Research article

Estimating solar irradiance using genetic programming technique and meteorological records

Rami Al-Hajj [1,*] **and Ali Assi** [2]

[1] Department of Math and Statistics for Engineering, American University of the Middle East, Egaila, Kuwait

[2] Department of Electrical and Electronics Engineering, International University of Beirut, Lebanon

* Correspondence: Email: rami.alhajj@aum.edu.kw; Tel: +965-94086716.

Abstract: Solar irradiance is one of the most important parameters that need to be estimated and modeled before engaging in any solar energy project. This article describes a non-linear regression model based on genetic programming technique for estimating solar irradiance in a specific region in the United Arab Emirates. The genetic programming is an evolutionary computing technique that enables automatic search for complex solutions. The best nonlinear modeling function that can estimate the global solar radiation on horizontal will be developed taking into account measured meteorological data. A reference approach to model the solar radiation is first presented. An enhanced approach is then presented which consists of multi nonlinear functions of regression in a parallel structure where each function is designed to estimate the global solar irradiance in a specific seasonal period of the year. Statistical analysis measures have been used to evaluate the performance of the proposed approaches. The obtained results are comparable with the outcomes of models developed by other researchers in the field.

Keywords: global solar radiation; estimation; genetic programming; evolutionary computation; meteorological data; statistical analysis; RMSE; MBE

1. Introduction

With the increased concern and interest in energy preservation and environmental protection, the world today is moving into a new era; transition from almost total dependence of the fossil fuel to an increased use of alternative sources of energy. Solar radiation is one of the promising and

potential renewable energy sources especially in regions like UAE.

An accurate and detailed long-term knowledge of the available global solar irradiance on horizontal surfaces is of a major importance for the design and development of solar energy systems in a given region. Information about solar radiation can be obtained by installing expensive measuring sensors (pyranometers) at as many locations as possible in this region thus, requiring daily maintenance and data acquisition; consequently, increasing the cost of collecting solar radiation data. In most of the cases, the potential sites for solar energy implementation are not covered by measuring stations, especially in the deserted regions. Many countries do not have sufficient network of weather stations for collecting solar data. For such regions, empirical models have to be developed using meteorological data from available measurement stations. These models are then used to estimate solar irradiance values at other locations in the region where solar energy systems are planned [2].

UAE is among countries having potential for solar energy where the solar irradiance has significant strength, the average annual solar hours is approximately 3568 h (i.e. 9.7 h/day), which corresponds to an average annual global solar irradiance of approximately 2285 kWh/m^2 (i.e. 6.3 kWh/m^2 per day) [2].

Numerous researchers have developed statistical and empirical regression models to predict the monthly average daily global solar irradiance in their regions using various weather parameters [3–20]. The mean daily sunshine duration and air delta temperature were the most available and commonly used parameters. The most popular model developed by researchers was the linear model by Angström-Prescott. This model establishes a linear relationship between global solar irradiance and sunshine duration taking into account extra-terrestrial solar irradiance and the theoretical maximum daily sunshine hours. Many studies with empirical regression and machine learning models were presented in the literature for many regions around the world. Recently, different models predicting global solar irradiance using various meteorological and climatological variables have been published [16–37].

Assi et al. [25,26] used four meteorological 12-years data between 1995 and 2007 to train and validate a Feed Forward ANN-based estimation system of solar radiation in Al-Ain city in UAE. The authors examined several MLP architectures and tested more than twelve alternatives based on various derivatives of back-propagation training algorithms.

Antonanzas et al. [27] presented a new methodology to build parametric models for the estimation of global solar irradiation. The models were adjusted to specific on-site characteristics based on an evaluation of the variables importance. Authors have adjusted general parametric models such as the Bristow and Campbell BC models [27] with the on-site particularities. The presented methodology was appropriate for the investigated case study. The daily range of maximum and minimum temperatures, the logical variable of rainfall, and the daily mean wind speed were among the parameters that showed higher correlation with solar irradiance and that were included in the newly developed models

Ahmed and Adam [28] applied a feed forward back- propagation neural network on weather data measured at Qena-Egypt during the year 2007. The proposed approach used location coordinates and sunshine hours to estimate monthly average daily global solar radiation. The authors presented a comparative study between the described MLP-based approach and other empirical models. Based on their experimental results, authors showed the advantages of the MLP-based estimation technique for solar radiation estimation over the existing empirical regression models. Khatib et al. [29] developed a feed forward multi-layer perceptron with four inputs: longitude,

latitude, day of the month, and sunshine radiation to predict the clearance index. The clearance index helped in calculating the solar irradiation. The models used long term solar radiation data for 28 sites in Malaysia measured between years 1984 and 2004.

