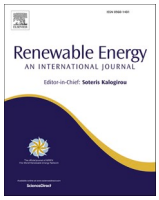
Renewable Energy 216 (2023) 119086

Contents lists available at ScienceDirect

Renewable Energy

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



Random Forest model to predict solar water heating system performance

I. Lillo-Bravo [a,*], J. Vera-Medina [b], C. Fernandez-Peruchena [b], E. Perez-Aparicio [a], J.A. Lopez-Alvarez [a], J.M. Delgado-Sanchez [c]

[a] Department of Energy Engineering, University of Seville, Seville, Spain
[b] Solar Thermal Energy Department, National Renewable Energy Centre. Pamplona (CENER), Spain
[c] Department of Applied Physics I, University of Seville, Seville, Spain

ARTICLE INFO

Keywords:
Solar thermal energy systems
Artificial intelligence
Random Forest
Solar water heating

ABSTRACT

This research proposes a Random Forest RF model to replace the experimental tests required by the ISO 9459–5:2007 for predicting the annual energy supplied and the solar fraction covered by a thermosiphon solar water heating system (TSWHS) for the same locations and daily load volumes that this standard. 38 TSWHS have been tested according to the procedures outlined in the standard ISO 9459-5 and two more have been selected from the Solar Keymark database to get the training and testing data set. From these, data from 36 of the TSWHS were used for RF model training, while data from the remaining four TSWHS were used for its testing. To assess the performance of the RF model, three statistical indicators were calculated: mean absolute percentage error (MAPE), mean absolute error (MAE) and the determination coefficient (R-square). Results show MAPE between 2.94% and 5.86% for the annual energy supplied and the solar fraction and R-Square between 0.995 and 0.998 for the annual energy supplied and between 0.973 and 0.976 for the solar fraction for all locations and daily load volume. Consequently, the RF model could be used successfully to replace the experimental tests required by the Standard.

1. Introduction

More of the renewable energy technologies are cost-effective today in an increasing number of markets. Furthermore, these technologies play a significant role in mitigating the climatic effects of the use of the energy. As a result, requiring the adoption of these technologies is on rise. For example, Directive 2010/31/EU [1] states the implementation of Nearly Zero-Energy Buildings (NZEBs), and has been further enhanced by the Directive (EU) 2018/844 [2], which facilitates the cost-effective transformation of existing buildings into nearly zero-energy buildings. Among the renewable energy technologies, solar thermal system is one of the most widely used in the building sector.

Most of the solar thermal systems installed worldwide are used for the production of hot water in buildings. Worldwide, more than three quarters of all solar thermal systems installed are thermosiphon solar water heating systems (TSWHS), and the rest are forced circulation solar heating systems. In these systems, water is heated directly by solar radiation and circulated naturally through the system by convection. They are simpler and less expensive than pumped solar heating systems, which use pumps to circulate water or any alternative heat transfer fluid through the system. Thermosiphon systems are more common in warm climates, such as Africa, South America, Southern Europe, and the Middle East and North Africa (MENA) countries [3], where the demand for hot water is high and solar radiation is an abundant and free resource. In these regions, TSWHS can be considered a reliable and cost-effective source of hot water without any external energy input.

Predicting the annual energy production by the TSWHS is challenging due to the complexity of those systems, the many factors that influence their performance [4], including the size and orientation of the collector, the thermal properties of the absorber and insulation [5], the efficiency of the heat transfer fluid, the flow rate of the water [6], the ambient temperature, solar irradiance [7], and the daily load volume of hot water required [8]. Furthermore, these factors interact in a non-linear relationship [9,10], making the prediction of the system energy output a complex task [11].

To ensure that TSWHS meet the technical required specifications, they must be experimentally tested according to the test standards, before they are commercially available. By testing TSWHS according to these standards, manufacturers and installers can ensure that their products are reliable, energy-efficient, and meet the expectations of their customers. On the European market, TSWHS must pass the

* Corresponding author. Department of Energy Engineering, Camino de los Descubrimientos, s/n, 41092, Seville, Spain.
E-mail address: isidorolillo@us.es (I. Lillo-Bravo).

https://doi.org/10.1016/j.renene.2023.119086
Received 28 April 2023; Received in revised form 10 July 2023; Accepted 25 July 2023
Available online 25 July 2023

Nomenclature

A	Collector aperture area (m^2)
A_C*	Effective collector loop area coefficient (m^2)
ATE	Athens
C_S	Heat capacity coefficient of the store (MJ/K)
DJ	Heat exchanger double jacket
DJH	Heat exchanger double jacket + Helix
D_L	Mixing constant coefficient
DAV	Davos
$f_{sol,EXP}$	Annual solar fraction covered by the TSWHS obtained experimentally in the laboratory (%)
$f_{sol,TRA}$	Annual solar fraction covered by the TSWHS obtained from the RF Model in the training phase (%)
$f_{sol,TEST}$	Annual solar fraction covered by the TSWHS obtained from the RF Model in the testing phase (%)
GWo	Collector thermal insulation of Glass Wool
H	Heat exchanger Helix
HI	Storage tank above the collectors
LO	Storage tank behind the collectors
MWo	Collector thermal insulation of Mineral Wool
PFo	Collector thermal insulation of Polyisocyanate
PGWo	Collector thermal insulation of Polyurethane and Glass Wool
PI	Tank thermal insulation material of Polyisocyanate
PMWo	Collector thermal insulation of Polyurethane and Mineral Wool
PU	Tank thermal insulation material of Polyurethane
$Q_{L,TEST}$	Supplied annual energy by the TSWHS obtained experimentally in the laboratory (MJ/year)
$Q_{L,TRA}$	Supplied annual energy by the TSWHS obtained from the RF model in the training phase (MJ/year)
$Q_{L,TEST}$	Supplied annual energy by the TSWHS obtained from the RF model in the testing phase (MJ/year)
RF	Random Forest model
RWo	Collector thermal insulation of Rock Wool
S_C	Collector loop stratification parameter
SEL 1.1	Collector selective absorber with Tinox treatment
SEL 1.2	Collector selective absorber Bluetec Eta-Plus treatment
SEL 1.3	Collector selective absorber with Mirotherm treatment
SEL 2	Collector selective absorber with Mirosol treatment
SEL 3	Collector selective absorber with Chromium oxide treatment
SEL 4	Collector selective absorber with polyester powder coating treatment
SEL 5	Collector selective absorber with PVD treatment
STO	Stockholm
u_C*	Heat-loss coefficient of the collector loop parameter ($W/m^2 \cdot K$)
U_S	Total store heat loss coefficient (W/K)
V	Storage volume (l)
WUR	Wurzburg

European Standard Tests, EN 12976:2019 [12]. EN 12976-2:2019 standard establishes the procedures for the estimation of the annual energy production (Q_L) and solar fraction (f_{sol}) for four European locations (Athens, Davos, Stockholm and Wurzburg) for daily load volumes between 0.5 and 1.5 times the tank volume of the TSWHS, based on experimental tests, jointly with the "*In-situ*" software [13]. The European Standard efficiency test refers to two International Standards ISO 9459-2:1995 [14] and ISO 9459–5:2007 [15]. Those tests required a qualified laboratory and can be time-consuming and expensive.

