sustainability

MDPI

Article

Prognostication of Shortwave Radiation Using an Improved No-Tuned Fast Machine Learning

Isa Ebtehaj [1], Keyvan Soltani [2], Afshin Amiri [3], Marzban Faramarzi [4], Chandra A. Madramootoo [5] and Hossein Bonakdari [1,*]

[1] Department of Soils and Agri-Food Engineering, Université Laval, Québec, QC G1V 0A6, Canada; isa.ebtehaj.1@ulaval.ca
[2] Department of Civil Engineering, Razi University, Kermanshah 6714967346, Iran; keyvansoltanii@gmail.com
[3] Department of Remote Sensing and GIS, University of Tehran, Tehran 1417935840, Iran; afshinamirii@yahoo.com
[4] Rangeland and Watershed Management Group, Faculty of Agriculture, Ilam University, Ilam 69315516, Iran; m.faramarzi@ilam.ac.ir
[5] Department of Bioresource Engineering, McGill University, Quebec, QC H9X 3V9, Canada; chandra.madramootoo@mcgill.ca
* Correspondence: hossein.bonakdari@fsaa.ulaval.ca; Tel.: +1-418-656-2131

Abstract: Shortwave radiation density flux (SRDF) modeling can be key in estimating actual evapotranspiration in plants. SRDF is the result of the specific and scattered reflection of shortwave radiation by the underlying surface. SRDF can have profound effects on some plant biophysical processes such as photosynthesis and land surface energy budgets. Since it is the main energy source for most atmospheric phenomena, SRDF is also widely used in numerical weather forecasting. In the current study, an improved version of the extreme learning machine was developed for SRDF forecasting using the historical value of this variable. To do that, the SRDF through 1981–2019 was extracted by developing JavaScript-based coding in the Google Earth Engine. The most important lags were found using the auto-correlation function and defined fifteen input combinations to model SRDF using the improved extreme learning machine (IELM). The performance of the developed model is evaluated based on the correlation coefficient (R), root mean square error (RMSE), mean absolute percentage error (MAPE), and Nash–Sutcliffe efficiency (NSE). The shortwave radiation was developed for two time ahead forecasting (R = 0.986, RMSE = 21.11, MAPE = 8.68%, NSE = 0.97). Additionally, the estimation uncertainty of the developed improved extreme learning machine is quantified and compared with classical ELM and found to be the least with a value of ±3.64 compared to ±6.9 for the classical extreme learning machine. IELM not only overcomes the limitation of the classical extreme learning machine in random adjusting of bias of hidden neurons and input weights but also provides a simple matrix-based method for practical tasks so that there is no need to have any knowledge of the improved extreme learning machine to use it.

Keywords: water resources; Daymet V3; Google Earth Engine; improved extreme learning machine (IELM); sensitivity analysis; shortwave radiation flux density; sustainable development

Citation: Ebtehaj, I.; Soltani, K.; Amiri, A.; Faramarzi, M.; Madramootoo, C.A.; Bonakdari, H. Prognostication of Shortwave Radiation Using an Improved No-Tuned Fast Machine Learning. *Sustainability* **2021**, *13*, 8009. https://doi.org/10.3390/su13148009

Academic Editor: Roger Jones

Received: 12 June 2021
Accepted: 16 July 2021
Published: 17 July 2021

1. Introduction

Shortwave radiation is of essential significance in climate research since it controls the complete energy exchange between the land/ocean surface and atmosphere [1]. It plays a crucial role in biogeochemical, physical, ecological, and hydrological processes [2]. Shortwave radiation is the energy source that causes photosynthesis, transpiration, evaporation, and other significant process connected to agriculture systems. It is incredibly variable (both temporally and spatially) on the earth's surface. Precise shortwave radiation is essential for evapotranspiration models, which are employed to construct irrigation plans to improve crop yield while saving water and minimizing herbicide, fertilizer, and pesticide applications [3].

Reliable calculation of shortwave radiation is essential as a significant component of the energy budget to understand global change [4–7]. In the clear sky, most of the shortwave radiation is caused by direct sunlight. Shortwave radiation is divided into direct and diffuse components. The direct component is the energy that emanates directly from the sun's rays, and particles and molecules disperse the diffuse component in the air [8]. The amount of diffused shortwave radiation is affected by the height of the sun. Shortwave solar input stands as an essential component of the surface energy balance and is considered to be the principal source of energy on Earth [9–12]. The available energy for hydrological processes such as evaporation and transpiration is strongly affected by solar radiation. Additionally, biological phenomena such as photosynthesis and the carbon cycle are also dependent on solar radiation (direct and diffused radiation) [13,14].

Re-analysis datasets are constantly being improved with increasing access to observational data and advances in modeling and data assimilation systems. The use of re-analysis data has considerable potential for studying areas with a shortage of terrestrial data. Re-analysis data are a combination of ground observations, field surveys, and various models [15]. Different re-analysis datasets are available on a continental and global scale today, e.g., Twentieth-Century Reanalysis, Daily Surface Weather and Climatological Summaries (Daymet), and the Global Land Data Assimilation System. These datasets can provide continuous data and compensate for the gap between terrestrial data.

The Daymet model includes a set of tools designed to calculate estimates of daily weather parameters in Canada, the United States, and Mexico. The main advantages of Daymet are (1) covering the vast majority of North America, (2) providing data on a daily scale for a long period starting from 1980, (3) low spatial resolution (i.e., one kilometer), which is higher than other Reanalysis data. According to the mentioned reasons and increasing the modeling accuracy, Daymet was used to provide shortwave radiation flux density (SRFD) information.

In the last decade, applying the machine learning-based approach to modeling nonlinear complex problems in hydrology and environmental science has attracted many scholars [16–19]. The main advantages of these approaches are high accuracy, low human intervention, and continuous improvement [20–22]. One of the commonly used machine-learning-based techniques is the feedforward neural network, which has been successfully applied in different fields of science. The main training algorithm in this approach is backpropagation. The key advantage of this method is the nonlinear mapping of the independent input variables and dependent output variable(s), which overcomes the limitation of the classical regression-based approaches. It should be noted that although non-linear mapping has several advantages, it preserves some of the limitations of the parent independent input variables. For example, it can preserve the expected value of the autocorrelation function but not the higher-order joint moments and time-asymmetry [23]. The feedforward neural network trained with the backpropagation algorithm is a well-known machine learning method. According to easy implementation, suitable performance, and inherent simplicity [24], it has been successfully applied in different fields of science [25–28]. To overcome the drawbacks of this approach, including slow convergence, time-consuming training [29], and trapping in local minima that leads to low generalizability [30], the extreme learning machine (ELM) [31] was introduced. The ELM is a single-layer feedforward neural network with a fast training process. The main pros of this algorithm are high accuracy, robustness, least user intervention, rapid learning rate, a learning process that requires only a single iteration, and high generalization [32]. However, the main drawback of the ELM is the random generation of the two main matrices, including bias of hidden neurons and input weights. To remove the main limitation of the ELM, an improved version of the ELM is developed based on the orthogonal of the random generation matrices and the definition of an iteration parameter.

According to current knowledge about the influence of non-renewable energies on the environment, renewable energy sources have attracted scholars for their research interests. It is principally because renewable energies are stabilized by natural procedures, which

do not contribute to climate change, global warming, and greenhouse gases. With the advancement of science and the discovery of solar radiation as a sustainable source of renewable energy, it has been considered to overcome the problems caused by fossil fuels. Moreover, the Sustainable Development Goals, which include various economic, social, and environmental goals, provide a global framework that the United Nations member states are committed to achieving [33]. The seventh goal of sustainable development is to improve access to clean and affordable energy, which aims to ensure that sustainable, reliable, modern, and affordable energy is accessible to everyone. To achieve this goal, energy consumption and efficiency must be controlled and monitored. Solar radiation as a renewable energy source can contribute to the seventh goal of sustainable development [34]. In this regard, it is necessary to develop practical and appropriate models for each region to achieve sustainable development goals. This study tries to establish a suitable model for SRFD estimation in Nunavik, which is one of the coldest regions of the world, so that it can be used to play an effective role in environmental processes such as vegetation management, water resources management, lake management, control of changes in land use, and so on.

The primary purpose of the current study is to build an improved version of the ELM known as IELM for monthly short-term prediction of the Shortwave Radiation Flux Density (SRFD) in the Nunavik region, Quebec, Canada. It has been shown that key hydrological-cycle processes (such as evapotranspiration, temperature, and precipitation) exhibit the so-called long-range dependence that also affects the fractal short-range dependence [35], and thus, prediction. The novelty of this study is four-fold. (1) Introducing the IELM to overcome the main constraint of the ELM in random generation of the bias of hidden neurons and input weights matrices: In this model, iteration parameters and calculation of orthogonal random generation matrices are considered to find the most reliable results in terms of simplicity and accuracy simultaneously. The main advantages of this model are high generalization capability and fast training samples for a large number of iterations. (2) Coding a JavaScript-based code in the Google Earth Engine environment to extract SRFD data from the gridded Daymet product: Daymet supplies long-term and continuous estimates of daily weather and climatology factors generated using ground-based observations through statistical modeling techniques. (3) Time series-based modeling of the SRFD is used without requiring other independent variables to develop a simple model. (4) Apractical matrix-based equation for practical applications is provided.

2. Materials and Methods

2.1. Case Study

SRFD is the main component of energy exchanges between the atmosphere, the Earth's surface, and the ocean. Therefore, it affects the context of temperature, atmospheric and oceanic circulation, and the hydrological cycle [1]. The selected case study is located in the northern part of Quebec in Canada. The Eastern Hudson Bay Basin is situated in the east of Hudson Bay (longitudes 76°40′ W–71°30′ W and latitudes 54°50′ N–57°10′ N), which is a part of the Nunavik area (Figure 1). Nunavik lies in both the subarctic and Arctic climate zones. Due to the climatic conditions of this area, access to the site is complex, and measuring hydrological variables in this area is not only easy but also costly. The basin's elevation is between zero and 594 m relative to the mean sea level, and its average elevation is 300 m. The most important natural phenomena of this basin are clearwater lakes composed of two separate lakes that occupy the middle part of the basin. The deepest part of these lakes is 178 m. Annual precipitation in this basin is between 600 to 852 mm with an average of 726 mm.

