environmental sciences proceedings

MDPI

Proceeding Paper

Short-Term Precipitation Forecasting Based on the Improved Extreme Learning Machine Technique †

Isa Ebtehaj [1], Hossein Bonakdari [2,*], Bahram Gharabaghi [3] and Mohamed Khelifi [1]

1 Department of Soils and Agri-Food Engineering, Université Laval, Quebec City, QC G1V 0A6, Canada; isa.ebtehaj.1@ulaval.ca (I.E.); mohamed.khelifi@fsaa.ulaval.ca (M.K.)
2 Department of Civil Engineering, University of Ottawa, 161 Louis Pasteur Private, Ottawa, ON K1N 6N5, Canada
3 School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada; bgharaba@uoguelph.ca
* Correspondence: hbonakda@uottawa.ca; Tel.: +1-613-562-5800
† Presented at the 7th International Electronic Conference on Water Sciences, 15–30 March 2023; Available online: https://ecws-7.sciforum.net.

Abstract: In this study, an improved version of the Extreme Learning Machine, namely Improved Weighted Regularization ELM (IWRELM), is proposed for hourly precipitation forecasting in multi-steps ahead. After finding the optimal values of the proposed method, including the number of hidden neurons, activation function, weight function, regularization parameter, and the effect of the orthogonality, the IWRELM model was calibrated and validated. Thereafter, the calibrated IWRELM model was used to estimate the precipitation up to ten hours ahead. The results indicated that the proposed IWRELM (R = 0.9996; NSE = 0.9993; RMSE = 0.015; MAE = 0.0005) has acceptable accuracy in short-term hourly precipitation forecasting up to ten hours ahead.

Keywords: Extreme Learning Machine (ELM); hourly precipitation; Improved Weighted Regularization Extreme Learning Machine (IWRELM); machine learning, Quebec; real-time forecasting; water resource management

Citation: Ebtehaj, I.; Bonakdari, H.; Gharabaghi, B.; Khelifi, M. Short-Term Precipitation Forecasting Based on the Improved Extreme Learning Machine Technique. *Environ. Sci. Proc.* **2023**, *5*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 15 March 2023

1. Introduction

As a fundamental hydrological variable, precipitation significantly contributes to the land surface and atmospheric processes. Recently, there have been many applications for precipitation forecasting, such as pollutant concentration level monitoring, flood forecasting, and more. Forecasting precipitation is challenging for meteorological scientists due to the precipitation timing and quantity variability. As a result of its persistence and complexity, rainfall forecasting has piqued the interest of academics. Moreover, potential flooding as the result of snow melt and heavy precipitation in early Spring in Canada [1,2], causing significant fatalities and economic damage. Therefore, a quantitative and accurate precipitation forecast can be helpful in formulating appropriate measures and reducing the risk of floods and landslides leading to property damage and loss of life.

It is recognized that existing models use complex statistical models, which are often neither computationally nor economically feasible, or downstream applications are not affected by them. It is therefore being explored as a possible solution to these shortcomings to use machine learning algorithms in combination with the time series concept. The most well-known Machine Learning (ML) is the Artificial Neural Network (ANN) trained with the backpropagation (BP) training algorithm. Although this method has been used by many scholars, it has limitations, such as slow convergence, time-consuming training process, stuck in local minima, overfitting, and low generalization proficiency [3–6]. The Extreme learning machine (ELM) was introduced by Huang et al. [7] as a single-layer feed-forward Neural Network to overcome the classical ANN. The main advantages of

ELM are its high generalization capability [7,8], rapid training process, and least adjustable parameters by the user.

Along with all these advantages, the random determination of more than 66% of the parameters related to the final model (i.e., two matrices of bias of hidden neurons and input weights) [9] is one of the limitations of this method. Ebtehaj et al. [9] suggested a simple iterative process to reduce the effect of random initialization of these two matrices. Moreover, Deng et al. [9] considered the regularization term in the loss function of the ELM. They defined a weighting process to improve the ELM generalizability in the presence of outliers resulting in Weighted Regularized ELM (WRELM). A comparison of the WRELM with the first version of the ELM introduced by [7] proved the higher ability of the WRELM [8,10]. To the authors' best knowledge, no study has been carried out on the application of the WRELM in hourly precipitation forecasting.

In the current study, a computer program was coded in MATLAB environment to develop an improved version of the WRELM (i.e., IWRELM) by taking advantage of the iterative process introduced in [9] and WRELM [10]. The introduced IWRELM is applied for hourly precipitation forecasting. Different parameters of the IWRELM, including the orthogonality effect, the regularization parameter, the weight function, the activation function, and the number of hidden neurons, have been optimized through different defined models. Moreover, various models have been defined for multi-step ahead forecasting of the precipitation in Quebec City, Canada.