Artificial Neural Networks (ANNs) [16] are a promising alternative to those experimental tests when it comes to predicting the performance of TSWHS to those experimental tests. ANNs are a type of machine learning algorithm that can learn the non-linear relationships between the input variables (like its constructive parameters, climatic conditions, and daily load volume of hot water) and the output variables (such as energy output and performance). Previous studies [17–19] reported potential advantages of ANNs in modeling these energy systems, such as high accuracy, generalization capabilities, and short computing time, over other theoretical and experimental modeling techniques. Recent research on surrogate models based on various machine learning techniques has demonstrated their potential in addressing numerous technical issues, making them a valuable predictive tool. [20–22]. Testing these models further enhances their reliability and usefulness.

Therefore, ANNs has been applied to complex non-linear engineering problems in different real-world applications. ANNs might represent a cost-effective alternative to experimental tests as they do not require expensive infrastructures. However, they must be trained and tested against experimental data to ensure their accuracy and reliability.

Specifically, in the research field of solar energy systems, there are some previous studies reporting the advantages of the ANN models for predicting and optimizing the performance. Ammar et al., [23] reviewed the applications of ANN as an intelligent system-based method for predicting the performance of different solar energy devices' performance like such as solar collectors, solar assisted heat pumps, solar air and water heaters, photovoltaic/thermal systems, solar stills, solar cookers, and solar dryers. Farkas [24], Kalogirou, [25] and Hamdan [26] used an ANN model to simulate the thermal behavior of flat plate collectors for its control operation. All of them found that ANN models could accurately predict the performance of these devices and outperform traditional modelling techniques in terms of accuracy and computational efficiency.

Kalogirou et al. [27] developed a model for solar domestic water heating systems using artificial neural networks. The model estimated the useful energy extracted from a solar water heating system and the temperature rise in the stored water under the stated physical parameters of the system and the weather conditions. The accuracy of the predictions was within 7–9%. Moreover, Kalogirou et al., [28] uses an ANN to predict the long-term performance of TSWHS. This research estimates the thermal energy output of the TSWHS for a draw-off quantity equal to the tank volume and the thermal energy output of the system and the average amount of hot water per month at demand temperatures between 35 and 40 °C. In this research, the ANN input requires the results of the experimental tests of the Standard ISO 9459-2:1995 and the proposed ANN is an alternative method to the simulation software supplied with the standard but cannot replace the experimental tests.

Dikmen et al. [29] compare ANN, adaptive neuro-fuzzy inference system, and genetic algorithm to predict and maximize the thermal performance of an TSWHS with an evacuated tube solar collector. The inputs of the models were solar radiation, ambient temperature, mean storage tank temperature, and tilt angle. The results showed that ANN is recommended to model that type of solar collector.

Muhammad et al. [30] present a comprehensive study to compare three machine learning models, Random Forest (RF), Decision Trees and Support Vector Regression, to predict the useful hourly energy from a solar thermal system. They found that amongst the algorithms studied, RF has the lowest mean square error. One of the advantages of the RF is its ability to use different types of inputs and evaluate their relevance, making it a versatile and flexible tool for analyzing complex data sets. In addition, as an ensemble learning method, RF is less likely to overfit than single decision trees, which can lead to more accurate predictions and better generalization to new data sets.

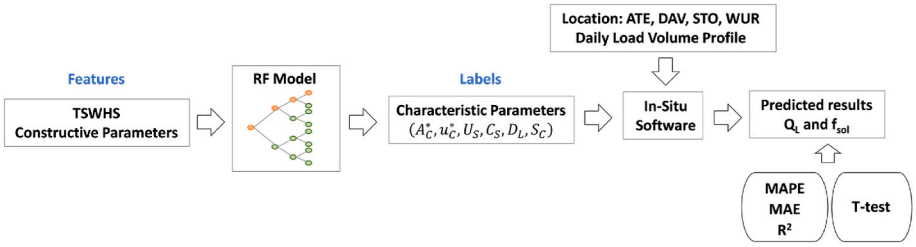


Fig. 1. Procedure to estimate Q_L and f_{sol} with the RF model and the In-Situ software.

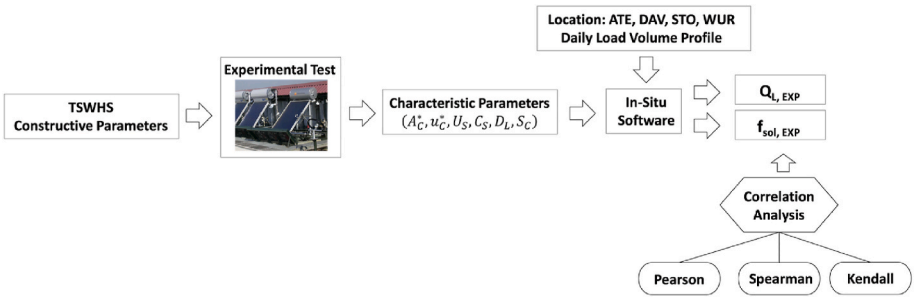


Fig. 2. Procedure to estimate QL and fsol using experimental tests and the In-Situ software.

The RF model, devised by Breiman [31], is a flexible machine learning algorithm consisting of a large number of individual decision trees generated by subsets of the training data. The optimal split for each node is identified from a set of randomly chosen candidate variables in the tree building procedure. Random forests are widely used in a variety of applications, such as solar and wind forecasting [32,33], chemo informatics [34], ecology [35], genomic data analysis [36] and solar resource [37] among others. Random Forest algorithm offers mainly accuracy and robustness with respect to noise and outliers in the input data, which makes it suitable for predicting the behavior of new equipment. Also, its combination of multiple trees improves its ability to generalize to new data reducing overfitting. These advantages outweigh the disadvantages of Random Forest, as the complexity of its interpretation or training and prediction time due to its use of computational resources and memory.

Hao Li et al. [38] proposed a machine learning-based high--throughput screening method to show the potential application to the design and optimization of solar energy systems with tube solar collectors where the selection of the inputs data, a throughput screening process and validation steps are proposed. Therefore, with an optimized RF model the best set of constructive characteristics could be selected to allow the prediction of the annual energy supplied (Q_L) and the solar fraction (f_{sol}) covered by the solar thermal. Finally, with a proper experimental database to train and test the model, a predictive model can be acquired.

This research proposes a RF model, as an alternative to conducting the experimental tests required by the ISO 9459–5:2007, to predict, jointly with the In-Situ software of the Standard, the annual energy supplied (Q_L) and the solar fraction (f_{sol}) covered by the TSWHS. The predictions are made for four different locations and daily load volumes that are specified in the Standard. How the Standard requires experimental test and the In-Situ software to estimate the performance of the TSWHS, the novelty of this paper is to replace the experimental tests with a RF model using the constructive parameters of the thermal solar system as input variables and using the In-Situ software.

In addition, this RF model would facilitate the work of design engineers to customize the TSWHS design to meet the specific hot water demand according to each climatic condition and the daily load demand. This approach can potentially save time and resources compared to traditional experimental testing methods. Fig. 1 shows the procedure to estimate Q_L and f_{sol} with the RF model and the In-Situ software.

The research is presented as follows: Section 2 describes the RF model and the characteristics of the thermosyphon solar systems. Section 3 shows the experimental results and the methodology followed for training and testing the full procedure described in Fig. 2 (RF model and In-Situ software). Conclusions are then presented in Section 4.