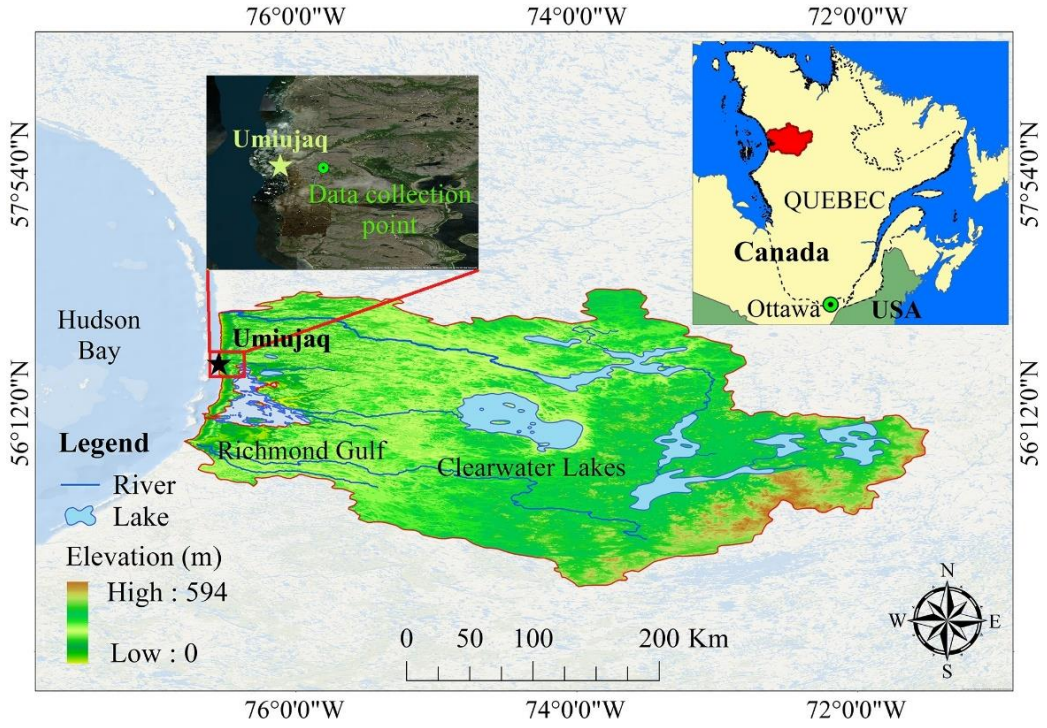

Figure 1. The geographic location of the study area.

Based on land cover data (MOD12Q1, 500 m × 500 m) obtained from NASA Land Processes Distributed Active Archive Center, open shrubland mostly covers the northern parts of the area with a height of one to two meters, the southern parts mostly include savannas areas with 10% to 30% tree cover, and most of the cover is grasslands in coastal regions (Figure 2).

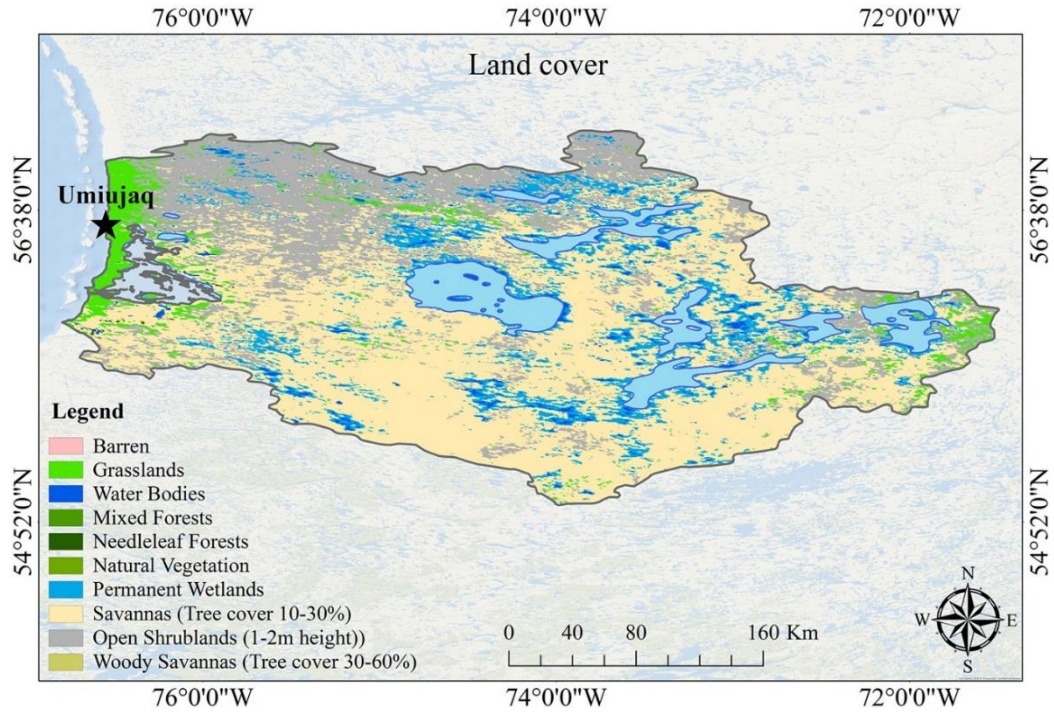

Figure 2. Land cover of the study area.

Solar energy is a clean source that does not introduce any pollutants to the environment, and it may even overcome the problems caused by fossil fuels. It should be noted that this is mostly possible if the solar energy is combined with a water-energy nexus battery (e.g., hydroelectric dam) to be effective and sustainable as an energy source. In this regard, its efficient use requires practical and effective knowledge [36]. SRFD data provide information about the impact of solar energy on the Earth in a given area over a period of time. It should be noted that due to the high cost of observing and measuring, this parameter is not easily accessible. Therefore, there is a need to create an alternative for estimating this data as well as predicting it [37,38]. Over the past decade, SRFD measurements as well as sunshine have declined in western Canada and elsewhere. Since the amount of SRFD is critical for calculating the evapotranspiration, soil melt, snowmelt radiation, and other hydrological cycle components, it is necessary to provide practical and accurate methods to estimate it [39]. Because a large volume of the study area is composed of different vegetation and water, solar radiation has an influential role in the area's environment. Changes in solar radiation in each region may affect different parameters such as plant growth period, photosynthesis, and melting outside the natural time of ice and snow (especially important due to the cold region) [40,41].

Since the study area is one of the cold regions and its period of frost and snowfall is high, the management of melting snow and ice is necessary for this region. Therefore, comprehension of snow and ice melting levels is essential for the proper management of available water resources, including short-term or seasonal flow forecasts, for hydrological studies and ice–snow mass balance. SRFD is an effective parameter in this regard [42,43].

2.2. Daymet V3 (Daily Surface Weather and Climatological Summaries)

Daymet V3 provides daily ground-level weather parameters in North America (Canada, Mexico, USA, Hawaii, and Puerto Rico) (https://daymet.ornl.gov/, accessed on 31 March 2020). This data is available with a spatial resolution of 1000 m and a time interval of one day. Daymet includes data on daily minimum 2-meter air temperature (°C), daily maximum 2-meter air temperature (°C), precipitation, the partial pressure of water vapor, duration of the daylight period, snow water equivalent, and shortwave radiation flux [44].

Shortwave radiation flux density (SRFD) was extracted from the gridded Daymet product for the study area. These data were processed in the Google Earth Engine environment and provide a monthly time series to apply as input in subsequent analysis. Google Earth is free, it has up-to-date maps and data, it is available on a wide array of devices, it is incredibly detailed, and it is very user-friendly. This dataset covers the period from January 1980 to December 2020. To access the SRFD data provided in the Google Earth Engine by NASA, a JavaScript-based code was offered to extract data for the desired location in Nunavik.

Google Earth Engine is a cloud platform for global spatial data analysis that allows the processing of large amounts of data in various fields, including drought, natural disaster water management, deforestation, vegetation, agriculture, soil studies, climate monitoring, and conservation of the environment [45]. The Google Earth Engine provides convenient conditions for developing algorithms and receiving results quickly with easy access and a user-friendly environment. It improves accessibility and usability by offering Earth observation data to a wide range of research fields. Additionally, the Google Earth Engine cloud platform also provides access to relevant data and scripts for users who do not have the necessary data or computing tools [46,47].

For the current study, 468 SRFD monthly data between 1981–2019 were used to model SRFD, so that 336 samples (from January 1982 to December 2008) were considered for the training phase, while the other 132 samples (from January 2009 to December 2019) were applied to check the performance of the calibrated model. The different values of the SRFD for both phases are provided in Figure 3. Additionally, the statistical characteristics of the total, training, and testing data are presented in Table 1.

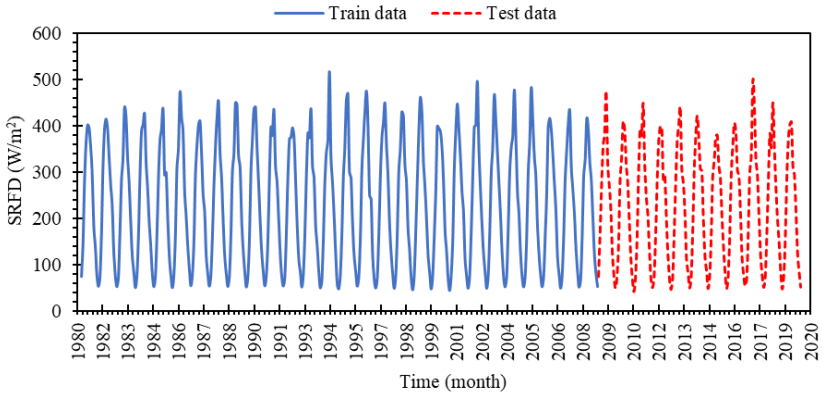

Figure 3. SRFD time series plot for test and training data.

Table 1. Statistical attributes of SRFD data.

	Nbr.	Min.	Max.	1st Q.	Median	3rd Q.	Mean	σ(n)	γ_1	γ_2
Total	468	41.89	516.18	106.77	243.38	350.39	231.67	131.50	0.11	−1.31
Train	336	43.41	516.18	106.78	245.02	356.37	232.74	131.06	0.08	−1.38
Test	132	41.89	497.03	105.92	232.70	329.67	227.24	128.67	0.08	−1.28

Nbr.: Number of data, Min. and Max.: minimum and maximum of data, 1st Q. and 3rd Q.: first and third quartiles, σ(n): standard deviation, γ1: skewness, γ2: kurtosis.

2.3. Improved Extreme Learning Machine (IELM)

The backpropagation is a well-established training algorithm for the feed-forward neural network to solve many nonlinear complex problems leading to acceptable results [48]. However, similar to any machine learning-based approach, the BP has some drawbacks with its generalization and implementation, including local minima, learning rate, overfitting, low generalization ability [49], and time-consuming training [50].

To overcome the abovementioned limitations, Huang et al. [30] presented the extreme learning machine (ELM) as a training algorithm for a single-layer feedforward neural network. Using ELM is so simple that no parameters need to be set other than defining the network architecture. Therefore, many of the complexities of tuning parameters in gradient algorithms are not present in this algorithm. Additionally, as the modeling speed in ELM is so high, most of the training takes a short time with a large amount of data, which is not easy to model with the classical neural network; it takes about a few minutes. Therefore, the most important advantages of this method are the least user intervention, the learning process needing a single iteration, robustness, the fast learning rate, avoiding local minimizations, high generalization, and high accuracy [51].

The modeling process in ELM consists of three main stages: (1) random determination of input weights and bias matrices of hidden neurons, (2) calculation of the hidden layer outputs matrix using randomly generated matrices and activation function, and (3) calculation of output weights through a linear process. Accordingly, among three different matrices that need to be quantified, only the output weights are calculated analytically, and the other two matrices are randomly set. The least-square solution of a linear system is considered to define the output weights matrix.