2. Materials and Methods

2.1. Study Area

The data used in the current study were recorded hourly at the station of Sainte-Catherine-de-la-Jacques-Cartier from 12/14/1994 to 10/31/2022 (Latitude: 46.8378 & Longitude = −71.6217). The collected data were divided into training and testing stages with a ratio of 50:50. The minimum, average, standard deviation, and maximum of the dataset are zero, 0.0843, 0.5574, and 18.4 mm/h, respectively. All measurements in the dataset were made by the "Ministère de l'Environnement et de la Lutte contre les changements climatiques, de la Faune et des Parcs" [11] of Quebec, Canada.

2.2. Improved Weighted Regularized Extreme Learning Machine (IWRELM)

Suppose we have a dataset with N training samples, x_i (i = 1, 2, …, S) representing the inputs and y_i representing the outputs associated with those inputs. Assuming that the f(x) is used as the activation function and there are h neurons in the hidden layer, the mathematical relationship specified by the IWRELM to map the input variables to the output variables can be stated as follows:

$$\sum_{j=1}^{h} z_j f(\mathbf{o}_j \cdot \mathbf{x}_i + A_j) = y_i, \qquad i = 1, 2, .., S \tag{1}$$

where z_j is the vector of the output weight, f(x) denotes the activation function, $\mathbf{o}_j$ is the input weight matrix, A_i is the bias of hidden neurons, $\mathbf{x}_i$ and y_i denote the input and output variables, respectively, S denotes the number of samples, and h is the number of hidden neurons.

Based on Equation (1), which consists of S equations, a matrix representation of the equation can be expressed as:

$$\mathbf{G}\mathrm{z} = \mathbf{y} \tag{2}$$

$$z = [z_1, ..., z_S]^T \tag{3}$$

$$\mathbf{y} = [y_1, ..., y_S]^T \tag{4}$$

$$\mathbf{G}=\begin{bmatrix} f(\mathbf{o}_1\cdot\mathbf{x}_1+A_1) & \cdots & f(\mathbf{o}_h\cdot\mathbf{x}_1+A_L) \\ \vdots & \ddots & \vdots \\ f(\mathbf{o}_1\cdot\mathbf{x}_S+A_1) & \cdots & f(\mathbf{o}_h\cdot\mathbf{x}_S+A_L) \end{bmatrix}_{S\times h} \tag{5}$$

Due to the random definition of the output weight (i.e., *O*) and bias of hidden neuron (i.e., A), the matrix G is known. Therefore, the only unknown variable in Equation (2) is z. Because the matrix G is, in most cases, not a square matrix, it is not possible to calculate z using Equation (2) directly [7]. To solve this problem, it is easier to use the minimization of the loss function to calculate the optimal least-squares. The loss function in IWRELM is defined as follows:

$$E=\min_z C\left\|\mathbf{W(y\text{-}Gz)}\right\|_2^2+\left\|z\right\|_2^2 \tag{6}$$

where *C* is the regularization parameter and *W* is the weight of each sample through the weighting process in IWRELM. In the weighting process, Equation (6) is considered as Equation (7) and assigned different weights to all samples so that the samples with the least error receive the most weight and vice versa:

$$E=\min_z C\left\|(\mathbf{y}\text{-}\mathbf{G}z)\right\|_2^2+\left\|z\right\|_2^2 \tag{7}$$

The solution of the output weight (i.e., w) from Equation (7) is as follows:

$$\hat{z}=\left(\mathbf{G}^T\mathbf{G}+\mathbf{I}/C\right)^{-1}\mathbf{G}^T\mathbf{y} \tag{8}$$

Here, I is the identity matrix. Moreover, C should not be zero. It could be a positive value less than 1 (0 < C < 1). The applied functions for the weighting process are defined in Table 1. The process of calculating the output weight using Equation (7), the weighting process by the functions provided in Table 1, and the recalculated loss function based on Equation (6) is repeated according to the iteration number defined by the user.

Table 1. Weight functions applied in IWRELM.