2. Material and methods

2.1. RF model description

Random Forest is a flexible machine learning algorithm that consists of a large number of individual decision trees generated by subsets of the training data to make predictions. Each tree in the forest is generated using a subset of the training data, and the optimal split for each node is chosen from a randomly chosen subset of candidate variables. This helps to reduce the overfitting and improve the generalization performance of the model. In this particular case, the RF model was developed using the R programming language (version 4.0.2) which supports regression, and it is implemented as in the original RF algorithm of Breiman [31], but in addition includes extra implementations of extremely randomized trees and quantile regression forests, being also faster than the original algorithm, and the "Ranger" Package (random forest generator, version 0.12.1) [39]. The package was used to build the forests with the estimated response variances as a splitting rule, and a total of 500 trees were generated in each forest. The tree type used is the estimated response variances. RF overfitting may be caused by different reasons. As mentioned above, if we fit the training data too closely (trees are too deep), we lose the generalization. Consequently, to avoid overfitting we have tuned the hyper-parameters of the algorithm, as number of trees, number of features to be considered for each tree, the split rule (i.e., the rule by which each split is considered in a tree), or the maximum depth of the tree, among others.

To obtain the training data 36 TSWHS were experimentally tested in our experimental facilities [15], according to the Standard ISO 9459-5:2007. Considering the TSHWS characteristic parameters, the climatic conditions in four different locations (Athens, Davos, Stockholm and Wurzburg), and the daily load volume as inputs in the In-Situ software, the annual energy supplied ($Q_{L,EXP}$) and the solar fraction ($f_{sol,EXP}$) as final results have been calculated.

Table 1
Features selected from constructive parameters.

Equipment	Features
Collector	Collector absorber type
	Material of the collector thermal insulation
	Thickness of the collector thermal insulation
	Aperture Area
Storage Tank	Volume of the tank thermal insulation
	Material of the tank thermal insulation
	Thickness of the tank thermal insulation
Heat Exchanger	Type
	Exchange surface
Collector Loop Tube	Length of the thermal insulation
	Thickness of the thermal insulation
Others	Profile of the tank respect to the collector

Table 2
Feature selected for each characteristic parameter.

Features	Characteristic Parameter
Collector absorber type	A_C^* (m^2)
Collector thermal insulation material	u_C^* $(W \cdot m^{-2} \cdot K^{-1})$
Collector thermal insulation thickness	
Profile type	
Aperture area	
Heat exchanger surface	
Collector loop tube length	
Collector insulation thickness.	
Tank volume	U_S (WK^{-1})
Tank thermal insulation material	C_S $(MJ \cdot K^{-1})$
Tank thermal insulation thickness	
Collector absorber type	D_L $(-)$
Profile type	S_C $(-)$
Heat exchanger type	
Aperture area	
Tank volume	
Volume/Area ratio	
Heat exchanger surface.	

In order to fit the RF model, the first step was to define the features from the list of constructive parameters. For the sake of clarity, features are also known as input variables that are used to predict the outcome or the dependent variable (Q_L and f_{sol}). Table 1 shows the constructive parameters selected using as input to the RF model. Once the independent variables were defined, the next step was to collect and prepare the data for training and testing the RF model jointly with the In-Situ software. This involved collecting data from the constructive parameters (Table 1) for a set of samples.

With the propose of improving the selection of the features of the RF model, the construction parameters that have a significant influence on each of the characteristic parameters have been previously selected. Table 2 shows the features selected from Table 1 for each TSWHS characteristic parameter defined in the Standard ISO 9459-5:2007: A_C^*, u_C^*, U_S, C_S, D_L and S_C. These characteristics parameters are assigned as labels of the RF model, and they are one of the inputs in the In-Situ software.

Table 1 shows the pool of available parameters for each TSWHS. Table 2 shows the features selected from Table 1 used as input in the RF model, distinguishing them according to each of the characteristic parameter to be determined. The selection criterion has been to include all the available constructive parameters that could have some kind of influence on the characteristic parameter, according to the definition of the parameter itself, and without previously discarding any of them.

To train the RF model jointly with the In-Situ software features from 36 different TSWHS systems are used as input. Each datasheet of construction parameters are used to predict the annual energy ($Q_{L,TRA}$) and the solar fraction ($f_{sol,TRA}$) at four different locations (Athens, Davos, Stockholm, and Wurzburg) with adverse climatic conditions. Additionally, for each previous state the solar thermal system modified the daily load volume in the range of 0.5–1.5 times its nominal volume (12800 data). Finally, the predicted values ($Q_{L,TRA}$ and $f_{sol,TRA}$) will be compared to results obtained experimentally in the laboratory ($Q_{L,EXP}$ and $f_{sol,EXP}$) to evaluate their correlation.

Secondly, a new phase was carried out to test the RF model with the In-Situ software. In this second stage, 4 commercial TSWHS were used (constructive parameters detailed in Table 4). Two of them (ID 37 and 38) have been experimentally tested according to the Standard ISO 9459-5:2007 in our testing laboratory and the required information for the other ones (ID 39 and 40) was obtained from the manufacturer datasheet in the Solar Keymark database [40], with license number 011-7S2865A, model GreenoneTec TSC 160 [41] and 011-7S2817A, model Zelios Thermo CF-GR 150/1 TR/TT/TT DT [42] respectively. In order to test the process, the prediction of the annual energy ($Q_{L,TEST}$) supplied by the solar system and the solar fraction ($f_{sol,TEST}$) covered by the solar system is compared to $Q_{L,EXP}$ and $f_{sol,EXP}$ values respectively, to analyze their correlation.

2.2. TSWHS characteristics

As explained, 36 TSWHS from different manufacturers have been tested according to the Standard ISO 9459-5:2007 in the accredited solar system testing laboratory of the School of Engineering of the University of Seville (Spain), all of them closed, with flat plate collectors, horizontal tanks, and without auxiliary heating. Details of each TSWHS used during training phase can be found in Table 3; TSWHS specifications used during RF testing phase can be found in Table 4. This laboratory is accredited according to the European Standard EN 12976:2019 and the Standard ISO 9459-5:2007. The data acquisition and test equipment used in our test facility (pyranometer, ambient temperature sensor, anemometer, flow-meter, inlet and outlet temperature sensors) are calibrated in an accredited laboratory, according to the accuracy and precision requirements shown in Standard ISO 9459-5:2007. All tests are performed with the TSWHS installed according to the manufacturer's installation instructions. The sequence is shown in Fig. 2.

The experimental testing procedure for TSWHS described in Standard ISO 9459-5:2007 consists of two sequences (S-Sol and S-Store), in which certain parameters such as solar irradiation, inlet and outlet water temperature, ambient temperature, and flow rate are samples. Based on these parameters, the annual energy supplied by the solar system ($Q_{L,EXP}$) is calculated using characteristic parameters and the In-situ software. The calculations take into consideration the referenced locations and the daily load volumes. The draw-off flow rate is fixed in the Standard at 10 l/min. A more detailed description of these tests and the constructive parameters can be found in the reference [43]. Training TSWHS samples correspond to 90% of the data (36 systems), and remaining 10% (4 systems) of the data samples were used for testing the total process. The TSWHS used to collect experimental data are the full portfolio of equipments available in our experimental facilities, which corresponds to a wide range of commercial equipments available in the current market.

