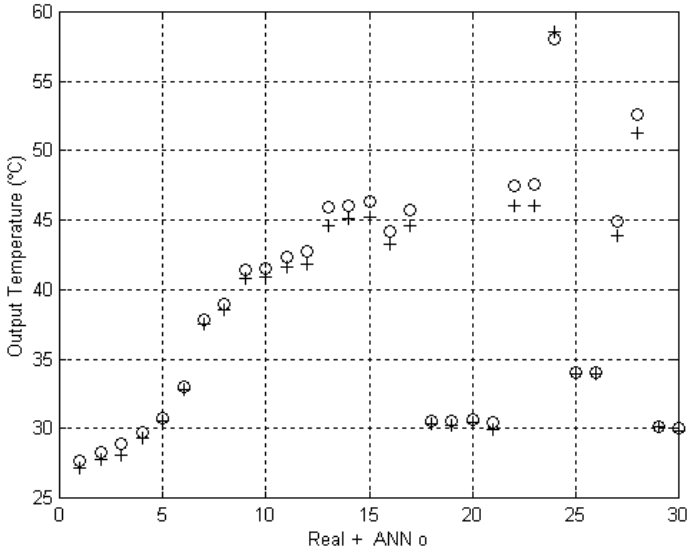

Figure 9. Results of the validation process of the net trained with 84 data

Table 3. Comparison of errors obtained in the process of validation for different nets

	Error (°C) with 84 sets	Error (°C) with 603 sets
Minimum	0.043265	0.02185
Maximum	1.475292	0.70706
Average	0.625579	0.27365

Error values obtained in the validation process of the net trained with 84 data are greater than the ones of the net trained with 603 data; however, its average error is lower than 1°C, as recommended by INMETRO (National Institute of Metrology and Industrial Quality - Brazil). Above all, the number of iterations in the training with 84 and 603 data, with the last configuration, are, respectively, 412800 and 7.7 million.

3.6. Verification of results

For the analysis by means of linear regression, Equation (9) has been used:

$$\eta = F_{R}(\tau\alpha)_{e} - F_{R}U_{L}\frac{(T_{in} - T_{amb})}{G} \tag{9}$$

$F_R(\tau\alpha)_e$ equals 66.662 and F_RU_L, 809.89. F_R corresponds to collector heat removal factor, $(\tau\alpha)_e$, to transmittance absorptance product and U_L, to collector overall loss coefficient. T_{in} is the input water temperature, T_{amb}, the ambient temperature and G, the solar irradiance. Equation (9) calculates efficiency when linear regression is used.
The solar collector efficiency can be calculated by means of ANN since the output water temperature is obtained. Once trained, a net can estimate values of output water temperature for parameter values not considered or not collected to the training process and, consequently, the efficiency can be calculated. Tab. I.2 (appendix) shows a sample of real efficiency values and their respective estimated values.

For the 84 data (containing T_{in}, T_{amb}, G, T_{out}) used by the net in its training, the real efficiency values have been calculated using Equation (1). Linear regression and the net (trained with those 84 data) have also been considered to calculate efficiency, with the procedures already mentioned. With real and estimated values of efficiency, errors of each one of both techniques have been calculated. Tab. 4 shows a comparison between those error values of ANN and linear regression.

Table 4. Error values in the calculation of solar collector efficiency

Efficiency calculation	Errors of ANN (%)	Errors of LR (%)
Minimum	0.0003	0.0671
Maximum	8.9414	10.5667
Average	2.2710	2.1041
Standard deviation	2.0374	2.0823

Even though the average error obtained in the linear regression technique is lower than the one obtained in ANN technique, the standard deviation value indicates that the error values of ANN are less dispersed than the error values of linear regression. It can be also noticed that a smaller better-defined training set has not prejudiced the generalization capacity of the net, once its errors are satisfactory.

4. Conclusions

A possible use of Artificial Neural Networks has been proposed in this work for the representation of a solar collector, alternatively to other techniques already used with the same purpose. Fig. 8 graphically represents the performance of the trained net in the training process, and Tab. 3 shows its errors in the validation process, with 30 parameters not seen during the training. Efficiency values, calculated using outputs of the trained net, are shown in Tab. 4.

Analyzing Tab. 3 and Tab. 4, it can be noticed that the size reduction of the training set, by means of statistical techniques, has maintained the generalization capacity and low error values of the net, besides its contribution in decreasing the time spent in the training. Table 4 also shows a comparison between ANN and linear regression in the calculation of solar collector efficiency. Linear regression errors are more dispersed than ANN ones, indicating that ANN are an advantageous alternative. Thus, ANN have been considered a satisfactory alternative in the proposed problem.

Since the size reduction of the training set has been profitable and the use of ANN instead linear regression presents advantages, future works in the same area are being planned. One of them includes the use of other techniques in the selection of data for the training set. In other future works, ANN will be used in the modeling of many kinds of solar collectors, each one with its own geometrical parameters and manufacturing characteristics.