No.	Function Definition
1	$\mathrm{w_i}=\begin{cases}1 & \left\|1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z})/\mathrm{IQR}\right\|\le 2.5\\ 2\times(3-\left\|1.349(\mathbf{y}\text{-}\mathbf{G}\mathrm{z})/\mathrm{IQR}\right\|) & 2.5\le\left\|1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z})/\mathrm{IQR}\right\|\le 3\\ 10^{-4} & \text{Otherwise}\end{cases}$
2	$\mathrm{w}=(\mathrm{abs(r)}<1\times(1-\mathrm{r}^2)^2;\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(4.685\times\mathrm{IQR})$
3	$\mathrm{w}=1/\max(1,\mathrm{abs(r)});\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(1.345\times\mathrm{IQR})$
4	$\mathrm{w}=(\mathrm{abs(r)}<\mathrm{pi})\times\sin(\mathrm{r})/\mathrm{r};\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(1.339\times\mathrm{IQR})$
5	$\mathrm{w}=1/(1+\mathrm{abs(r)});\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(1.4\times\mathrm{IQR})$
6	$\mathrm{w}=1/(1+\mathrm{r}^2);\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(2.385\times\mathrm{IQR})$
7	$\mathrm{w}=\tanh(\mathrm{r})/\mathrm{r};\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(1.205\times\mathrm{IQR})$
8	$\mathrm{w}=\tanh(\mathrm{r})/\mathrm{r};\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(2.795\times\mathrm{IQR})$
9	$\mathrm{w}=1\times\mathrm{abs(r)}<1;\ \mathrm{r}=(1.349\times(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))/(2.985\times\mathrm{IQR})$
10	$\mathrm{w}=1/\max(0.0001,\mathrm{abs}(\mathbf{y}\text{-}\mathbf{G}\mathrm{z}))$

* IQR is the interquartile range.

3. Results and Discussion

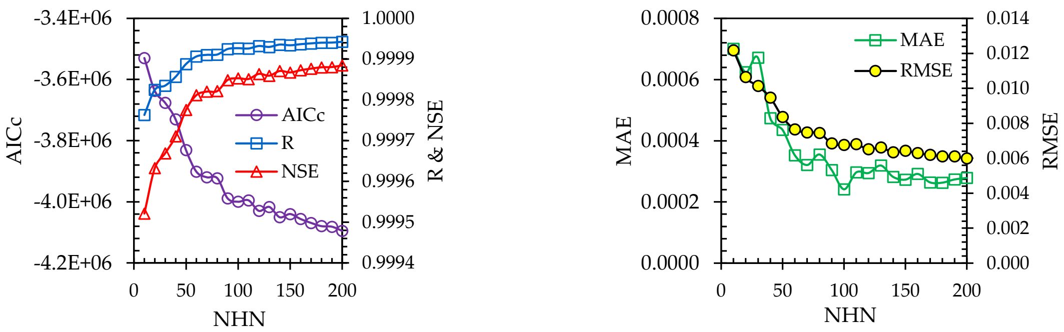
This section details the IWRELM-based modeling in finding the optimal parameters, including the number of hidden neurons (NHN), the activation function, the weight function, the regularization parameter, and the effect of the orthogonality. Finally, the performance of this model in multi-steps ahead precipitation forecasting is investigated using the found optimal values and functions.

Figure 1 shows the statistical indices of the developed IWRELM in hourly precipitation forecasting. Details of the statistical indices provided in Figure 1, including correlation coefficient (R) and Nash–Sutcliffe efficiency (NSE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), and Corrected Akaike Information Criteria could be found in recently published studies [12,13].

The precipitation at the next time is estimated using the value of this variable by considering one to three delays as follows:

$$Pr(t) = f(Pr(t-1), Pr(t-2), Pr(t-3)) \tag{9}$$

where Pr(t), Pr(t-1), P(t-2), and Pr(t-3) denote the precipitation at time t, t-1, t-2, and t-3, respectively.