Ramedani et al. [30] investigated two models based on Support Vector Regression technique (SVR) which is a type of Support Vector Machines (SVM) for predicting Global solar Radiation GSR in Tehran province. The authors examined two kernel functions for SVR: a radial basis function and a polynomial function. The authors designed and validated their approach on a measured daily data consisting on Temperature and sunshine parameters and belonging to seven-year period. The proposed approach, mainly the one based on radial basis function (SVR-rbf) showed better performance when compared to an ANN-based and a Neuro-Fuzzy based systems. Olatomiwa et al. [31] proposed a hybrid approach for predicting solar radiation based on SVMs coupled with the meta-heuristic Firefly algorithm FFA. The FFA has been applied to detect the optimal parameters of the SVM algorithm. The performance of the proposed approach showed superiority comparing to others based on ANN and GP when they have been tested on temperature and sunshine hour's data records collected from three different regions in Nigeria. Mohamadani et al. [32] presented a comparative evaluation among three soft computing methodologies for estimating global solar radiation in a specific region in Iran based on temperature measures. The developed models are an Adaptive Neuro-Fuzzy inference system ANFIS, a radial basis function SVR (SVR-rbf), and a polynomial basis function SVR (SVR-poly). The statistical analysis showed a superiority for the SVR-rbf over the two remaining examined models when validated on daily temperature measures. Kizi [33] proposed a Fuzzy-Genetic FG approach to model and predict solar radiation. The heuristic genetic algorithm has been used to find the optimum parameters of the Fuzzy inference method. The author used latitude, longitude, and altitude as inputs to the FG model to estimate one month ahead solar radiation in some regions in Turkey.

Recently, many researchers tried to instigate the most significant meteorological variables and parameters for estimating and predicting GSR [34,35]. Mohamadi et al. [34] examined the influence of meteorological parameters on horizontal GSR. They examined nine climatological parameters collected from three different cities in Iran. The authors applied an adaptive neuro-fuzzy inference technique in their selection procedure, and they determined the most influential parameters' combinations for each city and have concluded that it is not possible to introduce an optimal combination of inputs for all cities. They justified their conclusion by the fact that GSR, which is special for each region, depends on climate conditions and geographical location that are special for each region.

Demirhan and Atilgan [35] presented a robust coplot optimization approach coupled with a GP technique for solar radiation estimation. The robust coplot analysis technique has been applied on the measures of solar radiation and other related parameters to identify the optimal set of covariates in a data that consists of solar radiation, meteorological, and terrestrial variables. The main goal was to handle the multicollinearity problem that may exist among variables, and to eliminate the effect of outliers on the space of solar radiation modeling. The optimal set of covariates have been then used in a GP technique to construct monthly and yearly solar radiation estimation models. Pan I et al. [36] presented a GP-based approach for predicting solar radiation using six geographical and sunshine duration data from India. The authors introduced what they called Multi-Gene Genetic Programming (MGGP) models where each individual solution, named a gene, is composed by a weighted combination of sub-individuals named Single-Gene Genetic Programming (SGGP) models. The

authors indicated that the MGGP based approach has outperformed the other ones based on simple individual SGGP models as well as other classical regression models.

The current article investigates the prediction of global solar irradiance on horizontal using evolutional computational technique, namely the genetic programming (GP). Recently, the GP techniques showed good performance and flexibility in modelling non-linear regression problems [38,39]. Practically, The GP demonstrates its advantages in dynamically building complex formulas (solutions), and its flexibility to choose a set of functions and operators that match the problem to be solved. Such flexibility is possible due to the fact that the structure of the binary trees that represent solution candidates can be dynamically changed during the evolutionary process. These characteristics give the GP the ability to skip out of the local minima problem commonly found in the neural networks models especially in their feed-forward structures with back-propagation training algorithms. In the solar radiation estimation literature, the Genetic Algorithm GA has been used to select the optimal parameters of machine learning based models [33], whereas the GP algorithm has been used as core models of estimation [35,36].

In this article, the design and validation of a new GP based approach to estimate global solar irradiance using meteorological data will be described. The main idea is to find the best model for the relation between a set of meteorological parameters and the solar irradiance on a specific geographical area. Two approaches have been validated: A reference approach that consists of one global model that estimates the solar irradiance with respect to four climatological parameters, and a second approach that consists of a set of several models in a parallel structure. Each model consists of a nonlinear function that is dedicated to estimate the global solar irradiance in a specific seasonal or bi-seasonal period of the year. The experimental results indicated the advantages of using such type of multi-model structure when dealing with a set of data with large variability during the year. The remaining of this article is organized as follows: In section 2, the genetic programming is explained as an optimization heuristic technique, the GPLAB toolbox of MATLAB® that has been used in the adopted approach is then introduced. In section 3, the GP based reference approach that consists of one estimation function is described. Then, an enhanced approach also based on GP is given. Moreover, the dataset used for the design and the validation of the proposed approaches is introduced and described in this section. Discussion of the results is presented in section 4. Finally, section 5 includes conclusions and future perspectives.

2.   Genetic Programming and GPLAB Toolbox

2.1. Genetic Algorithm

The GP is an extension of the conventional genetic algorithm [40,41]. Genetic Algorithm (GA) is a metaheuristic method usually used to find an optimal solution in optimization problems based on a natural selection process. The GA starts by an initial population of random individuals. Each individual is represented by what is called chromosome that is an array of genes. Each gene represents a parameter to be optimized. Every individual (chromosome) represents a possible solution of the optimization problem and has its own fitness measure. The fitness is a measure for each individual that indicates how the solution related to this individual is suitable to solve the problem.

The GA uses the so called genetic operators: crossover, mutation, and cloning to evolve from a

population to another until it reaches a population that consists of an optimal solution based on a chosen fitness objective function [40,41]. It starts by an initial random population of individuals. Then, in each generation, the GA performs the following steps:

- Select from the present population the individuals that have the best computed fitness.
- Use the best individuals to generate the next population by the crossover of those individuals. This is similar to the biological reproduction based on natural selection. A new individual has part of its genes coming from the first parent and the other part from the second parent.
- Based on a computed probability, a mutation operator may be applied to one or many chromosomes by changing the value of one of its genes. Similarly, and based on another computed probability, a chromosome with good fitness may be cloned and promoted to evolve to the next generation.