To model a problem, it is first necessary to specify the inputs and outputs of the problem. For the input matrix with d independent inputs and the output of the problem with m, the outputs are $InV_i \in R^{DM}$ and $Ta_i \in R^z$, respectively. The number of S training samples are defined as $\{(InV_i, Ta_i)\}_{i=1}^{S}$. By considering the f(·) as activation function and NHN as number of hidden neurons, the output of the feedforward neural network is considered as follows:

$$O_i = \sum_{j=1}^{NHN} \beta_j f(a_j \times InV_i + b_j,), \qquad i = 1, 2 \ldots, S \tag{1}$$

where NHN is the number of hidden neurons, S is the number of training samples, $\beta_j \in R^z$ is the output weight matrix that connects the jth hidden node to the corresponding hidden node, z is the number of the output variable, and O_i and InV_i are the output and input variables, respectively. $b_j \in R$ is the bias of the jth hidden neuron, and $a_j \in R^{DM}$ is the input weights that link the input variables to the jth hidden node and $a_j \cdot InV_i$ denotes the inner product of the a_j and InV_i. The mathematical form of the different activations applied in the current study are as follows:

Sine

$$f(OutW, b, InV) = \sin(OutW \times InV + b) \tag{2}$$

Radial Basis Function

$$f(OutW, b, InV) = \exp(-(OutW \times InV + b)^2) \tag{3}$$

Triangular Basis Function

$$f(OutW, b, InV) = \begin{cases} 1 - |OutW \times InV + b| & OutW \times InV + b \geq 0 \\ 0 & otherwise \end{cases} \tag{4}$$

Hardlimit

$$f(OutW, b, InW) = \begin{cases} 1 & If\, OutW \times InV + b \geq 0 \\ 0 & otherwise \end{cases} \tag{5}$$

Sigmoid

$$f(OutW, b, InV) = \frac{1}{1 + \exp(-(OutW \times InV + b))} \tag{6}$$

Tangent hyperbolic

$$f(OutW, b, InV) = \tanh(OutW, b, InV) = \frac{\exp(2(OutW \times InV + b)) - 1}{\exp(2(OutW \times InV + b)) + 1} \tag{7}$$

The matrix-based form of Equation (1) with N separated equations is defined as follows:

$$T\beta = O \tag{8}$$

where T, β, and O ($O = [O_1, \ldots, O_N]^T$) denote hidden neurons' output, output weight, and output, respectively. The hidden neurons output matrix (T) is defined as follows:

$$\begin{aligned} &T(OuTW_1, \ldots, OuTW_L, INV_1, \ldots, INV_N, b_1, \ldots, b_L,) \\ &= \begin{bmatrix} f(OuTW_1 \times INV_1 + b_1) & L & f(OuTW_{NHN} \times INV_1 + b_{NHN}) \\ M & O & M \\ f(OuTW_1 \times INV_{DM} + b_1) & L & f(OuTW_{NHN} \times INV_{DM} + b_{NHN}) \end{bmatrix}_{DM \times NHN} \end{aligned} \tag{9}$$

The two matrices T and β in Equation (8) are unknown. The H is calculated using input weights and the bias of hidden neuron matrices that both of them are randomly assigned. Therefore, T is calculated without the experience of training samples. Consequently, the output weights matrix (β) is the only unknown matrix that should be calculated through the training phase. To find this matrix, Equation (8) as a linear system should be solved.

The dimension of the T is DM × NHN. The NHN is generally higher than the input variables (DM), and therefore, the H is not a square matrix. Thus, finding the output weights matrix using Equation (8) is not simple [30]. To find the output weights matrix, the optimal least square solution of the output weights matrix by loss function minimization is calculated as follows:

$$E_{ELM} = MIN\|O - T\beta\| \tag{10}$$

where the output weights matrix calculated through the least square solution of the above equation is calculated as:

$$\hat{\beta} = T^+ y \tag{11}$$

where T^{+} denotes the Moore–Penrose generalized inverse of the T [52]. The solution of Equation (10) for NHN < S is as follows:

$$\hat{\beta} = (T^{T}T)^{-1}T^{T}O \tag{12}$$

The dimensions of the bias of hidden neurons, input weights, and output weights matrices are "1 × NHN", "DM × NHN", and "1 × NHN", respectively. Therefore, the number of all tuned parameters through the training phase of the ELM is as follows:

$$k = NHN + DM \times NHN + NHN = NHN(DM + 2) \tag{13}$$

The ratio of the randomly generated parameters that are related to the input weights (i.e., DM × NHN) and bias of hidden neurons (i.e., 1 × NHN) to the output weight (i.e., 1 × NHN) is as follows:

$$R = \frac{NHN + DM \times NHN}{NHN(DM + 2)} = \frac{NHN(d + 1)}{NHN(d + 2)} = 1 - \frac{1}{DM + 2} \tag{14}$$

If the number of input variables is one, the R is more than 0.66. Therefore, it is observed that two-thirds of the parameters tuned in the modeling using ELM are randomly determined, which has a significant impact on the modeling results. An inaccurate value of these parameters may reduce the generalizability of the developed model. Therefore, in this study, two competencies are performed on the original ELM: (1) considering iteration parameter for ELM and (2) applying the orthonormal basis for the range of input weights and bias of hidden neuron matrices. Using these two competencies, the new version of the ELM is named improved ELM (IELM). The flowchart of the developed IELM is presented in Figure 4.

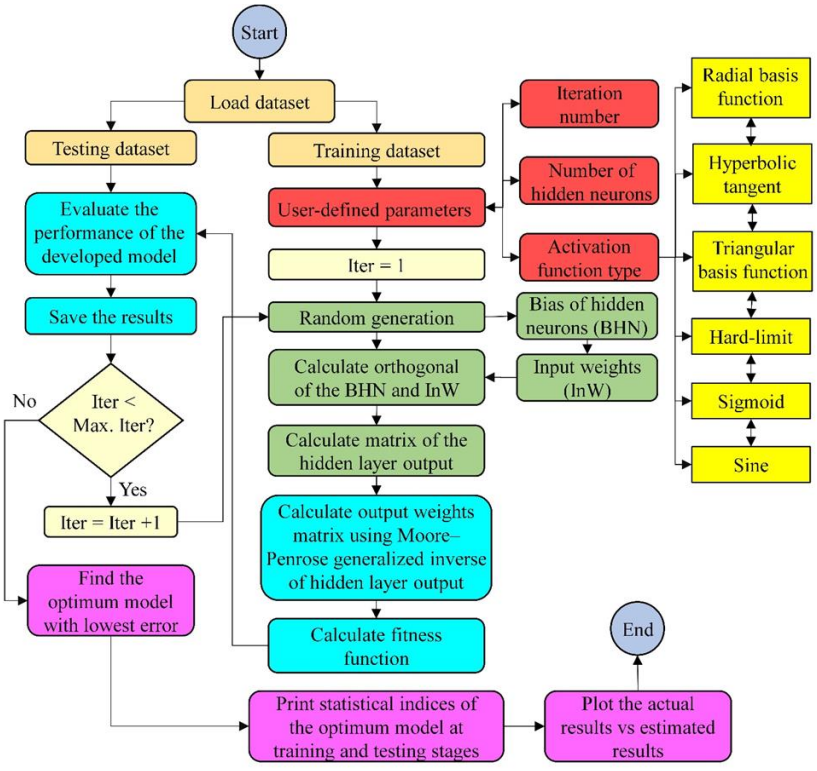

Figure 4. The flowchart of the developed IELM.

2.4. Uncertainty Analysis

The quantitative appraisal of the uncertainty analysis (UA) in the estimation of the shortwave radiation flux density (SRFD) has employed the developed improved extreme learning machine (IELM) instead of the original ELM. For a fair comparison of the different IELM-based models, the UA is utilized to the test data [53,54]. The use of test data in calculating and checking the UA of the developed IELM model has the advantage that its performance is examined for testing data without any role in model training. The generalizability of this model can be confirmed. The first step in UA is the calculation of the individual estimation error (IEE) as follows:

$$E_i = SRFD_{E,i} - SRFD_{O,i} \quad (15)$$

where $SRFD_{E,i}$, and $SRFD_{O,i}$ are ith estimated and observed SRFD, respectively, and E_i is the IEE of the ith sample. The IEE for all samples are applied to calculate mean estimation error (MEE) as follows:

$$\overline{E} = \sum_{i=1}^{S} E_i \quad (16)$$

where $\overline{E}$ is the MEE and S is the number of samples. Using the calculated $\overline{E}$ and IEE for all samples (E_i), the standard deviation of the estimation error (SDEE) is calculated as follows:

$$SDEE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(E_i - \overline{E})} \quad (17)$$

The negative (or positive) value of the MEE demonstrates that the estimator model underestimated (or overestimated) the observed values of the SRFD. To approximate a confidence band around the estimated values of an error, the ± 1.96 SDEE is calculated.

2.5. Workflow Approach

In this section, the workflow approach of the current study is presented. This workflow comprises four main steps: data collecting, model definition, tuning IELM parameters, and performance evaluation. Performing all four steps will lead to achieving the optimal model in SRFD estimation.

The first step is collecting data. To collect data, SRFD is extracted from the Daymet dataset by developing a JavaScript-based code in Google Engine Cloud. Using the developed code, the daily SRFD dataset from January 1981 to December 2019 is selected. The second step is the definition of the inputs to apply developed IELM-based MATLAB code to SRFD estimation. To do that, the auto-correlation function is employed [55]. Using this function, the most effective lags of the SRFD are found and defined as different combinations of these lags to find the best model. The third step is the definition of IELM parameters for all input combinations, defined in the previous step. To define IELM parameters, the type of activation function, the number of hidden neurons, and iteration number must be pre-defined by the user. It should be noted that the maximum number of hidden neurons should be considered. As the maximum allowable value of the hidden neurons should be considered, the number of optimal tuning parameters through the training phase will be less than the training samples. According to that, the number of columns and rows in the input weighs is identical to the number of input variables (InV) and the number of hidden neurons (NHN), respectively. The bias of hidden neurons (BHN) and output weights are two matrices with one column in which the number of rows is equal to NHN. The maximum allowable NHN is calculated as follows:

$$\text{Max. NHN} < \frac{TrSa}{InV + 2} \quad (18)$$

where TrSa is the number of training samples.

Besides the maximum allowable NHN, the activation function type and iteration numbers also must be pre-defined. The iteration number is considered in the range of 100 to 100,000. At the same time, the six different activation functions, including hyperbolic tangent, Sigmoid, hard limit, Sin, radial basis function, and triangular basis function, are investigated to find the optimum one. The final step of the IELM modeling is the performance evaluation of the developed models to find the optimum one. In this step, different statistical indices and uncertainty analysis are employed to find the optimum one. The schematic workflow for shortwave radiation flux density modeling is provided in Figure 5.