So, considering 40 equipments with 10 features each of them and the performance analyzed for 4 locations using 8 daily load volume, it means we are managing 12800 data.

3. Results and Discussion

This section details the prediction results for the annual energy supplied and the solar fraction covered by the solar system obtained using the RF model jointly with the In-Situ software. First, Fig. 3 shows the dependence of each constructive parameter on the annual energy ($Q_{L,EXP}$) of the TSWHS, respectively, estimated according the Fig. 1 from measurements in the laboratory. It is observed that there is no correlation between any input parameter and the output. This means that there is no linear relationship between the variables analyzed, so any change

Table 3
Constructive parameters of the tested TSWHS for the process training.

TSWHS ID	Absorber type	Collector thermal insulation		Profile	A (m^2)	V(l)	Tank thermal insulation		Heat exchanger		Collector loop tube	
		material	thickness (mm)				material	thickness (mm)	type	surface (m^2)	length (m)	thickness (mm)
1	SEL 3	MWo	40	High	4.14	300	PU	50	DJ	1.8	1.6	25
2	SEL 1.3	GWo	40	High	1.91	200	PU	50	DJ	0.97	3	9
3	SEL 1.3	GWo	40	High	3.82	300	PU	50	DJ	1.78	3	9
4	SEL 1.3	GWo	40	High	2.4	200	PU	50	DJ	0.97	3	9
5	SEL 1.3	GWo	40	High	1.9	149	PU	50	DJ	0.8	2.8	9
6	SEL 1.3	PMWo	38	High	2.58	209	PU	38	DJ	1.2	3.1	25
7	SEL 1.3	PMWo	38	Low	2.58	209	PU	38	DJ	1.2	3.5	25
8	SEL 1.2	GWo	20	High	3.54	301	PU	40	DJ	1.57	3.5	20
9	SEL 1.1	MWo	40	High	1.92	154	PU	43	DJ	1.16	3.05	20
10	SEL 4	RWo	55	High	1.95	192	PU	40	DJ	1.2	2.4	0
11	SEL 1.1	RWo	40	High	3.24	318	PU	50	DJ	1.8	3.75	35
12	SEL 2	PFo	20	High	2.08	158	PU	55	DJ	1	3.25	30
13	SEL 1.3	RWo	50	High	1.94	155	PU	50	DJ	0.73	1.2	9
14	SEL 1.3	RWo	50	High	3.88	314	PU	50	DJ	2.03	1.6	9
15	SEL 1.2	GWo	40	High	1.88	157	PI.	25	DJ	0.81	3.1	15
16	SEL 1.2	GWo	40	Low	3.76	317	PI.	25	DJ	1.78	3.1	15
17	SEL 1.3	GWo	60	Low	1.99	145	PU	50	DJH	0.48	2.1	20
18	SEL 1.3	GWo	25	Low	2.01	157	PU	50	DJ	0.78	2.6	9
19	SEL 1.3	RWo	40	High	2.4	210	PI	30	DJ	0.97	3.25	9
20	SEL 1.3	GWo	60	Low	3.98	282	PU	40	DJ	1.51	2.7	20
21	SEL 1.3	RWo	30	High	4.2	241	PU	50	DJ	1.32	1.1	9
22	SEL 1.3	MWo	30	High	4.46	282	PU	50	DJ	1.51	3	20
23	SEL 1.3	RWo	40	High	4.8	315	PI	30	DJ	1.78	3.9	9
24	SEL 1.1	GWo	50	High	1.94	155	PU	50	DJ	1.6	2.75	19
25	SEL 1.2	GWo	40	Low	3.76	300	PU	38	DJ	1.88	3	15
26	SEL 1.3	RWo	40	High	1.92	156	PU	40	DJ	0.74	2.9	9
27	SEL 1.3	MWo	35	High	4.72	314	PU	38	DJ	2	1.8	13
28	SEL 1.2	GWo	40	High	1.88	152	PU	25	DJ	0.81	3.5	13
29	SEL 1.2	MWo	40	High	2.65	170	PU	60	DJ	1.93	3.3	20
30	SEL 1.3	GWo	20	High	3.84	317	PU	50	DJ	1.66	3.35	9
31	SEL 1.2	MWo	30	High	4.02	282	PU	38	DJ	1.49	3	19
32	SEL 1.3	MWo	25	High	2.01	159	PU	50	DJ	0.87	2.5	14
33	SEL 1.1	GWo	50	High	2.01	211	PU	50	DJ	1.28	2.9	15
34	SEL 1.1	GWo	20	Low	2.44	164	PU	50	DJ	0.91	2.5	15
35	SEL 5	PGWo	40	Low	2.38	170	PU	50	H	2.42	2.4	13
36	SEL 1.1	RWo	15	Low	4.9	299	PU	50	DJ	1.78	2.8	15

Table 4
Constructive parameters of the TSWHS used for the process training.

TSWHS ID	Absorber type	Collector thermal insulation		Profile	A (m^2)	V(l)	Tank thermal insulation		Heat exchanger		Collector loop tube	
		material	thickness (mm)				material	thickness (mm)	type	surface (m^2)	length (m)	thickness (mm)
37	SEL 1.1	GWo	20	High	2.44	164	PU	50	DJ	0.91	2.8	15
38	SEL 1.3	MWo	20	High	3.84	294	PU	65	DJ	1.67	3.34	15
39	SEL 1.3	MWo	30	High	1.92	145	PU	50	DJ	1.3	3	15
40	SEL 5	MWo	25	High	2,01	14	PU	50	DJ	0.82	3	15

in the value of one parameter does not systematically increase or decrease the output parameter. They do not exhibit any clear pattern, and the learnt simple model may not be able to predict accurately. As there is a relationship between Q_L and f_{sol}., defined as $f_{sol} = Q_L/Q_d$, where Q_d is the annual heat demand, there is also no clear pattern between the constructive parameters and f_{sol}.

Fig. 3 show that there are no trivial correlations between the constructive parameters of the TSWHS with $Q_{L,EXP}$. The Pearson correlation coefficient is calculated from the experimental annual energy supplied and the solar fraction, and each constructive parameter is defined as input in the RF model. This coefficient is a measure of the linear relationship between two variables, and if the correlation coefficient is close to zero, it suggests that there is non-linear relationship between the variables. To ensure that there is a non-linear relationship between $Q_{L,EXP}$ or $f_{sol,EXP}$, and the constructive parameters, the Spearman correlation coefficient and the Kendall rank correlation coefficient have been calculated, as these coefficients are defined to quantify the strength of the monotonic relationship between variables which do not necessarily follow a straight line (details of the values obtained are presented in Tables SI1 to SI3 in the Supporting Information Section). As Pearson coefficient, Spearman and Kendall coefficients have a range from −1 to +1, where +/-1 indicates a perfect correlation, and values close to 0 indicate no correlation.

It is noted that the constructive parameters have a different importance score for each of the input features. All constructive parameters have a similar importance in all four locations for each constructive parameter, except for Athens, where the influence of the material collector thermal insulation and the collector tube length in the annual energy production do not follow the same trend as in the other locations. As expected, the collector area and the tank volume are the most significant constructive parameters on the annual energy production and solar fraction, followed by the interchange surface of the heat exchanger. The thickness of the collector tube and the thermal insulation of the material collector are positively related to both $Q_{L,EXP}$ and $f_{sol,EXP}$. However, the length and thickness of the collector tube and collector thermal insulation are negatively related to both $Q_{L,EXP}$ and $f_{sol,EXP}$.