This procedure is repeated until an optimum individual is found. Figure 1 illustrates the operations of crossover and mutation of two individuals to produce new individuals for the next generation.
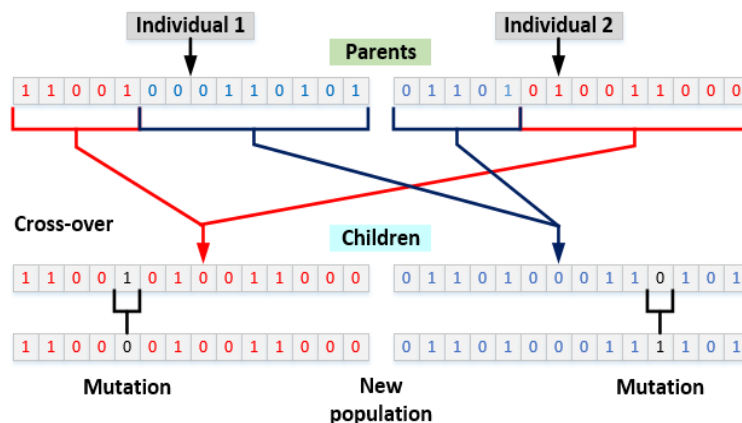

Figure 1. The crossover and mutation operators of the GA. The crossover produces two new children individuals from two parent individuals, and the mutation changes randomly the values of one or more than one gene.

2.2. Genetic Programming

The GP aims to find the best computer program (function) that is composed of both data and operators and that solves a specific problem [41,42,43]. A chromosome in GP is represented by a binary tree data structure where internal nodes represent algebraic and/or logical operators whereas the external ones represent numbers and parameters related to the problem to be solved. Figure 2 shows examples of binary trees that represent mathematical expressions/functions. A function can be considered as a computer program that consists of a set of data (terminals) and actions (operators).

The process of evolution in GP starts by an initial random population of chromosomes that represent possible solutions (functions) and tries to generate new populations subsequently [42].
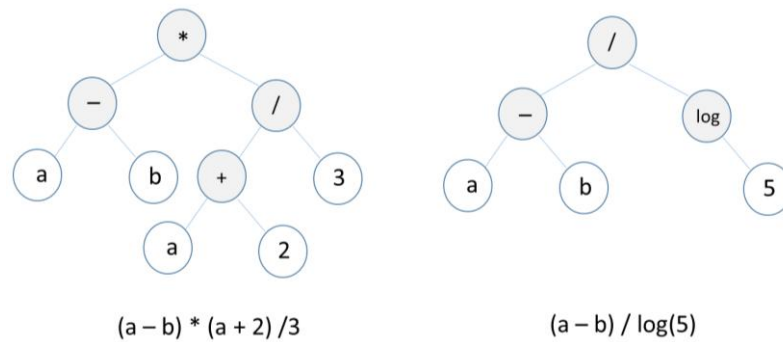
Figure 2. Examples of binary trees that represent algebraic functions.

The evolution is controlled by a fitness function that is equivalent to the objective function adopted in local heuristic search techniques. The fitness function is special for each optimization problem and allows the evaluation of fitness of each chromosome (solution) in a population. Figure 3 illustrates the effect of the crossover operator on two selected individuals.
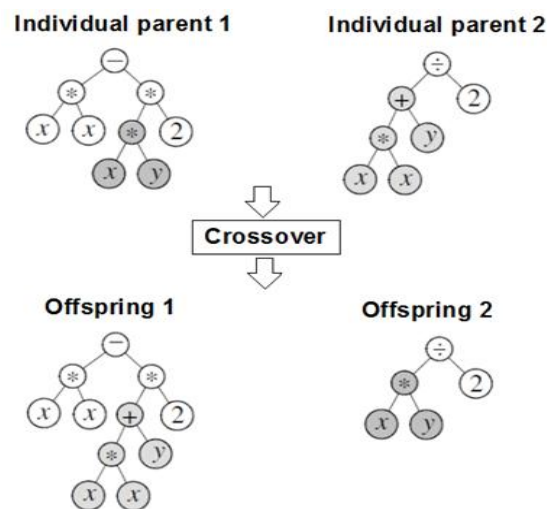

Figure 3. The crossover of two individuals to produce two new ones for the next generation.

2.3. GPLAB toolbox

GPLAB is a Genetic Programming toolbox for MATLAB® [44]. GPLAB provides most of the features and operators commonly used in GP. Its modular structure allows considering it as an extendable tool that is suitable for prototyping new techniques of heuristic local search in GP. GPLAB enables a set of facilities and features to handle and control the structure and the size of both: the chromosomes that represent individual solutions and the populations that represent sets of those individuals. In addition, GPLAB allows the dynamic control of the variable size of populations during run time. This feature is indeed important in case of limited computational resources [45,46,47]. Moreover, GPLAB implements a technique for automatically adjusting the probabilities of adopted genetic operators during runtime. This feature allows the use of the GPLAB toolbox as a test

workbench for new genetic operators

3. Genetic Programming Based Systems

In this work, a GP based approach to estimate solar irradiance is designed, implemented, and validated. In the first phase of this work, a reference system is proposed, which consists of a single function that can model the relation between solar irradiance and a set of climatological factors in a specific geographical area. The reference system showed promising performance. In the second phase, the performance of the reference system is analyzed and an enhanced one that consists of multiple independent models is suggested. Each model is dedicated to estimate the solar radiation amount in a specific seasonal period of the year. All components of the two proposed approaches were designed, implemented, and validated using the GPLAB toolbox.