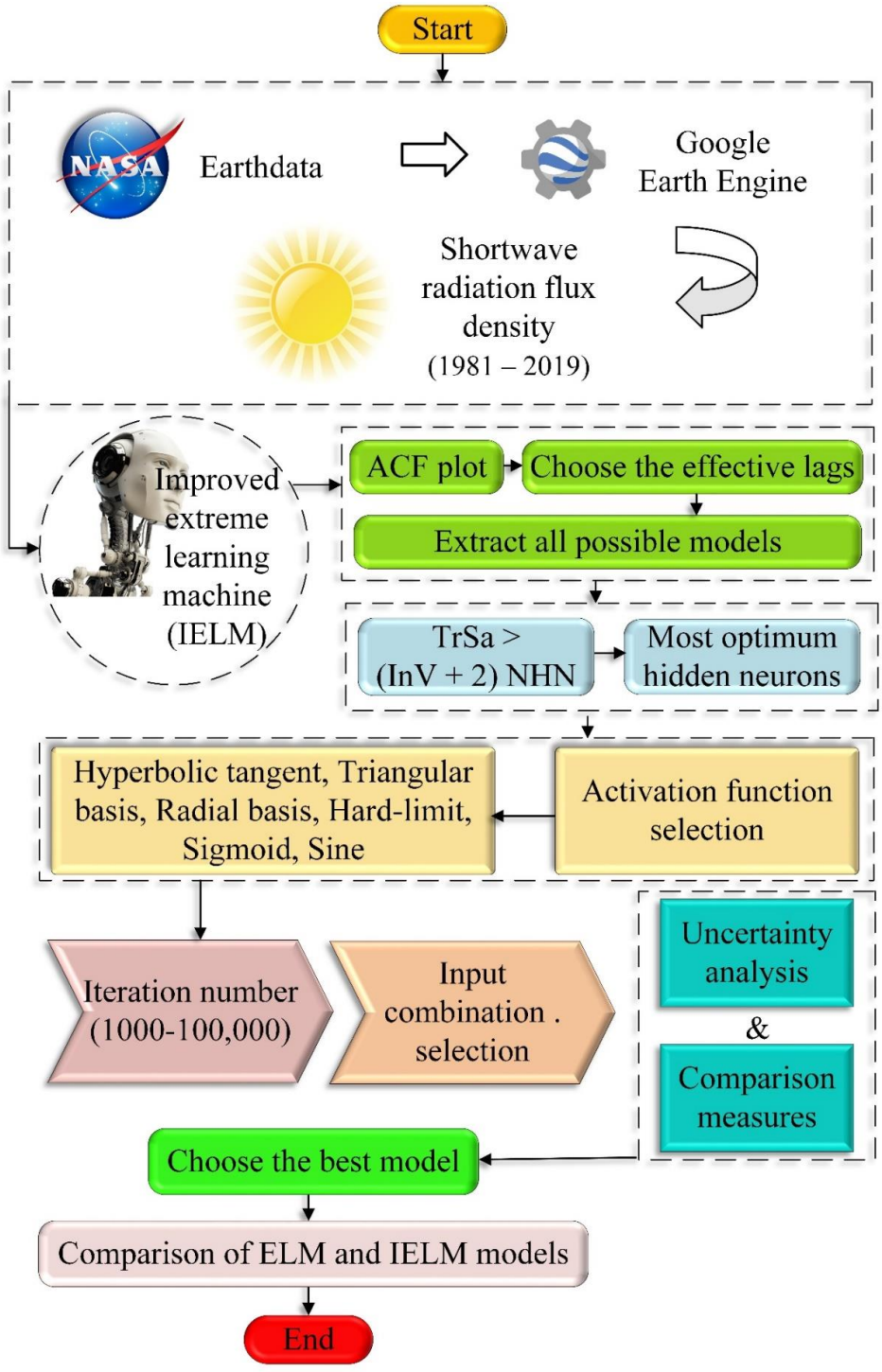

Figure 5. Schematic workflow for shortwave radiation flux density modeling.

2.6. Comparison Measures

In this section, four different statistical indices, including correlation coefficient (R), root mean squared error (RMSE), mean absolute percentage error (MAPE), and Nash–Sutcliffe model efficiency coefficient (NSE), are used to assess the performance of the developed IELM model in SRFD estimation. The RMSE index is of considerable importance in studying environmental and climatic parameters and is widely used in these fields [56]. However, this index is not practical enough to check the average performance of a model, which may be a misleading index of the average error [57]. Scale dependency is one of the main drawbacks of RSME, allowing outliers to have a significant impact on the obtained results of this index, which will make it difficult to evaluate the model, and any fraction of the data will cause fundamental changes in the results [58]. Therefore, it is necessary to use other indices along with it. Scholars often employ MAPE because of its intuitive understanding in terms of relative error [59]. In addition to these two relative (MAPE) and absolute (RMSE) indices, two correlation-based indices (i.e., R and NSE) are also used. According to the characteristics of each indicator, their simultaneous application can be an excellent approach to assess the efficiency of the developed IELM-based models in SRFD estimation. The mathematical definition of the R, RMSE, MAPE, and NSE is defined as follows:

$$R = \left(\frac{\sum_{i=1}^{S} \left(SRFD_{O,i} - \overline{SRFD}_O\right)\left(SRFD_{E,i} - \overline{SRFD}_E\right)}{\sqrt{\sum_{i=1}^{S} \left(SRFD_{O,i} - \overline{SRfD}_O\right)^2 \sum_{i=1}^{S} \left(SRFD_{E,i} - \overline{SRFD}_E\right)}} \right) \tag{19}$$

$$RMSE = \frac{1}{S}\sqrt{\sum_{i=1}^{S} \left(SRFD_{O,i} - SRFD_{E,i}\right)^2} \tag{20}$$

$$MAPE = \frac{100}{S}\sum_{i=1}^{S}\left(\left|\frac{SRFD_{O,i} - SRFD_{E,i}}{SRFD_{O,i}}\right|\right) \tag{21}$$

$$NSE = 1 - \frac{\sum_{i=1}^{S}\left(SRFD_{O,i} - SRFD_{E,i}\right)^2}{\sum_{i=1}^{S}\left(SRFD_{O,i} - \overline{SRFD}_O\right)} \tag{22}$$

where $SRFD_{O,i}$ and $SRFD_{E,i}$ are the observed and estimated values of the ith samples of the SRFD, respectively, N is the number of samples, and $\overline{SRFD}_O$ and $\overline{SRFD}_E$ are the mean of the observed and estimated SRFD, respectively.

3. Results and Discussion

3.1. SRFD Modeling

Before starting the modeling, the input parameters need to be determined. In the current study, historical values of the SRFD are used to estimate this parameter at future times. Indeed, the time series concept is used to solve the problem. Hence, effective lags are determined using the auto-correlation function. It indicates the correlations between the past and future values of the desired parameters (i.e., SRFD in the current study). The auto-correlation function of the SRFD time series was shown the most critical lags are 1 and 2. Additionally, Lag 12 is also evident as the periodic term. In addition to these three lags and to find a more reliable model, Lag 3 is also considered as one of the input parameters in the IELM-based modeling:

$$SRFD(t) = f(SRFD(t-1), SRFD(t-2), SRFD(t-3), SRFD(t-12)) \tag{23}$$

According to the above equation, the number of fifteen input combinations with one to four input variables is defined as provided in Figure 6.

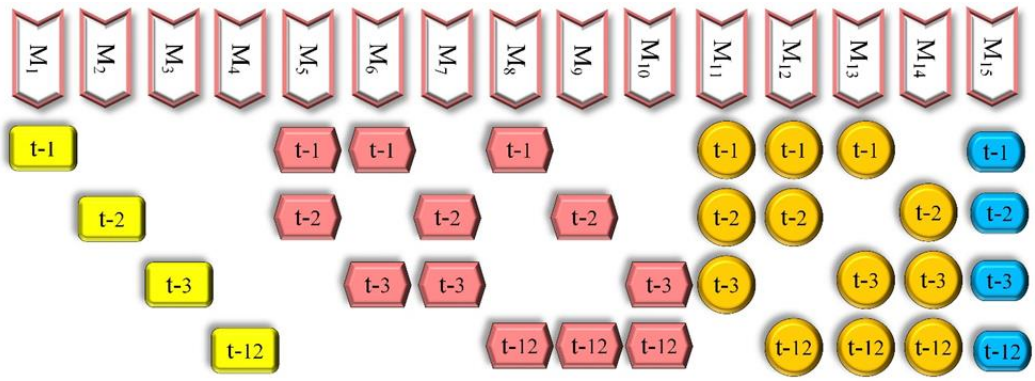

Figure 6. Input combinations of historical data defined for SRDF modeling.

3.2. Most Optimum Hidden Neurons

After defining different input combinations, finding the optimal number of neurons for SRFD estimation is required. For this purpose, Model 15 with SRFD (t − 1), SRFD (t − 2), SRFD (t − 3), and SRFD (t − 12) as input parameters (Figure 6) is employed. Additionally, by keeping other tuning parameters constant, including the activation function (i.e., Sigmoid) and iteration number (10,000 iterations), the optimal number of neurons is found. It should be noted that this number can be changed in the range of 1 to Max. NHN (Equation (18)). The reason for limiting the number of hidden layer neurons to Max is that increasing NHN could result in higher generalizability of the model so that if this limitation is not taken into account, the generalizability of this model may be doubtful.

The total number of data is 468, 70% of which have been selected as training samples, and 30% of the samples were considered testing samples. Taking into account the delays created in the modeling process, the value of TrSa is 336 (this number can be varied by changing the inputs). Given that the modeling process was initially performed using all lags (Model 15), the value of InV is equal to four. Consequently, the Max. NHN is calculated as 55 (Equation (18)).

The statistical indices of the developed IELM with different hidden neurons are provided in Figure 7. The minimum values of NSE and R are less than 0.6 and 0.8, respectively. As the number of the hidden layer neurons increases, the value of these two indices increases significantly. In NHN > 10, the value of both indices is more than 0.9, which is an acceptable value. Although the growth of the value of these two indices is also observed in most models with NHN > 10, the growth rate compared to NHN < 10 has a significant decrease. The upward trend presented in the correlation-based indices (i.e., R and NSE) is also observed as a similar downward trend in RMSE and MAPE (%). A significant point is a sharp decrease in the value of R and NSE as well as the increase in the RMSE and MAPE (%) at NHN = 31, which is significantly different from the values of its neighbors. One of the reasons for the performance of Model 31 may be the number of iterations, so for this model, the number of more iterations was examined, but there was no effective change in the performance of the model in the testing stage. Another reason can be the lack of accurate assignment of the input weights and bias of hidden neuron parameters, which has led to a significant reduction in the generalizability of this model. The results of this figure show that the best performance is obtained at NHN = 27 (R = 0.98; NSE = 0.96, RMSE = 25.02; MAPE (%) = 10.64). Therefore, the number of hidden neurons in the following modeling process is considered to be 27.