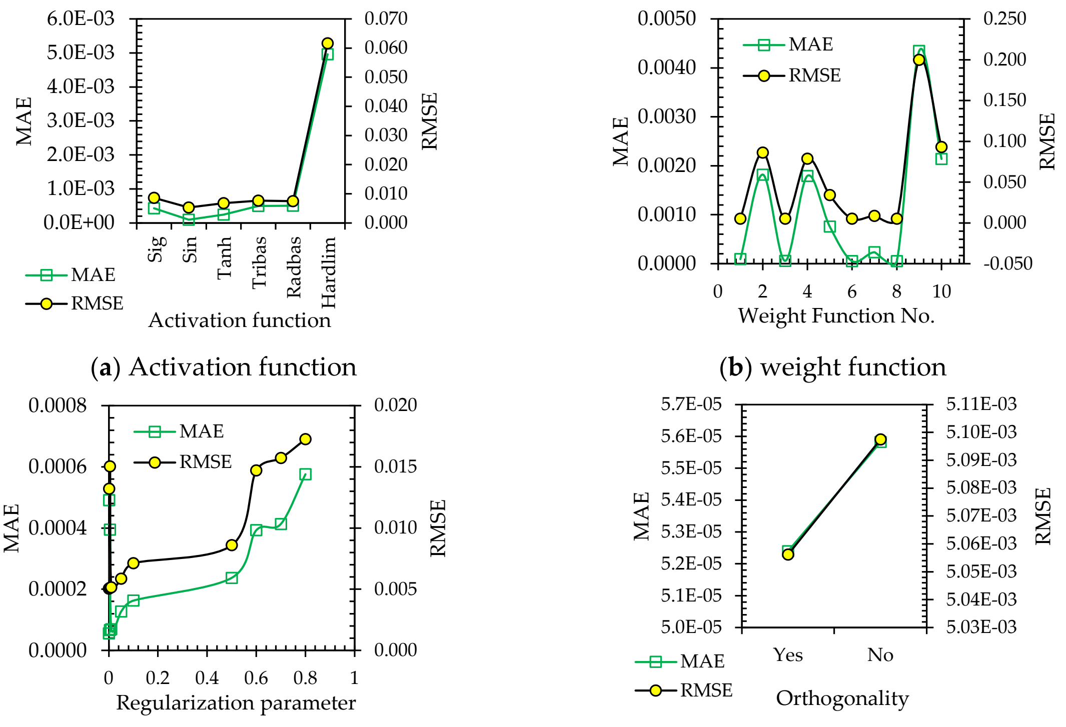
(**a**) AIC and Correlation-based indices (**b**) Absolute-based indices

Figure 1. Performance evaluation of the WRELM with the different number of hidden neurons.

The twenty different values in the range of [10, 200] were checked for NHN. Due to Figure 1b, it can be seen that the R and NSE indices are greater than 0.99, indicating a high level of accuracy in forecasting hourly precipitation with a different number of hidden neurons. Evaluating the effect of NHN in IWRELM indicates that increasing the number of hidden neurons generally enhances the model's accuracy. Based on Figure 1b, increasing the NHN leads to a significant reduction of the MAE and RMSE, especially for NHN of more than 100. For NHN less than 100, an increase of 10 units of NHN significantly changes the values of MAE and RMSE. For example, the MAE and RMSE of NHN = 10 (RMSE = 0.012 & MAE = 0.0007) are more than 10% and 12% lower than the value of these indices for NHN = 20 (RMSE = 0.0106 & MAE = 0.00062) while the difference between the values of these two indices for NHN = 200 and NH = 190 is less than 2%. However, in some instances, the increase in NHN did not correspond directly to the decrease in MAE and RMSE. For example, the values of these two indices corresponding to NHN = 110, 13, and 150 are higher than those corresponding to NHN = 100, 120, and 140, respectively. The reason for this could be related to the random determination of input weights and the bias of hidden neurons (Two main matrices in ELM-based methods such as WRELM), which include more than 66% of the total number of optimized values [8]. Considering that in the models presented in this study, the iteration number is equal to 1000, and when facing such conditions, its value is increased to 100,000, the best results related to IWRELM with different NHNs have been presented.

Considering that the values of regression-based indices in different models are not significantly different (similar to Figure 1), R and NSE are not provided to evaluate other parameters of IWRELM. Figure 2 shows the statistical indices of the IWRELM used for finding the optimal activation function, the weight function provided in Table 1, the regularization parameter (defined in Equation (6)), and the effect of orthogonality on the modeling performance. For the activation function, six different functions (i.e., Sigmoid (Sig), sine (Sin), tangent hyperbolic (Tanh), radial basis function (Radbas), triangular basis function (Tribas), and hard limit (Hardlim)) were compared. The RMSE and MAE of the IWRELM with the activation function Hardlim are more than 7 and 9 times higher than Sig and Radbas, respectively, as the functions that only performed better than Hardlim in these two indices.