In the design and validation phases, a meteorological dataset provided by the National Center of meteorology and Seismology (NCMS) in Abu-Dhabi—UAE is used. The dataset consists of daily data records for the period between 2004 and 2007. Each daily record includes the measures of: air temperature, wind speed, relative humidity, and sunshine duration. The dataset has been divided into two subsets; a design subset having records for the years between 2004 and 2006 inclusive and a test subset that includes records for the year 2007. Table 1 shows some samples of the dataset. In this table, the first four columns represent the four meteorological records.

Table 1. Samples of meteorological records (dataset).

Max Temperature	Mean wind Speed (knot)	Sun Hours	Mean Relative Humidity %	Total Radiation (kWh/sq.m)
25.7	5.6	8.9	68	462.8
24.5	7.1	9.6	67	459.3
23.0	5.4	10.2	54	515.5
25.1	6.4	10.1	47	505.0
23.6	8.3	1.6	63	230.6

3.1. The Reference System

The problem of developing an appropriate function that models the relation previously discussed, looks like a search problem for an optimal state that represents the expected function. In our case, the targeted function includes operand and operators. The operands consist of the climatological parameters and other constants that may appear in the resulting function. Besides, the set of operators may include arithmetic operators, exponential operators, and any other algebraic or non-algebraic ones like: square root, natural log, exponentials, etc.

The functions investigated in this work can be represented by binary trees data structures, where the internal nodes represent the operators whereas the terminals represent the operands. Figure 4 illustrates an example of a binary tree that represents a solution candidate. In Figure 4, X_1, X_2, X_3, and X_4 represent temperature, wind speed, sun hours, and humidity respectively.
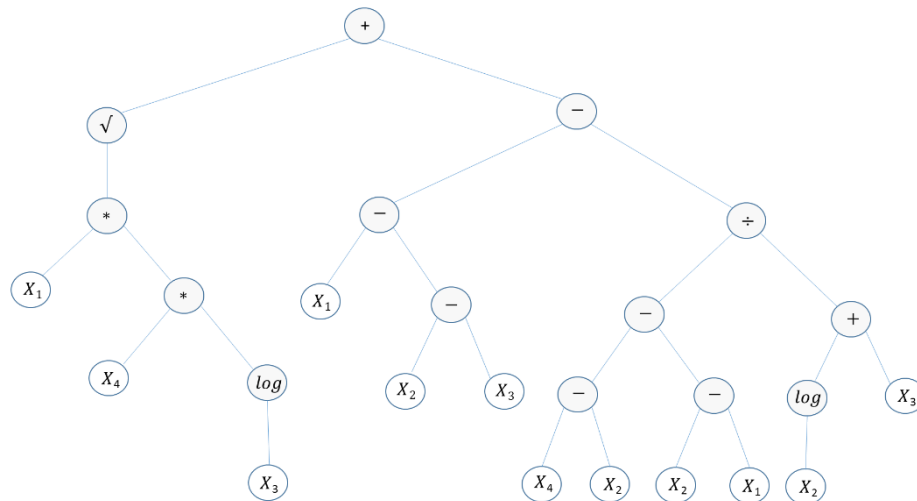
Figure 4. An example of a binary tree that represents a function for modeling (estimating) solar radiation.

Equation (1) shows the algebraic function that can be obtained using the binary tree of Figure 4. In this equation, *sr, tmp, sh, ws*, and *hum* stand for solar radiation, temperature, sun hours, wind speed, and humidity respectively. The internal nodes in the tree include the four arithmetic operators: addition (+), difference (-), and division (/), and multiplication (*), plus it includes the decimal log, as well as the square root ($\sqrt{\ }$). Figure 4 shows that the depth of the tree is equal to six. Actually, the maximum depth allowed for trees in each population is one of the adjustable parameters in a GP evolution process.

$$sr = \sqrt{tmp * hum * \log(sh)} + (tmp - (ws - sh)) - \left(\frac{(hum - ws) - (ws - tmp)}{\log(ws) + sh}\right) \tag{1}$$

In this work, fixed values for the following parameters have been adopted:
- The population size is set to be equal to 500 individuals. In our experiments the performance of optimization has been slightly affected by variations in the size of population.
- The fitness function that evaluates the efficiency of candidate solutions (chromosomes). A fitness function related to the root mean square error (RMSE) has been used.
- The depth of each of the binary trees that represents chromosomes (solution candidate) is set to be dynamic. The maximum value of that depth is chosen to be equal to six. GPLAB provides a technique that permits to start by an initial depth of trees that can be dynamically increased until a selected maximum value.

On the other hand, alternatives have been investigated:
- The probability of applying each of the genetic operators (crossover and mutation).
- The sampling method to select individuals from the current population to participate in generating new individuals for the next generation.

The first three columns in Table 2 show combinations of parameters adopted in designing the proposed approach. The implemented fitness function computes, for each individual, the RMSE between the set of exact output values available in the design dataset and the output values returned by that individual. Equation (2) describes the RMSE computation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(H_{pi} - H_i)^2}{N}} \tag{2}$$

Where H_{pi} represents the estimated value of global solar irradiance, H_i is the measured value that is available in the design dataset, and N is the total number of records in that dataset.