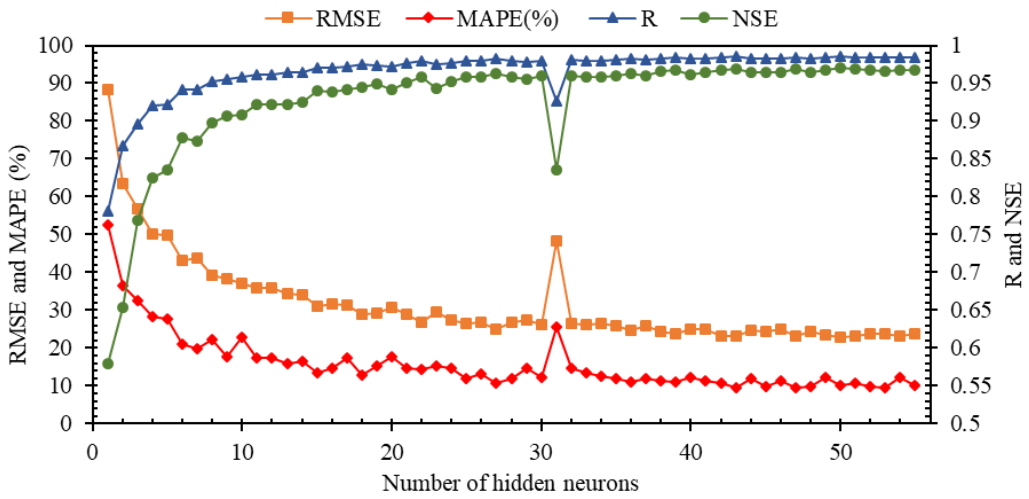

Figure 7. Statistical indices of the developed IELM with different hidden neurons.

3.3. Activation Function Selection

After determining the optimum number of hidden neurons, six activation functions, including hyperbolic tangent, sigmoid, radial basis function, triangular basis, hard limit, and sine, are evaluated in this section. The statistical indices of the IELM with defined input combinations in Model 15 and different activation functions are provided in Figure 8. As shown in this figure, although hyperbolic tangent, hard-limit, and sigmoid activation functions have close results, the RMSE value for sigmoid is improved by 12.2% and 4.2% compared to hard-limit and hyperbolic tangent, respectively, which indicates sigmoid function performs better than the other two functions. The RMSE for the other three functions (i.e., sine, radial basis, and triangular basis) is about ten times the value of this index for sigmoid. Moreover, the value of MAPE error index for sigmoid is 10.64% and for hard limit and hyperbolic tangent functions are equal to 12.44% and 11%, respectively. The NSE index shows the same results for the three mentioned activation functions. Additionally, the correlation coefficient value indicates the equality of this index in the hyperbolic tangent and sigmoid functions with a value of 0.98, while this value for the hard limit is 0.97. Therefore, it can be concluded that sigmoid has shown better performance compared to the other activation functions.

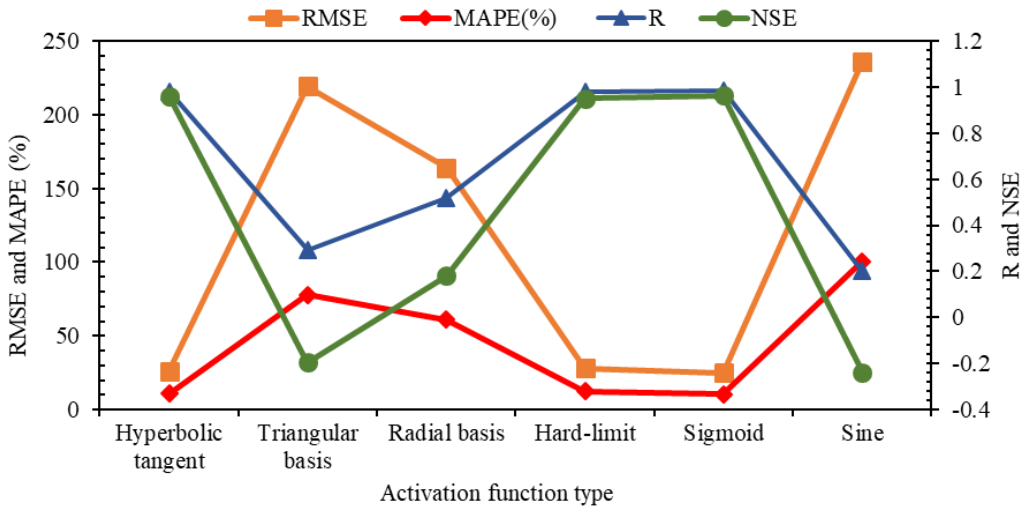

Figure 8. Statistical indices of the developed IELM with different activation functions.

3.4. Iteration Number

The third parameter is used to determine the iteration number. Many iterations in IELM modeling are used to overcome the problems caused by randomly determining the two matrices of bias of hidden neurons and input weights, which include at least 66% of the total parameters optimized during the training process (Equation (14)).

In the next step, the effect of iteration number in SRFD modeling was investigated. To investigate the impact of changes in the iteration number, a boxplot diagram was used to obtain the best number of iterations in the modeling process. In this study, 15 different values were used for the iteration number in the range of 10,000–100,000, similar to Table 2. The distribution of RMSE values for various iterations is shown in Figure 9. According to the high value of RMSE for some models, the maximum value of RMSE in this figure was plotted in two ranges, $[0, 2\times10^{7}]$ and [0, 160]. Additionally, the boxplot parameters, including minimum, maximum, median, first quartile (Q1), and third quartile (Q3), are provided in Table 2. The difference between Q1 and Q3 is known as IQR. Using IQR, the minimum and maximum values are calculated as Q1 − 1.5 × IQR and Q3 + 1.5 × IQR, respectively. Values smaller than the minimum and larger than the maximum are known as outliers. Indeed, if a number outside this range is recorded, it indicates that the amount of model error that is considered as RMSE in some cases might be very small or very large. Figure 9 shows that at iteration number = 30,000, the RMSE value also reaches $2\times10\text{-}^{7}$, a high value for this index. By limiting the error range to [0, 160], it can be seen that in all the values defined for the iteration number, a very large number of iterations record an error value greater than the maximum value (=Q3 + 1.5 Q IQR). Except for iteration numbers = 10,000 and 30,000, in other cases, the changes' range minimum, maximum, Q1, and Q3 are almost constant.

Table 2. Boxplot parameters for 15 iteration numbers.

No.	It. Number	Max	Min	Median	Q1	Q3
1	1000	50,000	29.93	47.46	42.15	54.57
2	2000	51,000	30.57	47.90	42.71	54.75
3	3000	1,201,000	28.47	47.83	42.69	55.16
4	5000	619,000	29.16	47.68	42.37	54.86
5	10,000	1000	21.11	62.40	47.64	81.13
6	15,000	619,000	29.16	48.02	42.66	55.16
7	20,000	498,000	24.83	47.84	42.60	54.94
8	30,000	20,386,000	44.18	99.93	82.81	120.11
9	40,000	435,000	24.47	47.71	42.57	54.79
10	50,000	1,019,000	27.10	47.72	42.53	54.83
11	60,000	997,000	26.75	47.79	42.59	55.00
12	70,000	1,714,000	25.68	47.66	42.48	54.81
13	80,000	2,108,000	25.91	47.79	42.59	54.99
14	90,000	1,327,000	26.21	47.72	42.48	54.80
15	100,000	2,506,000	26.44	47.76	42.57	54.92

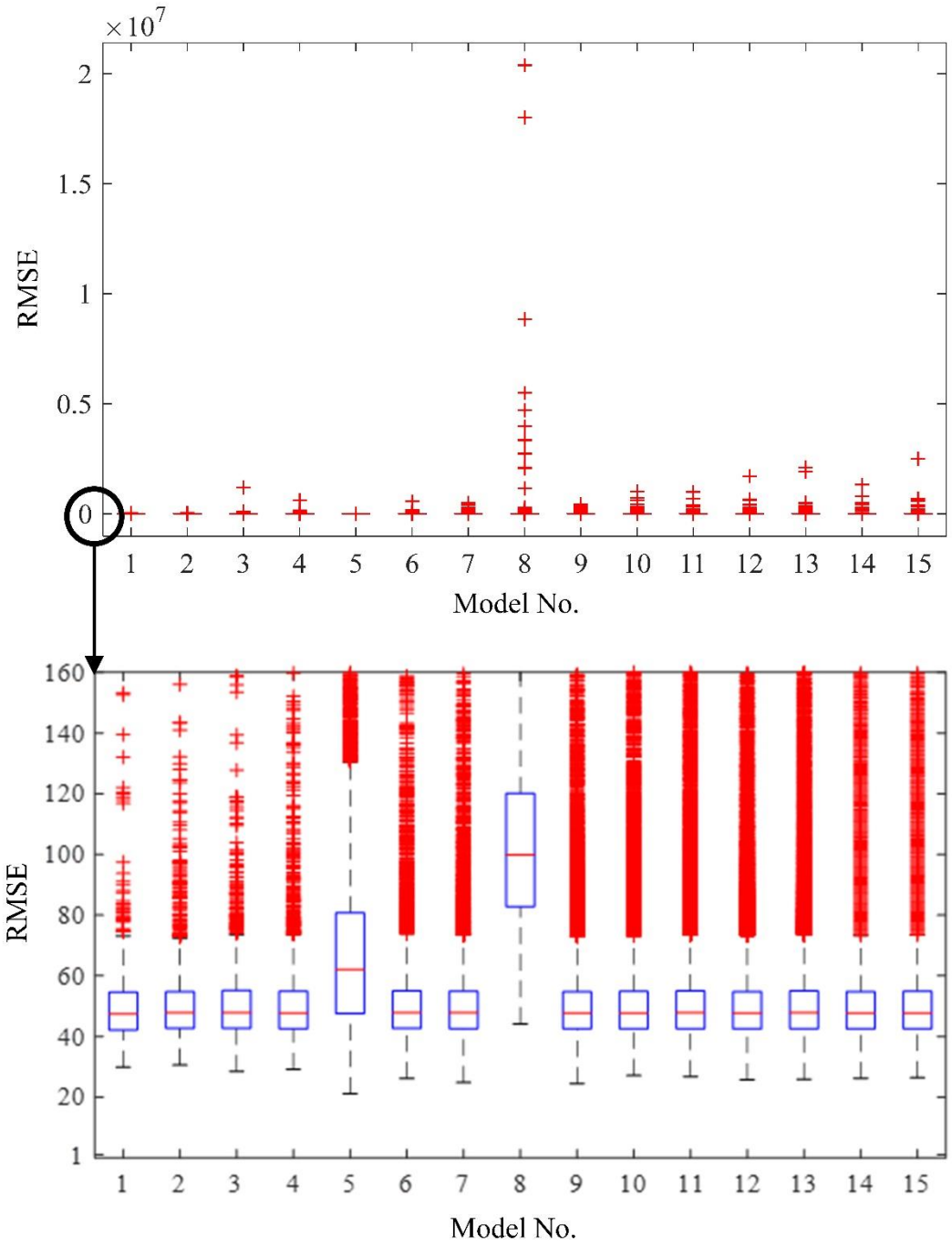

Figure 9. The distribution of the RMSE for different iteration numbers in SRFD modeling by IELM.

In Model 5 (iteration number = 10,000), the median is 62.4, which indicates that 50% of the data model from 10,000 runs is in the range of 47.64 to 81.13, while the lowest error recorded in this case is equal to 21.11. In all other models with different iteration numbers, the minimum error value is greater than 21.11, which indicates the better performance of Model 5 (Figure 9).

Figure 10 signifies the statistical indices of the best IELM-based model with the different iteration numbers. According to this figure, the lowest MAPE (%) was recorded for the iteration number = 10,000 (MAPE (%) = 10.64). The correlation-based indices (i.e., R and NSE) for all iteration numbers except 30,000 are very close together. The lowest and highest values of R are 0.973 and 0.982, respectively, and the lowest and highest of the NSE are 0.945 and 0.964, respectively. The lowest RMSE is related to the 20,000 iteration

(RMSE = 24.83). Considering the values of different indices, the iteration number = 10,000 is selected as the optimum value of the iteration number (R = 0.98, NSE = 0.96, RMSE = 25.02, and MAPE (%) = 10.64).