5. Acknowledgements

This project is financially supported by CEMIG (Energy Company of Minas Gerais).

6. References

Morrison G.L, and Ranatunga D.B.J, 1980., “Transient response of thermosiphon solar collectors”. Solar Energy 24, pp. 191.

Huang B.J., 1984. “Similarity theory of solar water heater with natural circulation”. Solar Energy 25, pp.105.
Kudish A.I., Santaura P., and Beaufort P., 2000, “Direct measurement and analysis of thermosiphon flow”. Solar Energy, Vol. 35, N° 2, pp. 167-173
Kalogirou S.A, 2000, “Thermosiphon solar domestic water heating systems: long term performance prediction using ANN”, Solar Energy, Vol. 69, N°2, pp. 167-174.
Kalogirou S.A., Panteliou S., Dentsoras A., 1999, “Modeling solar domestic water heating systems using ANN”, Solar Energy, Vol. 68, N° 6, pp. 335-342.
Duffie J.A. and Beckman W.A., 1999, “Solar Engineering of Thermal Processes”, 2a. Ed. John Wyley&Sons, Inc., USA.
Kovács Z.L., 1996, “Redes Neurais Artificiais”, Edição acadêmica São Paulo, Cap 5, pp.75-76. São Paulo. Brasil.
Bittencout F.R. and Zárate L.E., 2002, “Controle da Laminação em Redes Neurais, com Capacidade de Generalização, e Lógica Nebulosa via Fatores de Sensibilidade”, Congresso Brasileiro de Automática, Natal, RN, Brasil.
Ashrae 93-86 Ra 91, 1986. “Methods of Testing to Determine the Thermal Performance of Solar Collectors”, American Society of Heating, Refrigeration, and Air-Conditioning Engineers, Inc., Atlanta.

Appendix

Table I.1. Sample of the trainig set

T_{amb}	T_{in}	G	T_{out}
27.02	38.09	1003.38	45.04
24.99	23.72	767.35	29.95
24.07	23.06	921.81	30.35
26.76	37.92	1012.9	45
25.08	27.14	909.19	33.96
24.88	24.79	942.41	31.74
23.55	30.25	922.83	36.91
23.41	22.68	969.89	30.25
24.51	22.86	939.96	29.92
23.09	32.85	927.03	39.19

Table I.2. Comparison of efficiency values

Efficiency	ANN	Linear regression
0.649205	0.658952	0.660887
0.577049	0.565908	0.589940
0.571363	0.597309	0.581160
0.585377	0.595684	0.581360
0.619978	0.600094	0.649054
0.613144	0.657389	0.653534
0.603335	0.578951	0.607819
0.578527	0.595848	0.581105
0.657918	0.655221	0.656952
0.476594	0.561617	0.467298

Table I.3. Validation data set

T_{amb}	T_{in}	Irradiance	T_{out}	ANN T_{out}	ANN error	ANN error (%)
23.75	20.35	743.61	26.89	27.68	0.51	1.87
23.86	20.78	781.23	27.36	28.29	0.55	1.99
24.43	20.96	811.62	27.65	28.85	0.78	2.78
24.7	21.24	850.34	28.04	29.70	0.43	1.47
24.45	22.62	871.81	29.41	30.75	0.20	0.65
24.73	25.66	881.77	32.07	33.01	0.16	0.47
24.76	26.02	908.71	33.01	37.85	0.33	0.87
25.37	28.53	883.55	35.54	38.92	0.34	0.88
25.66	31.61	968.54	38.69	41.39	0.61	1.50
25.6	34.74	1007.28	41.38	41.48	0.62	1.52
25.92	34.98	1065.36	42.17	42.32	0.72	1.74
26.76	37.92	1012.9	45	42.73	0.85	2.04
26.95	38.1	1044.55	45.11	45.89	1.26	2.82
25.95	38.19	789.12	43.35	46.03	0.91	2.02
23.41	22.68	969.89	30.25	46.31	1.04	2.29
23.71	22.85	949.8	30.41	44.22	0.92	2.12
23.76	22.93	941.18	30.44	45.76	1.11	2.48
24.1	23.01	931.59	30.39	30.51	0.19	0.61
24.31	23.16	903.18	30.31	30.52	0.24	0.81
22.95	24.47	878.18	31.12	30.63	0.21	0.69
23.29	30.17	916.35	36.79	30.43	0.50	1.68
23.11	32.97	933.2	39.4	47.45	1.40	3.03
23.46	43.48	967.66	49.45	47.57	1.48	3.20
23.89	53.33	977.22	58.63	58.02	0.58	0.99
23.81	57.86	953.49	62.13	34.08	0.13	0.37
24.51	22.85	943.37	29.86	34.05	0.10	0.31
24.62	22.95	934.9	30.11	44.92	1.07	2.44
25.1	27.13	911.55	33.95	52.56	1.25	2.44