Table 2. Combinations of parameters with the fitness of best individuals in the last population for each combination.

Operators probability	Sampling method	Set of functions	Best fitness in last population
Crossover prob. : 0.85 Mutation prob. : 0.15	Roulette	{arithmetic , $\sqrt{}$, $\log_{10}$}	4.402
Crossover prob. : 0.85 Mutation prob. : 0.15	Tournament	{arithmetic , $\sqrt{}$, $\log_{10}$}	4.491
Dynamic	Roulette	{arithmetic , $\sqrt{}$, $\log_{10}$}	4.503
Dynamic	Tournament	{arithmetic , $\sqrt{}$, $\log_{10}$}	**4.174**

As for the probabilities of applying each of the genetic operators, the GPLAB allows either to fix the values of those probabilities or to dynamically compute them at each iteration during the run time. The computation in this case is based on the history of each operator in producing individuals with the best fitness and on statistics about the newly produced individuals [44]. The results presented in the right most column of Table 1 indicate that the best fitness value (in this case the lowest) is the one related to dynamic probabilities of operators and tournament sampling method. Figure 5 shows the binary tree associated to the best individual in the last population of the best combination. Equation (3) represents the function stored in that binary tree.

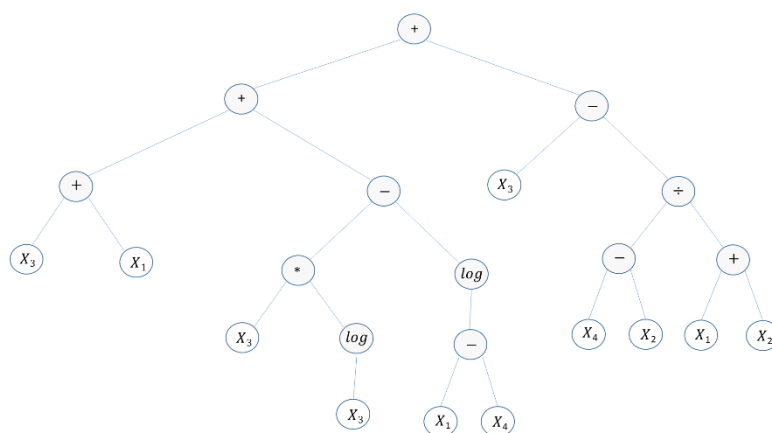$$sr = tmp + sh + sh * \log(sh) - \log|tmp - hum| + sh - \left(\frac{hum - ws}{tmp + ws}\right) \tag{3}$$

Figure 5. Binary tree related to the function of the fittest individual, where X_1 = temperature, X_2 = wind speed, X_3 = sun hours, and X_4 = humidity.

3.2. The Multi-Model System

The results obtained using the reference system show remarkable difference between the measured values and the estimated ones. Figure 7 compares the measured and the estimated monthly average daily global solar irradiance values.

The records of the design dataset have been investigated, and the values of each meteorological factor have been analyzed. Analysis showed that the values of some factors, especially the humidity, have wide variations, i.e. a large deviation around the average value over a year. Such variations make the search for an optimal model quite difficult.

One of the suggestions to improve the whole performance is by estimating the global average solar irradiance over relatively short period of time in a year by using a multi-model approach. The main idea is to find the function with best fitness for estimating the global solar irradiance for each seasonal period of the year. Applying this strategy lead to build proficient functions. A function has been built for each two consecutive months of the year. Thus, the multi-mode system consists of six nonlinear functions. Figure 6 illustrates the structure of the proposed approach.
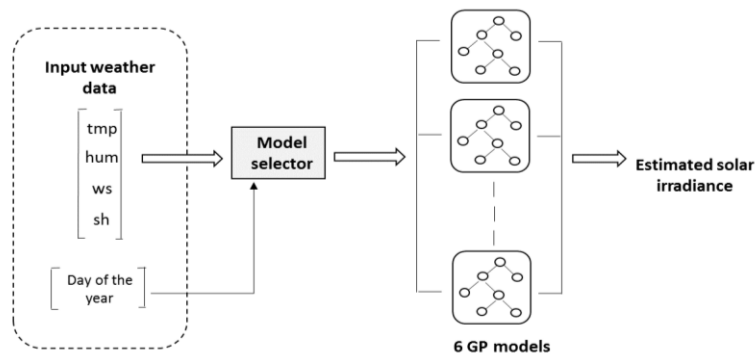

Figure 6. Structure of the multi-model approach.

The evolutional computation process is launched with the same combination of parameters described in the 1st and 4th rows of Table 2. The estimation performance is significantly improved. Table 3 shows the obtained enhancement in terms of fitness of the best individual when the multi-model strategy is applied. The best performance is obtained with dynamic probabilities of genetic operators, tournament sampling, and set of functions that contains arithmetic, algebraic and logarithmic operators.

Table 3. The fitness of the best individual of the last population for the multi-model approach.

Operators probability	Sampling method	Set of functions	Best fitness in last population
Crossover prob. : 0.85 Mutation prob. : 0.15	Roulette	{arithmetic , $\sqrt{}$, $\log_{10}$}	2.137
Dynamic	Tournament	{arithmetic , $\sqrt{}$, $\log_{10}$}	**1.218**

Error statistical analysis showed good improvement as will be described in the next section.