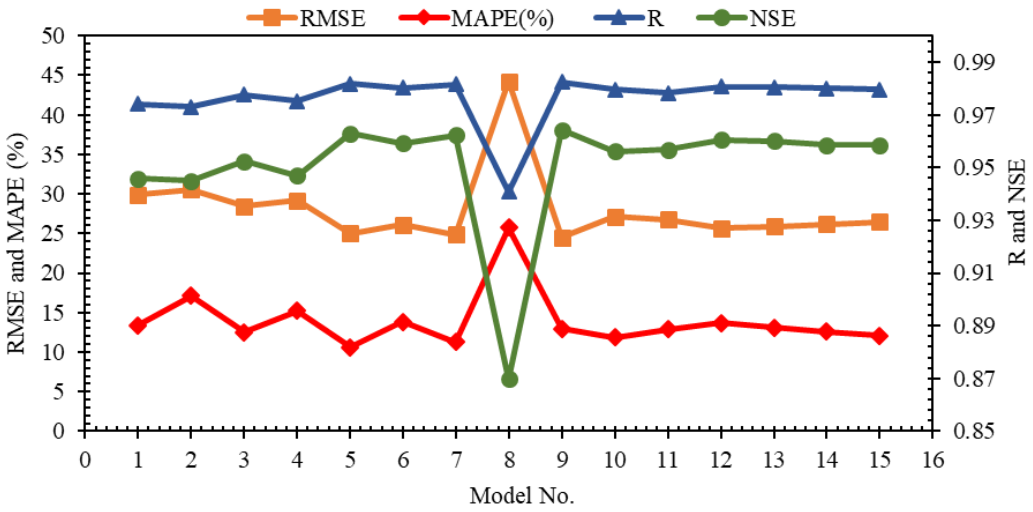

Figure 10. Statistical indices of the best IELM-based model with different iteration numbers.

3.5. Input Combination Selection

After determining the modeling parameters in IELM, the effect of each input is examined by considering 15 input combinations, defined in Figure 6. The presented models in this figure are divided into four general categories: models with one, two, three, and four inputs, which include four, five, four, and one input combinations, respectively.

Models 1 to 4 are single-input models that include Lag 1, 2, 3, and 12. Among the models presented with a single input, it is observed that the highest value of correlation-based indices (i.e., R and NSE) is related to Model 4 (R = 0.983, NSE = 0.966) so that the closest value of these two indices to Model 4 corresponds to Model 1 (R = 0.892; NSE = 0.756). The results of these two indices in Lag 2 and Lag 3 are significantly different from the other two lags because in Lag 3, the NSE index has a negative value (NSE = -0.691). The value of this index in Lag 2 (NSE = 0.106) is approximately 10% of the value of this index in Lag 12. In addition to these two indices, Model 4 (i.e., Lag 12) also offers the best performance in RMSE and MAPE indices, so that the modeling relative error is less than 9% (MAPE (%) = 8.88), while the value of this index for Models 2 and 3, whose inputs are Lag 2 (MAPE (%) = 88.23) and Lag 3 (MAPE (%) = 90.28), is about ten times the value of this index in Model 4. According to the explanations provided, Lag 12 has the most effectiveness, and Lag 1 is in the second place for one-step-ahead modeling of SRFD by considering only one lag.

Although Lag 2 had the weakest performance among all lags, its combination with Lag 1 (Model 5) (R = 0.956; RMSE = 37.645; MAPE (%) = 20.038; 0.911) increased the performance of Model 1, which uses only Lag 1 so that the values of R, RMSE, MAPE, and NSE indices have increased by 7.24%, 34.42%, 55.88%, and 20.47%, respectively, compared to Model 1.

The combination of Lag 3 with Lag 1 has also led to Model 6 (R = 0.975; RMSE = 28.833; MAPE (%) = 13.352; 0.948), which is more accurate than Model 1. According to the results presented for single-input models, the weakest performance was observed for Lag 3. Therefore, it was expected that using Lag 2 in combination with Lag 1 would provide better performance than combining Lag 3 with Lag 1, but Model 6, whose inputs are Lag 1 and Lag 3, compared to Model 5, whose inputs are Lag 1 and Lag 2, provided better performance. Therefore, it is concluded that to achieve the optimal model, it is necessary to examine the synergy of different parameters with each other.

Considering that in single-input models, the best performance was obtained for Lag 12, which was significantly different from other lags, it is expected that the combination of this lag with one of the other lags (i.e., Lags 1, 2, and 3 as Models 8 to 10, respectively) leads to a more accurate result compared to Models 5 to 7. The results presented in Figure 11 show that the performance of Model 4, whose only input was Lag 12, has been slightly improved by Models 8 to 10 (two-input models). In Models 8 to 12, correlation based-indices (i.e., R and NSE) and RMSE experienced an increase in accuracy compared to Model 4, which has only one input (i.e., Lag 12). Still, for MAPE, this trend is not incremental in all three models. In Models 8 and 10, which use Lag 1 and Lag 3 (respectively) as input in addition to Lag 12, the MAPE is increased by about 2.5%, while in Model 9, which uses Lags 2 and 12 to estimate SRFD, the value of this index has decreased. Therefore, it is concluded that the simultaneous combination of Lag 12 and Lag 2 offers the best performance between models with 1 to 2 inputs. As the number of model inputs increases to 3 and 4 (Models 11 to 15), it is observed that in all models, correlation-based indices decrease, and both RMSE and MAPE indices are increased. Therefore, it is concluded that Model 9 is the best input combination for SRFD estimation.

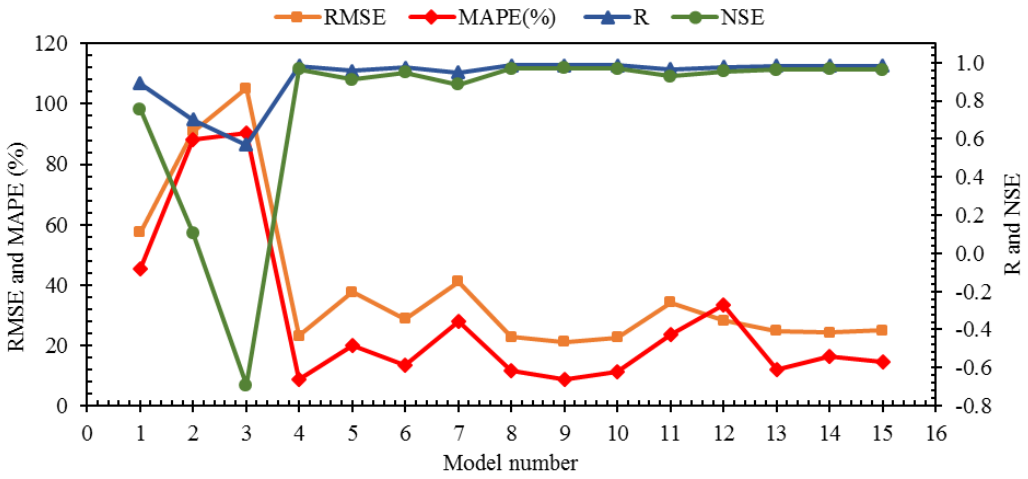

Figure 11. Error and correlation coefficient values for different input combinations.

3.6. Comparison of the IELM with ELM

Figure 12 illustrates the scatter plot of the observed and predicted SRFD by ELM and IELM. According to this figure, most of the estimated samples by IELM are in the range of ±10%, while the results provided by ELM in different SRFD ranges have many errors. Most of the estimated samples by the ELM in the range of 180–350 are fixed so that in this range, the estimated points do not follow the 45-degree line and are presented almost in a horizontal line.

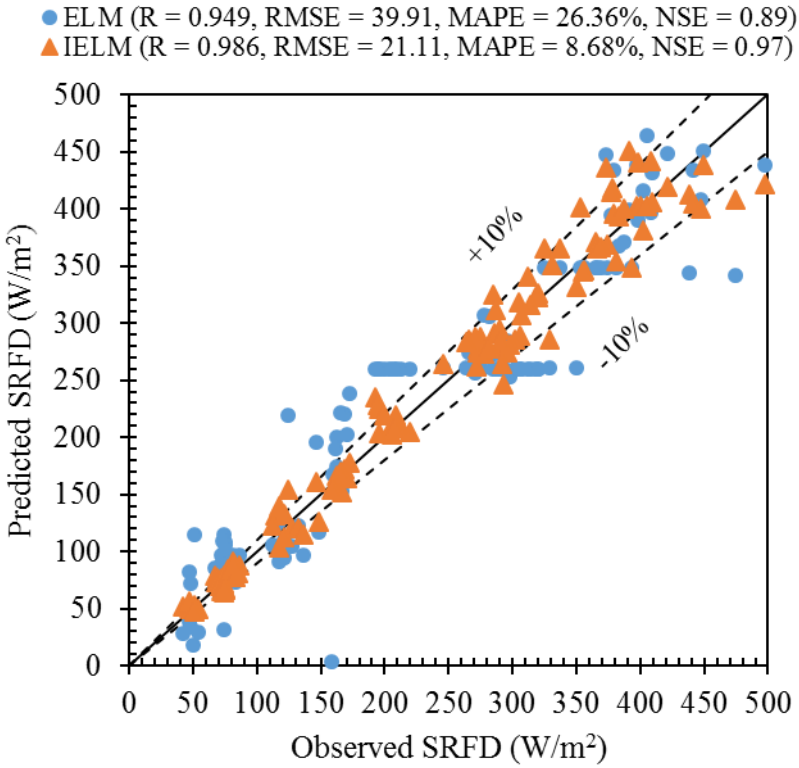

Figure 12. Scatter plot of the observed and predicted SRFD by ELM and IELM.

In both ELM and IELM methods, the errors are presented in both under and overestimate forms, and the only significant difference is the modeling error values. The indices presented in the figure show that correlation-based indices (i.e., R and NSE) in the IELM method have increased by about 3.9% and 9% compared to ELM. An increase of about 47% and more than 17% was observed in RMSE and MAPE, respectively. Therefore, it is determined that the developed method in the current study (i.e., IELM) has well overcome the limitations of the ELM method.

3.7. Uncertainty Analysis of the IELM Models Versus ELM

The uncertainty analysis (UA) results are summarized in Table 3. The mean estimation error (MEE), standard deviation of estimation error (SDEE), 95% estimation error interval (EEI), and width of uncertainty band (WUB) are provided in this table. According to this table, the MEE for $IELM_4$ and ELM are positive, while the value of this index for others is negative. Therefore, it is concluded that these two models over-estimated the SRFD, while the other models underestimated the SRFD. Given that the positive and negative values of the error are added together by considering their sign, the use of this model cannot be used as an index to check the accuracy of the model. Therefore, other indices need to be evaluated. The SDEE for all IELM models indicates that this index's lowest and highest values are associated with 21.12 and 105.47, related to the $IELM_9$ and $IELM_3$, respectively. The widths of uncertainty bands of the IELM are in the ranges of [±17.56, ±3.64] so that the lowest and highest ones are related to the $IELM_3$ and $IELM_9$, respectively. The widths of uncertainty bands for ELM is ±6.9, which is 52% higher than the best of the IELM model (WUB ($IELM_9$) = ±3.64).