It has been demonstrated that the input and output variables do not follow one strict linear relationship. In this case, Radom Forest is a

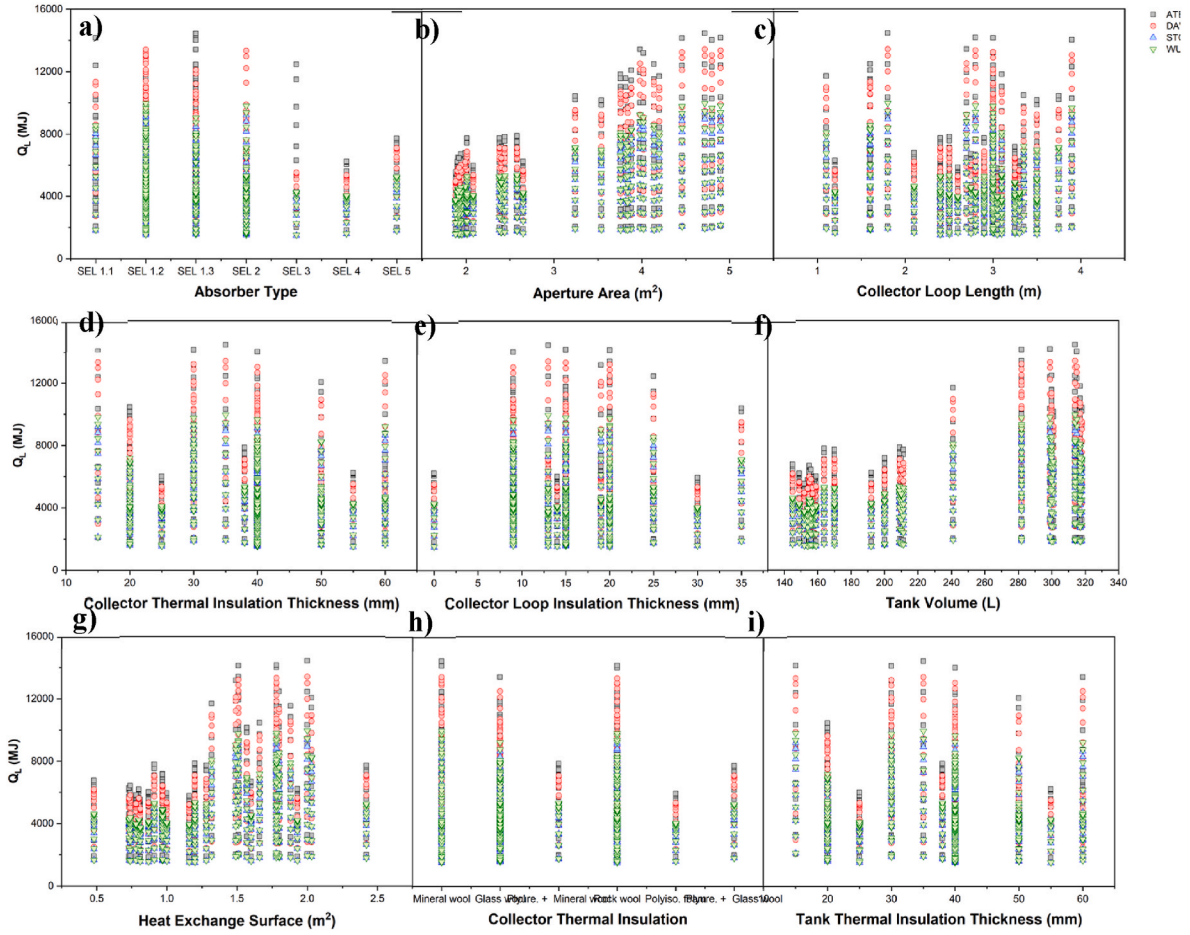


Fig. 3. Correlation between the constructive parameters of the TSWHS to $Q_{L,EXP}$ for four locations, obtained experimentally in the laboratory: a) absorber type; b) aperture area; c) collector loop length; d) collector thermal insulation thickness; e) collector loop insulation thickness; f) tank volume; g) heat exchange surface; h) collector thermal insulation; i) tank thermal insulation thickness.

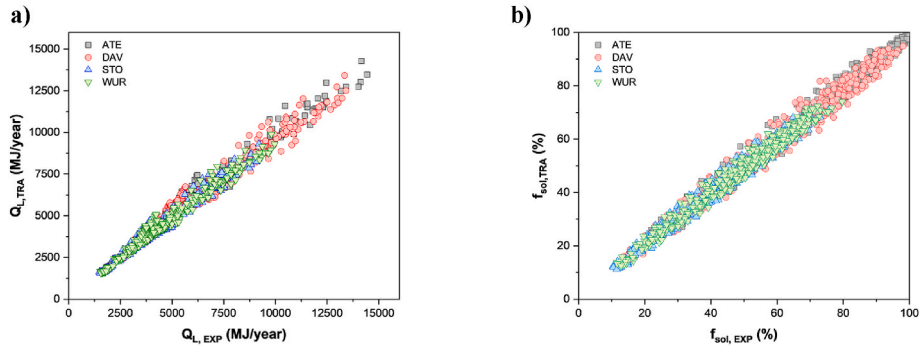


Fig. 4. (a) Correlation between the annual energy supplied ($Q_{L,EXP}$) experimentally in the laboratory and the predicted values in the training phase ($Q_{L,TRA}$) of the Random Forest model at four locations: Athens, Davos, Stockholm, and Wurzburg. (b) Correlation between experimental solar fraction covered by the solar system ($f_{sol,EXP}$) and predicted values in the training ($f_{sol,TRA}$) phase of the Random Forest model, for the same locations as in (a).

recommended option to predict the thermal response of the TSWHS system, because it can capture complex interactions between the requirement of any assumption about the functional form of the relationship, in contrast to linear regression models. The first stage to develop the Random Forest consists in the training.

Fig. 4 show the correlation between the annual energy ($Q_{L,EXP}$) supplied by the solar system and the solar fraction ($f_{sol,EXP}$) covered by the solar system, experimentally obtained in the laboratory, and the values of $Q_{L,TRA}$ and $f_{sol,TRA}$ from the prediction of the trained RF model for the four locations defined in the Standard, Athens, Davos, Stockholm and Wurzburg for different daily volume, jointly with In-Situ software in both cases, as shown in Figures 1 and 2.