4. Results

The estimation performance of the suggested approach was assessed through a statistical analysis of error. The analysis was conducted by computing the RMSE and Mean Bias Error MBE that measure the variation of estimated values against the measured available ones. Low RMSE and MBE values are desired and indicate an accurate estimation. The RMSE computation is described earlier in equation (2), whereas the MBE computation is described in equation (3).

$$MBE = \frac{\sum_{i=1}^{N}(H_{pi} - H_i)}{N} \tag{3}$$

Table 4 compares the values of RMSE of the best two reference models and the new model that consists of parallel multi-functions.

Table 4. The RMSE and MBE of the best functions.

Operators probability	Best fitness in last population	RMSE	MBE
Operators prob.: [0.85, 0.15] Sampling: roulette Model: reference model	4.402	1.862	1.322
Operators prob. : dynamic Sampling: tournament Model: reference model	4.174	1.026	0.68
Operators prob. : dynamic Sampling: tournament Model: Parallel	**1.218**	**0.210**	**0.052**

Figure 7 and Figure 8 show the measured and estimated values of monthly average daily global solar irradiance for the best reference model and the multi- model. The later shows better performance in estimating monthly average daily global solar irradiance.
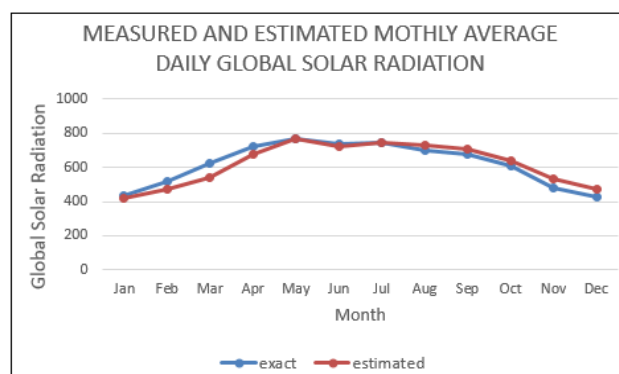

Figure 7. Monthly average daily global solar irradiance estimation of the best reference system.
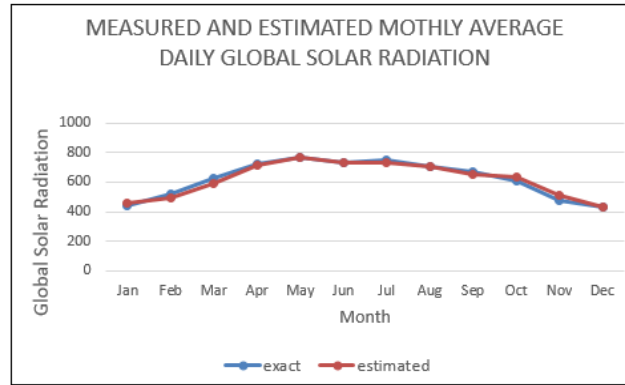
Figure 8. Monthly average daily global solar irradiance estimation of the multi-model system.

Figure 9 shows the binary tree that represents one of the best functions in the multi-model approach. Equation (4) represents the function of the binary tree shown in this Figure.
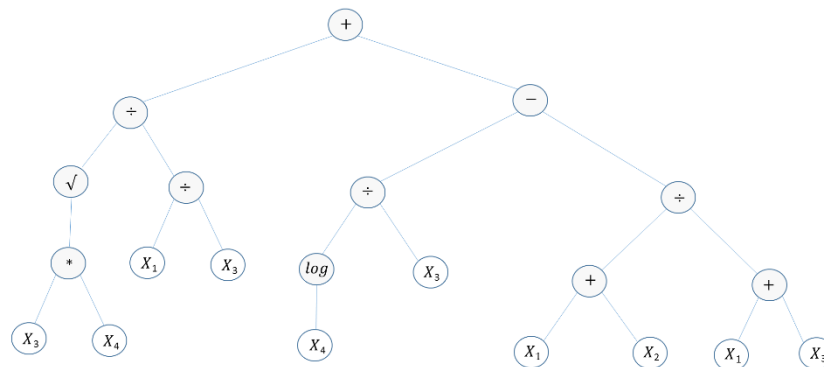

Figure 9. Binary tree that represents one of the functions in the multi-model system.

$$sr = \frac{\sqrt{sh * hum}}{\frac{tmp}{sh}} + \frac{log\,(hum)}{sh} - \frac{tmp + ws}{tmp + sh} \tag{4}$$

The suggested genetic programming based approaches show comparable performance with respect to other empirical regression and neural models. Table 5 compares the RMSE of the results obtained by the suggested approach to those obtained by other models conducted by other groups.

Table 5. The RMSE of the multi-functions model compared to other models.

Model	RMSE
Quadratic Regression model	0.214
Logarithmic Regression model	0.259
ANN 3-10-4-1	0.374
ANN 3-20-8-1	0.391
GP model with parallel functions	**0.210**

5.   Conclusions and Perspectives

This article described new approaches for estimating global solar irradiance using meteorological records. The suggested methods are based on a GP heuristic technique. The first method (reference system) consists of estimating the nonlinear function that can model the relation between solar irradiance and four meteorological parameters. The performance of the reference system is promising. An enhanced model that consists of multi-function was proposed and it showed better performance with respect to the first method. The performance of the proposed approaches was evaluated using statistical analysis measures.