Table 3. Uncertainty analysis of IELM versus ELM.

Model	MEE	SDEE	95% EEI	WUB
$IELM_1$	−2.40	57.56	(−11.91 · · · 7.11)	±9.51
$IELM_2$	−3.55	91.00	(−18.65 · · · 11.54)	±15.10
$IELM_3$	−4.86	105.47	(−22.42 · · · 12.7)	±17.56
$IELM_4$	0.07	23.24	(−3.93 · · · 4.07)	±4
$IELM_5$	−4.71	37.48	(−10.93 · · · 1.51)	±6.22
$IELM_6$	−4.58	28.57	(−9.34 · · · 0.17)	±4.76
$IELM_7$	−5.83	40.91	(−12.64 · · · 0.99)	±6.81
$IELM_8$	−2.73	22.81	(−6.66 · · · 1.19)	±3.93
$IELM_9$	−1.73	21.12	(−5.36 · · · 1.91)	±3.64
$IELM_{10}$	−3.01	22.43	(−6.87 · · · 0.84)	±3.86
$IELM_{11}$	−4.70	34.05	(−10.37 · · · 0.97)	±5.67
$IELM_{12}$	−1.63	28.37	(−6.52 · · · 3.25)	±4.89
$IELM_{13}$	−3.57	24.68	(−7.82 · · · 0.68)	±4.25
$IELM_{14}$	−4.36	23.97	(−8.49 · · · −0.23)	±4.13
$IELM_{15}$	−4.27	24.75	(−8.53 · · · −0.01)	±4.26
ELM	0.093	40.07	(−6.81 · · · 6.99)	±6.9

MEE = mean prediction error, SDEE = standard deviation of estimation error, 95% EEI = 95% estimation error interval, WUB = width of uncertainty band.

The universal form of the IELM is as follows:

$$\mathrm{SRFD(t)} = \left[\frac{1}{(1+\exp(\mathrm{InW}\times\mathrm{InV}+\mathrm{BHN}))}\right]^{\mathrm{T}}\times\mathrm{OuTW} \tag{24}$$

where the InW, InV, BHN, and OutW are the input weights, input variables, bias of hidden neurons, and output weights, respectively. The InW, InV, BHN, and OutW are defined as follows:

$$\mathrm{InV} = \begin{bmatrix} \mathrm{SRFD(t-2)} \\ \mathrm{SRFD(t-12)} \end{bmatrix}$$

$$\mathrm{BHN} = \begin{bmatrix} 0.0028 \\ 0.0798 \\ 0.0282 \\ -0.2781 \\ 0.0971 \\ 0.0618 \\ -0.2191 \\ -0.298 \\ -0.0738 \\ 0.2939 \\ -0.1102 \\ 0.061 \\ 0.1924 \\ -0.1088 \\ -0.1521 \\ -0.3477 \\ -0.2962 \\ -0.3353 \\ -0.2547 \\ -0.0322 \\ -0.2454 \\ -0.0749 \\ -0.1349 \\ -0.1087 \\ -0.2937 \\ -0.1236 \\ 0.0312 \end{bmatrix} \quad \mathrm{InW} = \begin{bmatrix} -0.3155 & 0.0897 \\ 0.1958 & -0.0233 \\ -0.0134 & 0.006 \\ 0.0698 & 0.183 \\ 0.1113 & 0.2568 \\ -0.1232 & -0.3535 \\ 0.0043 & 0.1263 \\ 0.2896 & 0.0352 \\ -0.2924 & -0.1488 \\ 0.2656 & -0.1874 \\ -0.2286 & 0.2698 \\ -0.113 & -0.2295 \\ 0.1747 & -0.0049 \\ 0.1458 & -0.1498 \\ 0.0484 & -0.202 \\ 0.3529 & -0.0543 \\ -0.3473 & -0.0036 \\ 0.1246 & -0.0774 \\ -0.2028 & -0.3509 \\ -0.0016 & 0.0059 \\ -0.019 & -0.1279 \\ 0.0384 & 0.313 \\ -0.0116 & -0.1706 \\ -0.2961 & 0.1584 \\ 0.0461 & -0.1289 \\ -0.1318 & -0.3874 \\ 0.2436 & -0.1908 \end{bmatrix} \quad \mathrm{OutW} = \begin{bmatrix} 10.6 \\ 34962.01 \\ -228.72 \\ 31039.42 \\ 32876.78 \\ 0.0013 \\ 30746.25 \\ 32791.86 \\ 0.0427 \\ 8.6694 \\ -12.476 \\ 8.1247 \\ -133843.68 \\ -35.032 \\ 573.72 \\ -61940.9 \\ -38.53 \\ 36.8 \\ 1.68\times10^{-5} \\ 944.98 \\ 417363.84 \\ 32980.31 \\ -399887.66 \\ 63.81 \\ -2.006 \\ 2.87\times10^{-5} \\ -13.614 \end{bmatrix}$$

4. Conclusions

Shortwave radiation flux density (SRFD) modeling can be a key factor in the better control of environmental parameters such as evaporation, transpiration, use in solar structures, etc. In the current study, an improved version of the ELM (IELM) is proposed to overcome the limitation of this method due to random generation of the two matrices, including input weights and bias of hidden neurons. Additionally, the Google Earth Engine environment is employed to extract the monthly satellite-based data of the SRFD from the gridded Daymet product. The datasets were from January 1981 to December 2019. Because the time series concept was used in the SRFD modeling, using the SRFD historical data, the most important lags were found to determine the input combinations. The auto-correlation function was applied to find the most effective lags. This function indicated that Lags 1, 2, 3, and 12 are the most effective ones. Using these lags, fifteen different input combinations were obtained. The best input combination was found through the modeling phase, with Lags 2 and 12 as input variables. It should be noted that the sigmoid was found as the best activation function, and the optimum number of hidden neurons was 27. By considering different iteration numbers from 1000 to 100,000, the best results were obtained at 10,000. Comparison of the developed IELM (R = 0.986, RMSE = 21.11, MAPE = 8.68%, NSE = 0.97) with original ELM (R = 0.949, RMSE = 39.91, MAPE = 26.36%, NSE = 0.89) proved the higher performance of the developed ones. Additionally, the uncertainty analysis results indicated that the widths of uncertainty bands for IELM (WUB = $\pm$3.64) are almost half of this index's calculated value for the original ELM (WUB = $\pm$6.9). A matrix-based equation for the optimum IELM-based model is provided to calculate SRFD for one month ahead using Lags 2 and 12. Using this model in predicting the amount of SRFD can be a practical solution in energy and water resources management. In the current situation, where environmental factors and climate change are occurring, using this type of model can be an effective way to advance the management of goals and achieve a roadmap in the future with a sustainable development approach. In the current study, the developed model was checked for only one station. The developed IELM could be applied to other stations and other real-world problems. The developed IELM applied an iterative process to overcome the limitation of the classical ELM in random generation of the two main matrices (i.e., input weights and bias of hidden neurons). However, an implementation of the developed ELM to overcome the mentioned drawback could also be made in a follow-up study by using the new developed evolutionary algorithms such as sperm swarm optimization (SSO), conscious neighborhood-based crow search algorithm (CCSA), and other evolutionary-based algorithms to optimize the randomly generated parameters for the input weights and bias of hidden neurons matrices. Moreover, the ACF was applied to track the most effective lags dependence structure of the SRFD. As the ACF has a large statistical bias that tends to underestimate the second-order dependence structure, it is recommended to apply alternative methods.