Those results show a good correlation between the experimental values and the prediction values of $Q_{L,EXP}$ versus $Q_{L,TRA}$ and $f_{sol,EXP}$ versus $f_{sol,TRA}$ for all climatic conditions and daily load volume. So, according to these figures, the RF model with the In-Situ software predict well the annual energy and the solar fraction for all locations. To assess the performance of the RF model on training phase, three common Key Performance Indicators (KPI) are calculated: mean absolute percentage deviation (MAPE), mean absolute error (MAE) and the determination

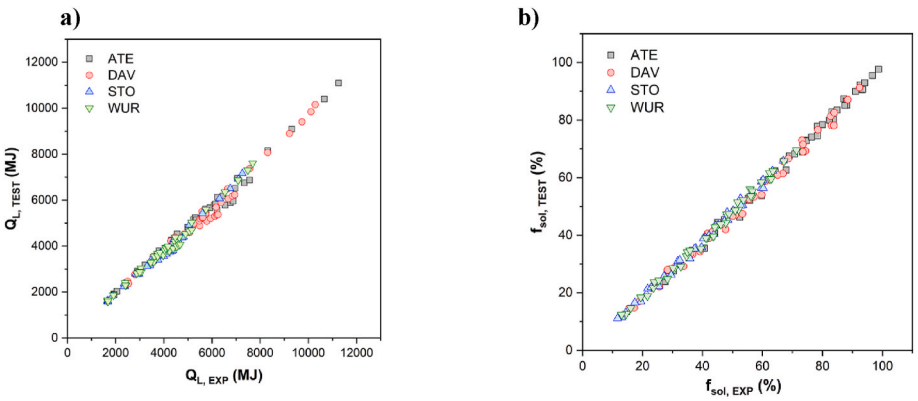


Fig. 5. (a) Annual energy supplied (Q_L), and (b) the solar fraction (f_{sol}) covered by the solar system predicted with RF model and experimentally in the laboratory obtained during the testing phase for four different locations: Athens, Davos, Stockholm, and Wurzburg.

Table 5
Statistical Key Parameter Indicators obtained during the training of the RF model.

	Q_L			f_{sol}		
Location	MAPE (%)	MAE (MJ/year)	R^2	MAPE (%)	MAE (%)	R^2
Athens	3.41	159.6	0.987	3.41	2.4	0.899
Davos	5.72	400.3	0.968	5.72	3.4	0.905
Stockholm	5.54	263.7	0.974	5.54	2.4	0.903
Wurzburg	5.00	250.4	0.970	5.00	2.3	0.904

coefficient (R^2):

a) Mean Absolute Percentage Error (MAPE) represents a measure of prediction accuracy of a forecasting method in statistics, expressed as:

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad \text{(Eq 1)}$$

where y_i is the actual value, $\hat{y}_i$ is the forecast value, and N is the total number of measurements.

b) Mean Absolute Error (MAE) is another measure of the average distance between the predicted and the experimental values. It is a good indicator for models where outliers are expected.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad \text{(Eq 1)}$$

c) R-square (R^2) represents the proportion of the variance in the independent variable that is explained by the independent variables in the model. An R^2 score of 1 indicates a perfect fit between predicted and experimental results, while a score of 0 indicates no relationship between the independent and dependent variable.

$$R^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad \text{(Eq 3)}$$

Table 3 summarizes these KPIs calculated for all climatic conditions and different daily load volume within the Standard range. A high correlation is observed between experimental and predicted datasets during the training phase. R-square of $Q_{L,TRA}$ vs $Q_{L,EXP}$ varies from 0.968 to 0.987 and R-square of $f_{sol,TRA}$ vs $f_{sol,EXP}$ varies from 0.899 to 0.905 (close to 1 for both parameters). This result is aligned with Fig. 5. It is likely that the RF model prediction fit well with the experimental data obtained in the laboratory. Data summarized in Table 5 also point out that MAPE and MAE parameters are different for each location. This result might suggest that the training RF model may be more accurate for some locations than others; for example, Athens (high solar irradiance and high ambient temperature over the year) shows the best operational conditions to run this model meanwhile Davos (high solar irradiance but low ambient temperature) is the worst scenario.

As shown in Table 5, MAPE values are the same for Q_L and f_{sol} as expected. It is consistent due the definition of MAPE (Eq. 13) and the solar fraction. $f_{sol,}$ which relates Q_L and Q_d, and Q_d has the same value for the training and experimental testing procedure.

Once the training phase was completed and the RF model has been trained, a new batch of samples, that is separate from the data used for the training, was used to make predictions and test the proposed model. Two requirements were sought for the new data used: first, they should have the same characteristics variable as the training dataset, while being entirely separate from it; and second, the testing dataset must be representative of the data that the model is expected to solve in the real world. It is important to point out that the dataset used for this second phase (testing) is only used for evaluation purposes, but any hyperparameter of the RF model, which defines the number of trees used in the model, has been tuned to avoid overfitting the model to the testing data. 4 commercially available TSWHS in the market were selected for this second phase (constructive parameters described in Tables 3 and 4).

The annual energy supplied (Q_L) and the solar fraction (f_{sol}) testing data are plotted in Fig. 5, respectively. The correlation observed on these data collection fits well and, they have the same tendency of the previous collection data used during the training phase of the RF model. Note the new data overlapped with previous results obtained during the training phase. As it was previously explained, a good fit of the testing data is not a necessary and sufficient condition to conclude that the RF model will perform well on any new data in the future. To ensure the model reliability of the model, additional KPI (MAPE and MAE) to R-square were evaluated, following the same analysis procedure than during the training phase.

By comparing these KPIs for both the testing and training results, we can get a good idea of how well the RF model is performing on both the training and testing data. It is observed high correlation between experimental and simulated datasets. R-square of Q_L varies from 0.968 to 0.987 and R-square of f_{sol} varies from 0.977 to 0.981 (close to 1 for both parameters). It is likely that a prediction model based on this correlation fits well. However, it is important to assess the quality of the fit and the performance of the trained model on new and unseen data. To characterize the quality of the training model R-square parameter is not sufficient as correlation does not imply necessary causation. In order to achieve a better capability of the training model, a cross sectional analysis is performed for each location tested, using MAPE and MAE

Table 6
Statistical Key Parameter Indicators obtained during the testing of the RF model.

Location	Q_L			f_{sol}		
	MAPE (%)	MAE (MJ/year)	R^2	MAPE (%)	MAE (%)	R^2
Athens	2.94	157.6	0.998	2.94	2.2	0.975
Davos	5.86	325.9	0.995	5.86	3.4	0.974
Stockholm	4.96	192.9	0.995	4.96	2.1	0.976
Wurzburg	4.51	182.1	0.995	4.51	2.1	0.973

metrics.

MAPE and MAE results are presented in Table 6 for RF model testing phase. Both KPIs are lower than those obtained with the previous dataset in the training phase (Table 5), from which can be interpreted that overfitting to the training sets has been avoided. Overfitting occurs when the RF model learns the noise and idiosyncrasies of the training data rather than the underlying patters that generalize the new data. As conclusion, the developed RF model is able to memorize the training dataset and, more importantly, it is able to generalize to new data.

Once the RF model has been tested (i.e. it has been shown to be able to memorize the training set, and to predict new data), the simulated and experimental datasheets are compared to investigate if the experimental test might be replaced by the simulation. For this purpose, a paired *t*-test was carried out to determine the significance of differences in the mean and deviation $Q_{L,EXP}$ and $Q_{L,TEST}$ for each location and daily load volume tested. This test measures one parameter, such as Q_L, that undergoes four different locations and eight daily loads volume (0.5–1.5 times the nominal volume of TSWHS). For the *t*-test, there are two possible results depending on the p-value: if the p-value is less than the reference probability (0.05), the result is statistically significant and there is no null hypothesis. However, if the p-value is greater than the reference probability, the result is not significant. In this study, the *t*-test has been performed to analyze the mean and standard deviation. The test statistic for the paired-difference *t*-test is calculated as:

$$t = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \quad \text{Eq 4}$$

where $\overline{X}$ is the average difference, σ is the standard deviation of the difference, and n is the sample size. In paired tests, the null hypothesis is assumed when $\mu_0 = 0$, meaning that there is no difference between groups [41]. The significant probabilities obtained from the *t*-test are shown in Tables 7 and 8, including Standard Error of Mean (SEM) and Standard Deviation (SD).