The GP showed its advantages in dynamically building complex formulas that represent solution candidates for the problem to be solved. As an evolutionary process, the GP shows its ability to resolve the local minima problem and to converge toward a global minima. The problem of local minima is commonly found in the neural networks models especially in their feed-forward structures commonly used in the literature coupled with the classical back-propagation training algorithm. Moreover, the GP technique provides analytical expressions as solutions, like the expression given in Equation (3), which is not available in most of the machine learning techniques, for instance the neural and the neuro-fuzzy models. The later property is important for researchers to understand the contribution of each variable input in the calculation of the dependent variable output.

In our experiments, we controlled the well-known bloat (inflation) problem of GP by using a set of techniques provided by the GPLAB environment [44]. Some of those techniques consist of automatic resizing of the population in runtime to save computational resources. Those techniques are adequate in cases of complex problems when the complexity of the expressions' models increases dramatically during the evolutionary process.

A similar approach to our enhanced one has been proposed in [35] with two main differences: First, in that approach the input data has been pre-analyzed by using an optimization technique to handle the multicollinearity problem that may exist among the variables, which is not available in our approach that has been applied on data records of four meteorological variables. Second, the approach in [35] consists of a set of twelve models, each model is dedicated to estimate the solar radiation in a specific month of the year whereas our approach consists of six models. Each of our parallel models is devoted to estimate the solar radiation for a semi-seasonal period of the year. In general, the increase of number of learning-based models requires more training data records to adjust those models during the design phase which may not be always possible. One of our future suggestions is to automate the splitting of data into seasonal subsets based on an automatic learning technique in order to optimize the estimation performance.

In this work, the proposed GP approach is not compared to other learning based estimation techniques of the same type. Such comparison needs comparing the convergence time of each approach as well as the performance of estimation using the same data sets and same leaning parameters. This could be one of our future perspectives. On the other hand, the obtained results in this work are comparable to those obtained by mathematical regressions and neural models that were conducted by other research groups. Finally, the obtained results showed the advantage of using the parallel modular structure over the global one.

Three main future perspectives can be drawn for this work. The first one consists of finding a way to automatically splitting the data into seasonal or semi seasonal periods in order to optimize the performance of estimation. The second perspective consists of comparing or GP based approach with

other machine learning based techniques by using the same data sets. The third perspective consists of validating the proposed approach using new meteorological datasets with larger number of weather parameters in each record.

Acknowledgments

Conflict of Interest

All authors declare on conflicts of interest in this paper.

References

1. Kassem AS, Aboukarima AM, EL Ashmawy NM (2009) Development of Neural Network Model to Estimate Hourly Total and Diffuse Solar Radiation on Horizontal Surface at Alexandria City (Egypt). *J Appl Sci Res* 5: 2006-2015.
2. Assi A, Jama M (2010) Estimating Global Solar Radiation on Horizontal from Sunshine Hours in Abu Dhabi –UAE. *Advances in Energy Planning, Environmental Education and Renewable Energy Sources, 4th WSEAS international Conference on Renewable Energy Sources*, 101-108.
3. Podestá G, Núñez L, Villanueva C, et al. (2004) Estimating daily solar radiation in the Argentine Pampas. *Agr Forest Meteorol* 123: 41-53.
4. Almorox J, Benito M, Hontoria C (2008) Estimation of global solar radiation in Venezuela. *Interciencia* 33: 280-283.
5. Falayi E, Adepitan J, Rabiu A (2008) Empirical models for the correlation of global solar radiation with meteorological data for Iseyin, Nigeria. *Int J Phys Sci* 3: 210-216.
6. Fortin J, Anctil F, Parent L, et al. (2008) Comparison of empirical daily surface incoming solar radiation models. *Agr Forest Meteorol* 148: 1332-1340.
7. Togrul I, Togrul H (2002) Global solar radiation over Turkey: Comparison of predicted and measured data. *Renew Energy* 25: 55-67.
8. Walthall C, Dulaney W, Anderson M, et al. (2004) A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. *Remote Sens Environ* 92: 465-474.
9. Bakirci K (2009) Correlation for estimation of daily global solar radiation with hours of bright sunshine in Turkey. *Energy* 34: 485-501.
10. Tadros M (2000) Uses of sunshine duration to estimate the global solar radiation over eight meteorological stations in Egypt. *Renew Energy* 21: 231-246.
11. Al-Lawati A, Dorvlo A, Jervase J (2003) Monthly average daily solar radiation and clearness index contour maps over Oman. *Energ Convers Manage* 44: 691-670.
12. Zhou J, Yezheng Wu, Gang Y (2005) General formula for estimation of monthly average daily global solar radiation in China. *Energ Convers Manage* 46: 257-268.
13. Ball R, Purcell C, Carey S (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J* 96: 391-397.

14. Menges H, Ertekin C, Sonmete M (2006) Evaluation of solar radiation models for Konya, Turkey. *Energ Convers Manage* 47: 3149-3173

15. Şahin A (2007) A new formulation for solar irradiation and sunshine duration estimation. *Int J Energy Res* 31: 109-118.

16. Ulgen K, Hepbasli A (2002) Comparison of solar radiation correlations for Izmir, Turkey. *Int J Energy Res* 26: 413-430.

17. Angström A (1924) Solar and terrestrial radiation. *Quart J Roy Met Soc* 50: 121-125.

18. Boccol M, Willington E, Arias M (2010) Comparison of Regression and Neural Networks Models to estimate Solar Radiation. *Chilean J Agr Res* 70: 428-435.