Author Contributions: Conceptualization, I.E., K.S., and H.B.; methodology, I.E., A.A., and H.B.; software, I.E. and A.A.; validation, I.E., K.S., and H.B.; formal analysis, I.E., K.S., A.A., and H.B.; resources, K.S., A.A., and H.B.; data curation, K.S. and A.A.; writing—original draft preparation, I.E., M.F., C.A.M., and H.B.; writing—review and editing, I.E., M.F., C.A.M., and H.B.; visualization, I.E. and K.S.; supervision, H.B.; project administration, H.B.; funding acquisition, H.B. All authors have read and agreed to the published version of the manuscript. Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This research was funded by the Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grant (#RGPIN-2020-04583).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tang, B.; Li, Z.L.; Zhang, R. A direct method for estimating net surface shortwave radiation from MODIS data. *Remote Sens. Environ.* **2006**, *103*, 115–126. [CrossRef]
2. Wu, H.; Ying, W. Benchmarking machine learning algorithms for instantaneous net surface shortwave radiation retrieval using remote sensing data. *Remote Sens.* **2019**, *11*, 2520. [CrossRef]
3. Klassen, S.; Bugbee, B. Shortwave radiation. *Micrometeorol. AES Syst.* **2005**, *47*, 43–57.
4. Hatzianastassiou, N.; Matsoukas, C.; Fotiadi, A.; Pavlakis, K.G.; Drakakis, E.; Hatzidimitriou, D.; Vardavas, I. Global distribution of Earth's surface shortwave radiation budget. *Atmos. Chem. Phys.* **2005**, *5*, 2847–2867. [CrossRef]
5. Ceppi, P.; Zelinka, M.D.; Hartmann, D.L. The response of the Southern Hemispheric eddy-driven jet to future changes in shortwave radiation in CMIP5. *Geophys. Res. Lett.* **2014**, *41*, 3244–3250. [CrossRef]
6. Wang, B.; Zheng, L.; Liu, D.L.; Ji, F.; Clark, A.; Yu, Q. Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *Int. J. Climatol.* **2018**, *38*, 4891–4902. [CrossRef]
7. Rabehi, A.; Guermoui, M.; Lalmi, D. Hybrid models for global solar radiation prediction: A case study. *Int. J. Ambient Energy* **2020**, *41*, 31–40. [CrossRef]
8. Wallenberg, N.; Lindberg, F.; Holmer, B.; Thorsson, S. The influence of anisotropic diffuse shortwave radiation on mean radiant temperature in outdoor urban environments. *Urban Clim.* **2020**, *31*, 100589. [CrossRef]
9. Slater, A.G. Surface solar radiation in North America: A comparison of observations, reanalyses, satellite, and derived products. *J. Hydrometeorol.* **2016**, *17*, 401–420. [CrossRef]
10. Soares, J.; Alves, M.; Ribeiro, F.N.D.; Codato, G. Surface radiation balance and weather conditions on a non-glaciated coastal area in the Antarctic region. *Polar Sci.* **2019**, *20*, 117–128. [CrossRef]
11. Wei, Y.; Zhang, X.; Hou, N.; Zhang, W.; Jia, K.; Yao, Y. Estimation of surface downward shortwave radiation over China from AVHRR data based on four machine learning methods. *Solar Energy* **2019**, *177*, 32–46. [CrossRef]
12. Schwarz, M.; Folini, D.; Yang, S.; Allan, R.P.; Wild, M. Changes in atmospheric shortwave absorption as important driver of dimming and brightening. *Nat. Geosci.* **2020**, *13*, 110–115. [CrossRef]
13. Zhang, X.; Liang, S.; Zhou, G.; Wu, H.; Zhao, X. Generating Global Land Surface Satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data. *Remote Sen. Environ.* **2014**, *152*, 318–332. [CrossRef]
14. Ryu, Y.; Jiang, C.; Kobayashi, H.; Detto, M. MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. *Remote Sens. Environ.* **2018**, *204*, 812–825. [CrossRef]
15. Schellekens, J.; Dutra, E.; Martínez-de la Torre, A.; Balsamo, G.; Dijk, A.V.; Sperna Weiland, F.; Minvielle, M.; Calvet, J.C.; Decharme, B.; Eisner, S.; et al. A global water resources ensemble of hydrological models: The eartH2Observe Tier-1 dataset. *Earth Syst. Sci. Data* **2017**, *9*, 389–413. [CrossRef]
16. Zeng, L.; Xia, T.; Elsayed, S.K.; Ahmed, M.; Rezaei, M.; Jermsittiparsert, K.; Dampage, U.; Mohamed, M.A. A Novel Machine Learning-Based Framework for Optimal and Secure Operation of Static VAR Compensators in EAFs. *Sustainability* **2021**, *13*, 5777. [CrossRef]
17. Kamolov, A.A.; Park, S. Prediction of Depth of Seawater Using Fuzzy C-Means Clustering Algorithm of Crowdsourced SONAR Data. *Sustainability* **2021**, *13*, 5823. [CrossRef]
18. Haq, I.U.; Khan, Z.Y.; Ahmad, A.; Hayat, B.; Lee, Y.E.; Kim, K.I. Evaluating and Enhancing the Robustness of Sustainable Neural Relationship Classifiers Using Query-Efficient Black-Box Adversarial Attacks. *Sustainability* **2021**, *13*, 5892. [CrossRef]
19. Soltani, K.; Amiri, A.; Zeynoddin, M.; Ebtehaj, I.; Gharabaghi, B.; Bonakdari, H. Forecasting monthly fluctuations of lake surface areas using remote sensing techniques and novel machine learning methods. *Theor. Appl. Climatol.* **2021**, *143*, 713–735. [CrossRef]
20. Aissani, N.; Beldjilali, B.; Trentesaux, D. Use of machine learning for continuous improvement of the real-time heterarchical manufacturing control system performances. *Int. J. Ind. Syst. Eng.* **2008**, *3*, 474–497. [CrossRef]
21. Guyon, I.; Chaabane, I.; Escalante, H.J.; Escalera, S.; Jajetic, D.; Lloyd, J.R.; Macià, N.; Ray, B.; Romaszko, L.; Sebag, M.; et al. A brief review of the ChaLearn AutoML challenge: Any-time any-dataset learning without human intervention. *Workshop Autom. Mach. Learn.* **2016**, *64*, 21–30.
22. Bustillo, A.; Reis, R.; Machado, A.R.; Pimenov, D.Y. Improving the accuracy of machine-learning models with data from machine test repetitions. *J. Intell. Manuf.* **2020**, 1–19. [CrossRef]
23. Koutsoyiannis, D. Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes. *Hydrol. Sci. J.* **2019**, *64*, 1013–1037. [CrossRef]
24. Shamshirband, S.; Mosavi, A.; Rabczuk, T.; Nabipour, N.; Chau, K.W. Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 805–817. [CrossRef]
25. Alizamir, M.; Kim, S.; Zounemat-Kermani, M.; Heddam, S.; Kim, N.W.; Singh, V.P. Kernel Extreme Learning Machine: An Efficient Model for Estimating Daily Dew Point Temperature Using Weather Data. *Water* **2020**, *12*, 2600. [CrossRef]

26. Yaseen, Z.M.; Sulaiman, S.O.; Deo, R.C.; Chau, K.W. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *J. Hydrol.* **2019**, *569*, 387–408. [CrossRef]
27. Abbaa, S.I.; Elkiranb, G.; Nouranic, V. Improving novel extreme learning machine using PCA algorithms for multi-parametric modeling of the municipal wastewater treatment plant. *Desalin. Water Treat.* **2021**, *215*, 414–426. [CrossRef]
28. Ghazvinei, P.T.; Hassanpour Darvishi, H.; Mosavi, A.; Yusof, K.B.W.; Alizamir, M.; Shamshirband, S.; Chau, K.W. Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *Eng. Appl. Comput. Fluid Mech.* **2018**, *12*, 738–749. [CrossRef]
29. Bonakdari, H.; Ebtehaj, I. A comparative study of extreme learning machines and support vector machines in prediction of sediment transport in open channels. *Int. J. Eng.* **2016**, *29*, 1499–1506.
30. Ebtehaj, I.; Bonakdari, H.; Moradi, F.; Gharabaghi, B.; Khozani, Z.S. An integrated framework of Extreme Learning Machines for predicting scour at pile groups in clear water condition. *Coast. Eng.* **2018**, *135*, 1–15. [CrossRef]
31. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
32. Feng, Z.K.; Niu, W.J.; Tang, Z.Y.; Xu, Y.; Zhang, H.R. Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction. *J. Hydrol.* **2021**, *595*, 126062. [CrossRef]
33. UN General Assembly. Resolution Adopted by the General Assembly on 25 September 2015. 2015. Available online: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E (accessed on 31 March 2020).
34. Rebelatto, B.G.; Salvia, A.L.; Reginatto, G.; Daneli, R.C.; Brandli, L.L. Energy efficiency actions at a Brazilian university and their contribution to sustainable development Goal 7. *Int. J. Sustain. Higher Educ.* **2019**, *20*, 842–855. [CrossRef]
35. Dimitriadis, P.; Koutsoyiannis, D.; Iliopoulou, T.; Papanicolaou, P. A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes. *Hydrology* **2021**, *8*, 59. [CrossRef]
36. Sözen, A.; Arcaklioğlu, E.; Özalp, M. Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data. *Energy Convers. Manag.* **2004**, *45*, 3033–3052. [CrossRef]
37. Dorvlo, A.S.; Jervase, J.A.; Al-Lawati, A. Solar radiation estimation using artificial neural networks. *Appl. Energy* **2002**, *71*, 307–319. [CrossRef]
38. Kim, S.; Seo, Y.; Rezaie-Balf, M.; Kisi, O.; Ghorbani, M.A.; Singh, V.P. Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: Peninsula vs continent. *Theor. Appl. Climatol.* **2019**, *137*, 693–712. [CrossRef]
39. Shook, K.; Pomeroy, J. Synthesis of incoming shortwave radiation for hydrological simulation. *Hydrol. Res.* **2011**, *42*, 433–446. [CrossRef]
40. Gitelson, A.A.; Viña, A.; Masek, J.G.; Verma, S.B.; Suyker, A.E. Synoptic monitoring of gross primary productivity of maize using Landsat data. *IEEE Geosci. Remote Sens.* **2008**, *5*, 133–137. [CrossRef]
41. Che, Y.; Zhang, M.; Li, Z.; Wei, Y.; Nan, Z.; Li, H.; Wang, S.; Su, B. Energy balance model of mass balance and its sensitivity to meteorological variability on Urumqi River Glacier No. 1 in the Chinese Tien Shan. *Sci. Rep. UK* **2019**, *9*, 1–13. [CrossRef]
42. Hamlet, A.F.; Lettenmaier, D.P. Effects of climate change on hydrology and water resources in the Columbia River Basin 1. *J. Am. Water Resour. Assoc.* **1999**, *35*, 1597–1623. [CrossRef]
43. Pellicciotti, F.; Brock, B.; Strasser, U.; Burlando, P.; Funk, M.; Corripio, J. An enhanced temperature-index glacier melt model including the shortwave radiation balance: Development and testing for Haut Glacier d'Arolla, Switzerland. *J. Glaciol.* **2005**, *51*, 573–587. [CrossRef]
44. Thornton, P.E.; Thornton, M.M.; Mayer, B.W.; Wei, Y.; Devarakonda, R.; Vose, R.S.; Cook, R.B. *Daymet: Daily Surface Weather Data on a 1-km Grid for North America*; Version 3; ORNL DAAC: Oak Ridge, TN, USA, 2016. [CrossRef]
45. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
46. Sazib, N.; Mladenova, I.; Bolten, J. Leveraging the google earth engine for drought assessment using global soil moisture data. *Remote Sens.* **2018**, *10*, 1265. [CrossRef] [PubMed]
47. Soltani, K.; Ebtehaj, I.; Amiri, A.; Azari, A.; Gharabaghi, B.; Bonakdari, H. Mapping the spatial and temporal variability of flood susceptibility using remotely sensed normalized difference vegetation index and the forecasted changes in the future. *Sci. Total Environ.* **2021**, *770*, 145288. [CrossRef] [PubMed]
48. Bonakdari, H.; Ebtehaj, I. Verification of equation for non-deposition sediment transport in flood water canals. In Proceedings of the 7th International Conference on Fluvial Hydraulics, RIVER FLOW, Lausanne, Switzerland, 1 January 2014; pp. 1527–1533.
49. Bonakdari, H.; Ebtehaj, I.; Samui, P.; Gharabaghi, B. Lake Water-Level fluctuations forecasting using Minimax Probability Machine Regression, Relevance Vector Machine, Gaussian Process Regression, and Extreme Learning Machine. *Water Resour. Manag.* **2019**, *33*, 3965–3984. [CrossRef]
50. Bonakdari, H.; Moradi, F.; Ebtehaj, I.; Gharabaghi, B.; Sattar, A.A.; Azimi, A.H.; Radecki-Pawlik, A. A Non-Tuned Machine learning technique for Abutment Scour Depth in Clear Water Condition. *Water* **2020**, *12*, 301. [CrossRef]
51. Ebtehaj, I.; Bonakdari, H.; Shamshirband, S. Extreme learning machine assessment for estimating sediment transport in open channels. *Eng. Comput.* **2016**, *32*, 691–704. [CrossRef]

52. Rao, C.R.; Mitra, S.K. *Generalized Inverse of Matrices and Its Applications*; John Wiley & Sons Inc.: New York, NY, USA, 1971.
53. Azimi, H.; Bonakdari, H.; Ebtehaj, I. Gene expression programming-based approach for predicting the roller length of a hydraulic jump on a rough bed. *ISH J. Hydraul. Eng.* **2019**, 1–11. [CrossRef]
54. Ebtehaj, I.; Bonakdari, H.; Gharabaghi, B. A reliable linear method for modeling lake level fluctuations. *J. Hydrol.* **2019**, *570*, 236–250. [CrossRef]
55. Moeeni, H.; Bonakdari, H.; Ebtehaj, I. Integrated SARIMA with neuro-fuzzy systems and neural networks for monthly inflow prediction. *Water Resour. Manag.* **2017**, *31*, 2141–2156. [CrossRef]
56. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
57. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]
58. Shcherbakov, M.V.; Brebels, A.; Shcherbakova, N.L.; Tyukov, A.P.; Janovsky, T.A.; Kamaev, V.A.E. A survey of forecast error measures. *World Appl. Sci. J.* **2013**, *24*, 171–176. [CrossRef]
59. De Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. Mean absolute percentage error for regression models. *Neurocomputing* **2016**, *192*, 38–48. [CrossRef]