The p-value for all cases exceeded 0.05. As a result of significant tests, there was no significant difference in Q_L and f_{sol} predicted from the RF model and equivalent data experimentally obtained in the laboratory following the ISO 9549-5:2007 Standard. Therefore, we concluded that the annual energy supplied ($Q_{L,EXP}$) and the solar fraction ($f_{sol,EXP}$) can be evaluated directly from the RF model without the need of experimental tests, which require complex facilities and are highly expensive.

4. Conclusions

The paper details the feasibility of using a Random Forest model to predict the annual energy supplied and solar fraction covered by a TSWHS in comparison with the resulting values using the experimental procedure described in ISO 9549-5:2007. Different statistical measures were used to assess the use of RF models with the In-Situ software. Results show MAPE between 2.94% and 5.86% for the annual energy supplied and the solar fraction, MAE between 157.6 and 325.9 MJ/year for the annual energy supplied and between 2.1% and 3.4% for the solar fraction and R-Square between 0.995 and 0.998 for the annual energy supplied and between 0.973 and 0.976 for the solar fraction for all locations and daily load volume. Consequently, the RF model could be used successfully to replace the experimental tests required by the Standard, avoiding cost, experimental resources and time required by the experiments. All these advantages have a positive impact in the commercialization of this renewable energy technology.

In addition, the use of the RF model would facilitate the work of TSWHS designers to predict the relative influence of each of the constructive parameters on the energy performance of the solar system.

The performance of RF model with the In-Situ software will be enhanced in the future with more training samples. Finally, RF models will need to be developed for other types of factory made solar thermal systems, such as forced-circulation systems and integrated collector storage systems, or with other types of collectors, such as evacuated tube solar collector or unglazed solar collector.

CRediT authorship contribution statement

I. Lillo-Bravo: Conceptualization, Methodology, Writing – original draft, Investigation. **J. Vera-Medina:** Resources, Data curation, Software. **C. Fernandez-Peruchena:** Formal analysis, Data curation. **E. Perez-Aparicio:** Resources, Formal analysis. **J.A. Lopez-Alvarez:** Resources, Data curation. **J.M. Delgado-Sanchez:** Data curation, Methodology, Investigation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial

Table 7
Results of a statistical *t*-test (Q_L).

Location	Mean		SD		SEM		t Statistic		Prob > | t |	
	$Q_{L,EXP}$ (MJ/year)	$Q_{L,TEST}$ (MJ/year)	$Q_{L,EXP}$ (MJ/year)	$Q_{L,TEST}$ (MJ/year)	$Q_{L,EXP}$ (MJ/year)	$Q_{L,TEST}$ (MJ/year)	$Q_{L,EXP}$	$Q_{L,TEST}$	$Q_{L,EXP}$	$Q_{L,TEST}$
ATE	5501	5343	2074	2036	476	467	0.2363	0.2363	0.81452	0.81452
DAV	5999	5674	2012	2016	461	463	0.4888	0.4888	0.62096	0.62096
STO	4084	3892	1419	1399	326	321	0.4221	0.4221	0.67547	0.67547
WUR	4257	4075	1491	1476	342	339	0.3781	0.3781	0.70755	0.70755

Table 8
Results of a statistical *t*-test (f_{sol}).

Location	Mean		SD		SEM		t Statistic		Prob > | t |	
	$f_{sol,EXP}$ (%)	$f_{sol,TEST}$ (%)	$f_{sol,EXP}$ (%)	$f_{sol,TEST}$ (%)	$f_{sol,EXP}$ (%)	$f_{sol,TEST}$ (%)	$f_{sol,EXP}$	$f_{sol,TEST}$	$f_{sol,EXP}$	$f_{sol,TEST}$
ATE	77.281	75.053	7.752	8.085	1.17	1.85	0.8672	0.8672	0.39157	0.39157
DAV	59.142	55.738	10.001	9.994	2.29	2.30	1.0489	1.0489	0.30121	0.30121
STO	43.435	41.316	6.726	6.747	1.54	1.55	0.9694	0.9694	0.33879	0.33879
WUR	47.057	44.971	6.650	6.753	1.52	1.54	0.9589	0.9589	0.34398	0.34398

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to express our sincere appreciation to Dra. Ana Diaz Real for the valuable discussions provided throughout the development of the Random Forest model. Your contributions have greatly enriched our work, and we are grateful for the opportunity to engage in meaningful exchanges of ideas.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.renene.2023.119086.