19. Mohandes M, Rehman S, Halawani T (1998) Estimation of Global Solar Radiation Using Artificial Neural Networks. *Renew Energy* 14: 179-184.

20. Mohandes M, Balghonaim A, Kassas M, et al. (2000) Use of Radial Basis Functions for Estimating Monthly Mean Daily Solar Radiation. *Solar Energy* 68: 161-168.

21. Rehman S, Mohandes M (2008) Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* 36: 571-576.

22. Tasadduq I, Rehman S, Bubshait K (2002) Application of neural networks for the prediction of hourly mean surface temperature in Saudi Arabia. *Renew Energy* 25: 545-554.

23. Krishnaiah T, Srinivasa RS, Madhumurthy K, et al. (2007) A Neural Network Approach for Modelling Global Solar Radiation. *Appl Sci Res* 3: 1105-1111.

24. Elminir H, Areed F, Elsayed T (2005) Estimation of solar radiation components incident on Helwan site using neural networks. *Solar Energy* 79: 270-279.

25. Assi A, Al-Shamisi M, Jama M (2010) Prediction of Monthly Average Daily Global Solar Radiation in Al Ain City–UAE Using Artificial Neural Networks, *Proceedings of the 4th International Conference on Renewable Energy Sources* , *Tunisia*, 109-113.

26. Assi A, Al-Shamisi M (2010) Prediction of Monthly Average Daily Global Solar Radiation in Al Ain City–UAE Using Artificial Neural Networks. *Proceedings of the 25th European Photovoltaic Solar Energy Conference, Spain*, 508-512.

27. Antonanzas-Torres F, Sanz-Garcia A, Martınez-de-Pison F, et al. (2013) Evaluation and improvement of empirical models of global solar irradiation: Case study northern Spain. *Renew Energy* 60: 604-614.

28. Ahmed A, Adam M (2013) Estimate of Global Solar Radiation by using Artificial Neural Nework in Qena, Upper Egypt. *J Clean Energy Technol* 1(2): 148-150.

29. Khatib T, Mohamed A, Mahmoud M, et al. (2012) Estimating Global Solar Energy Using Multilayer Perception Artificial Neural Network. *Int J Energy* 6(1): 82-87.

30. Ramedani Z, Omid M, Keyhani A, et al. (2014) Potential of radial basis function based support vector regression for global solar radiation prediction. *Renew Sust Energy Rev* 39:1005-1011.

31. Olatomiwa L, Mekhilefa S, Shamshirband S, et al. (2015) A support vector machine‒firefly algorithm-based model for global solar radiation prediction. *Solar Energy* 115: 632-644.

32. Mohammadi K, Shamshirband S, Danesh AS, et al. (2016) Temperature-based estimation of global solar radiation using soft computing methodologies. *Theor Appl Climatol* 125: 101-112

33. Kisi O (2014) Modeling solar radiation of Mediterranean region in Turkey by using fuzzy genetic approach. *Energy* 64: 429-436

34. Mohammadi K, Shamshirband S, Kamsin A, et al. (2016) Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. *Renew Sust Energy Rev* 63: 423-434.

35. Demirhan H, Atilgan Y (2015) New horizontal global solar radiation estimation models for Turkey based on robust coplot supported genetic programming technique. *Energy Convers Manage* 106: 1013-1023.

36. Pan I, Pandey DS, Das S (2013) Global solar irradiation prediction using a multi-gene genetic programming approach. *J Renew Sust Energy* 5: 063129.

37. Baser F, Demirhan H (2017) A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation. *Energy* 123: 229-240.

38. Schwaerzel R, Bylander T (2006) Predicting currency exchange rates by genetic programming with trigonometric functions and high-order statistics. *Genetic and Evolutionary Computation Conference Gecco'06* 1: 955-956.

39. Agapitos A, Dyson M, Kovalchuk J, et al. (2008) On the genetic programming of time-series predictors for supply chain management, *proceedings of the 10th annual conference on genetic and evolutionary computation, Atlanta* 1: 1163-1170.

40. Srinivas M, Patnail L (1994) Genetic algorithms, A survey. *IEEE Computer* 27(6): 17-26.

41. Poli R, Langdon W, McPhee N, et al. (2007) Genetic programming, An introductory tutorial and a survey of techniques and applications. University of Essex, UK, Tech Rep CES-475. Available from: http://cswww.sx.ac.uk/staff/rpoli/technical-reports/tr-ces-475.pdf

42. Poli R, Langdon WB, McPhee NF (2008) *A field guide to genetic programming.* Available from: http://www.gp-field-guide.org.uk/

43. Riolo R, Vladislavleva E, Moore J (2011) *Genetic programming theory and practice IX*. Springer Science & Business Media, ISBN 1461417708.

44. Silva S, Almeida J (2003) A Genetic programming toolbox for MATLAB, Version 3. *ECOS Evolutionary and complex Systems Group*, University of Coimbra-portugal.

45. Silva S, Costa E (2005) Resource-Limited genetic programming: The dynamic approach. *Proceedings of GECCO-2005*, 1673-1680.

46. Silva S, Costa E (2004) Dynamic limits for bloat control - Variations on Size and Depth. *Proceedings of GECCO'04*, 666-677.

47. Silva S, Almeida J (2003) Dynamic maximum tree depth - A Simple technique for avoiding bloat in tree-based GP. *Proceedings of GECCO-2003*, 1776-1787.