References

[1] Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the Energy Performance of Buildings. Off. J. Eur. Union (2010), vol. 3, pp. 124-146.
[2] Directive (EU), 2018/844 of the European Parliament and of the Council of 30 May 2018 Amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on Energy Efficiency, Off. J. Eur. Union, 2018, pp. 75–91, 4.
[3] Werner Weiss, Monika Spörk-Dür, Solar Heat Worldwide. Global Market Development and Trends 2021 Detailed Market Figures 2020, IEA Solar Heating & Cooling Programme, 2022, https://doi.org/10.18777/ieashc-shw-2022-0001.
[4] N. Nikolić, N. Lukić, Theoretical and experimental investigation of the thermal performance of a double exposure flat-plate solar collector, Sol. Energy 119 (2015) 100–113.
[5] K. Balaji, P. Ganesh Kumar, D. Satkhivadivel, V.S. Vigneswaran, S. Iniyan, Experimental investigation on flat plate solar collector using frictionally engaged thermal performance enhancer in the absorber tube, Renew. Energy 142 (2019) 62–72.
[6] E.A. Kabeel, A. Khalil, S.M. Shalaby, M.E. Zayed, Investigation of the thermal performances of flat, finned, and v-corrugated plate solar air heaters, J. Sol. Energy Eng. 138 (5) (2016), 0510041-051004-7.
[7] P.K.C. Pillai, R.C. Agarwal, Factors influencing solar energy collector efficiency, Appl. Energy 8 (3) (1981) 205–2013.
[8] A.S. Vieira, R.A. Stewart, R. Lamberts, C.D. Beal, Residential solar water heaters in Brisbane, Australia: Key performance parameters and indicators, Renew. Energy 116 (2018) 120–132.
[9] Y.K. Rashidov, K.Y. Rashidov, I.I. Mukhin, K.T. Suratov, J.T. Orzimatov, S. S. Karshiev, Main reserves for increasing the efficiency of solar thermal energy in heat supply systems, Appl. Sol. Energy 55 (2019) 91–100.
[10] A. Yedilkhan, K. Murat, A. Beibut, K. Aliya, K. Ainur, M. Tumur, D. Azhibek, B. C. Babu, Mathematical justification of thermosyphon effect main parameters for solar heating system, Cogent Engineering 7 (1) (2020), 185629.
[11] K. Varun, U.C. Arunachala, P.K. Vijayan, Experimental demonstration of the performance of the novel thermosyphon heat transport device and comparison with CFD predictions, Int. J. Therm. Sci. 176 (2022), 107503.
[12] EN 12976-2, Thermal Solar Systems and Components - Factory Made Systems - Part 2: Test Methods, 2019.
[13] Version 2.7, In-Situ Scientific Software, Dynamic System Testing Program, 1996. ISS, Kriegerstr. 23 d, D-82110 Germering.
[14] ISO 9459-2, Solar Heating — Domestic Water Heating Systems — Part 2: Outdoor Test Methods for System Performance Characterization and Yearly Performance Prediction of Solar-Only Systems, 1995.
[15] ISO 9459-5, Solar Heating — Domestic Water Heating Systems — Part 5: System Performance Characterization by Means of Whole-System Tests and Computer Simulation, 2007.
[16] S.S. Haykin, Neural Networks and Learning Machines, Pearson, Upper Saddle River, NJ (USA), 2009.
[17] C. Correa-Jullian, E. López Droguett, J.M. Cardemil, Operation scheduling in a solar thermal system: a reinforcement learning based Framework, Appl. Energy 268 (2020), 114943.
[18] M. Vakili, S.A. Salehi, A review of recent developments in the application of machine learning in solar thermal collector modelling, Environ. Sci. Pollut. Res. Int. 30 (2) (2023) 2406–2439.
[19] C. Correa-Jullian, J.M. Cardemil, E. López Droguett, M. Behzad, Assessment of Deep Learning techniques for Prognosis of solar thermal systems, Renew. Energy 145 (2020) 2178–2191, https://doi.org/10.1016/j.renene.2019.07.100.
[20] M. Abedi, X. Tan, J.F. Klausner, M.S. Murillo, A. Benard, A Comparison of the Performance of a Data-Driven Surrogate Model of a Dehumidifier with Mathematical Model of Humidification-Dehumidification System, AIAA SCITECH 2023 Forum, 2023, pp. 2023–2329.
[21] M. Quartulli, A. Gil, A.M. Florez-Tapia, P. Cereijo, E. Ayerbe, I.G. Olaizola, Ensemble surrogate models for fast LIB performance predictions, Energies (14) (2021) 4115.
[22] L. Cai, L. Ren, Y. Wang, W. Xie, G. Zhu, H. Gao, Surrogate Models Based on Machine Learning Methods for Parameter Estimation of Left Ventricular Myocardium", vol. 8, Royal Society Open Science, 2021, 201121, https://doi.org/10.1098/rsos.201121.
[23] A.H. Elsheikh, S.W. Sharshir, M.A. Elaziz, A.E. Kabeel, W. Guilan, Z. Haiou, Modeling of solar energy systems using artificial neural network: a comprehensive review, Sol. Energy 180 (2019) 622–639, https://doi.org/10.1016/j.solener.2019.01.037.
[24] I. Farkas, P. Géczy-Vig, Neural network modelling of flat-plate solar collectors, Comput. Electron. Agric. 40 (1–3) (2003) 87–102, https://doi.org/10.1016/S0168-1699(03)00013-9.
[25] S.A. Kalogirou, Prediction of flat-plate collector performance parameters using artificial neural networks, Sol. Energy 80 (3) (2006) 248–259, https://doi.org/10.1016/j.solener.2005.03.003.
[26] M.A. Hamdan, E.A. Abdelhafez, A.M. Hamdan, R.A.H. Khalil, Heat transfer analysis of a flat-plate solar air collector by using an artificial neural network, J. Infrastruct. Syst. 22 (4) (2016) A4014004.
[27] S.A. Kalogirou, S. Panteliou, A. Dentsoras, Modeling of solar domestic water heating systems using artificial neural networks, Sol. Energy 65 (6) (1999) 335–342, https://doi.org/10.1016/S0038-092X(99)00013-4.
[28] S.A. Kalogirou, S. Panteliou, Thermosiphon solar domestic water heating systems: long-term performance prediction using artificial neural networks, Sol. Energy 69 (2) (2000) 163–174, https://doi.org/10.1016/S0038-092X(00)00058-X.
[29] E. Dikmen, M. Ayaz, H.H. Ezen, E.U. Küçüksille, A.S. Şahin, Estimation and optimization of thermal performance of evacuated tube solar collector system, Heat Mass Tran. (50) (2014) 711–719, https://doi.org/10.1007/s00231-013-1282-0.
[30] M.W. Ahmad, J. Reynolds, Y. Rezgui, Predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees, J. Clean. Prod. 203 (2018) 810–821, https://doi.org/10.1016/j.jclepro.2018.08.207.
[31] L. Breiman, Random Forests", vol. 45, Machine Learning, 2001, pp. 5–32, https://doi.org/10.1023/A:1010950718922.
[32] J.A. Garcia-Moya, J. Casado, I. Marco, C.M. Fernández-Peruchena, M. Gastón, Deterministic and Probabilistic Weather Forecasting", Research Report AEMET, 2016.
[33] A. Gonzalez-Arceo, M. Zirion-Martinez de Musitu, A. Ulazia, M. del Rio, O. Garcia, Calibration of reanalysis data against wind measurements for energy production estimation of building integrated savonius-type wind turbine, Appl. Sci. 10 (24) (2020) 9017, https://doi.org/10.3390/app10249017.
[34] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random Forest: a classification and regression tool for compound classification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (2003) 1947–1958.
[35] D.R. Cutler, T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, J. J. Lawler, Random forests for classification in ecology, Ecology 88 (2007) 2783–2792.
[36] X. Chen, H. Ishwaran, Random forests for genomic data analysis, Genomics 99 (6) (2012) 323–329.
[37] P. Jiménez-Valero, M. Larrañeta, E. López-García, S. Moreno-Tejera, M.A. Silva-Pérez, I. Lillo-Bravo, Synthetic generation of plausible solar years for long-term forecasting of solar radiation, Theor. Appl. Climatol. 150 (2022) 649–661, https://doi.org/10.1007/s00704-022-04163-9.
[38] H. Li, Z. Liu, K. Liu, Z. Zhang, Predictive power of machine learning for optimizing solar water heater performance: the potential application of high-throughput screening, Int. J. Photoenergy (2017), https://doi.org/10.1155/2017/4194251. Article ID 4194251.
[39] M.N. Wright, A. Ziegler, Ranger: a fast implementation of random forests for high dimensional data in C++ and R, J. Stat. Software 77 (1) (2017) 1–17, https://doi.org/10.18637/jss.v077.i01.
[40] Solar Keymark Database, 2023. https://solarkeymark.eu/database/. (Accessed 17 December 2022).
[41] GREENoneTEC Solarindustrie GmbH, Tested system: GREENoneTEC TSC 160, Licence number 011-7S2865A, https://solarkeymark.eu/database/. (Accessed 17 December 2022).
[42] S.p.A. Ariston, Tested system: Zelios theermo CF-GR 150/1 TR/TT/TT DT, Licence number 011-7S2817A, https://solarkeymark.eu/database/. (Accessed 17 December 2022).
[43] J. Vera-Medina, C. Fernandez-Peruchena, J. Guasumba, I. Lillo-Bravo, Performance analysis of factory-made thermosiphon solar water heating systems, Renew. Energy 164 (2021) 1215–1229, https://doi.org/10.1016/j.renene.2020.10.133.