Comparison of other functions indicated insignificant differences between them, so that the Sin outperformed the others. For the weight functions provided in Table 1, which is the most crucial feature of the IWRELM compared to the classical ELM, the difference between all the functions is remarkable so that the RMSE and MARE of function No. 9 are about more than four times greater than those for functions No. 1, 3, 6, 7, and 8. The performances of functions No. 1, 3, 6, 7, and 8 are very close, so function No. 3 outperformed others (i.e., functions No. 1, 6, 7, and 8) with a minimal difference, and this function is chosen as the optimal function. The regularization parameter is selected in the range of [0.0001, 0.8]. The results indicated that the best performance of the IWRELM is achieved in the lowest value of the regularization parameter (i.e., C = 0.0001). Moreover, the statistical indices provided in Figure 2d proved the importance of using the orthogonality function to define random initialized matrices (i.e., the bias of hidden neurons and input weights).

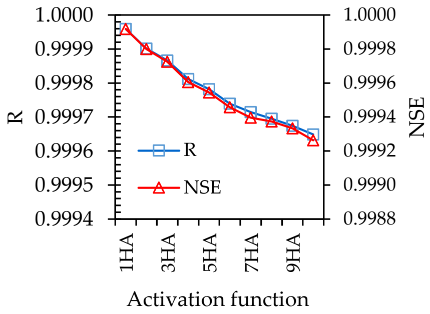


(**a**) Activation function (**b**) weight function

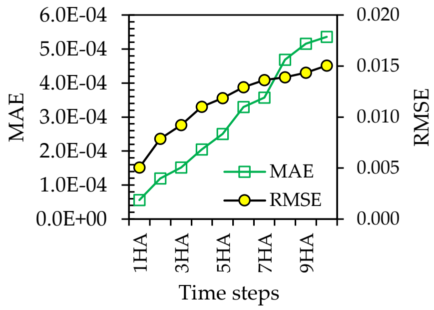
(**c**) Regularization parameter (**d**) Orthogonality

Figure 2. Investigation of the effects of (**a**) Activation function, (**b**) weight function, (**c**) regularization parameter, and (**d**) orthogonality on the IWRELM performance.

Figure 3 shows the statistical indices of the IWRELM in forecasting hourly precipitation from one to ten times ahead. To develop these models, three lags were considered as input variables like Equation (3) for one hour ahead (1HA). For two to ten hours ahead, the first inputs are Pr(t-2) to Pr(t-10), the second inputs are Pr(t-3) to pr(t-11), and the third ones are from Pr(t-4) to Pr(t-12), respectively. It can be seen that increasing the time ahead for forecasting the precipitation (i.e., from 1HAT to 10HA) decreases the values of two correlation-based indices (i.e., R and NSE) and increases the values of the RMSE and MAE indices. Also, the index values for 10HA precipitation forecasting, where the performance is almost the weakest, are acceptable (R = 0.99965; NSE = 0.99926; MAE = 0.00054; RMSE = 0.015). Based on the results of this study, the method developed here (i.e., IWRELM) has a high ability to forecast precipitation several hours in advance.



(**a**) Correlation-based indices



(**b**) Absolute-based indices

Figure 3. Assessment of the WRELM capacity in multi-steps ahead precipitation forecasting.

4. Conclusions

In this study, the Improved Weighted Regularization Extreme Learning Machine (IWRELM) was proposed for hourly forecasting of precipitation one to ten hours ahead. By calibrating this model for one hour ahead, the optimal values of the number of hidden neurons (e.g., 100), the regularization parameter (e.g., C = 0.0001), the most efficient activation function (e.g., Sine function), and the weighting function (e.g., function no. 3 in Table 1) were found. Using the optimal values of the different parameters, the model's performance was checked up to ten hours ahead. The results showed that the model's accuracy decreases with the increase of the forecast time ahead. Nevertheless, the model's accuracy in forecasting precipitation ten hours ahead remains acceptable (R = 0.9996; NSE = 0.9993; RMSE = 0.015; MAE = 0.0005). Since there has been no evaluation of the input variables, a feature selection can be used in future studies and a comparison of the model's performance with optimal inputs can also be performed.

Author Contributions: Conceptualization, I.E. and H.B.; methodology, I.E. and H.B.; software, I.E.; validation, I.E. and H.B.; formal analysis, I.E.; investigation, I.E.; data curation, I.E.; writing—original draft preparation, I.E. and H.B.; writing—review and editing, H.B., M.K. and B.G.; visualization, I.E.; supervision, H.B. and M.K.; project administration, H.B.; All authors have read and accepted the published version of the manuscript.

Data Availability Statement: All data used during the study were provided by a third party. Direct requests for these materials may be made to the provider, as indicated in the Acknowledgements.

Acknowledgments: The authors acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (#RGPIN-2020-04583) and the "Fond de Recherche du Québec- Nature et Technologies", Québec Government (#B2X–315020). Moreover, the authors would like to thank the "Ministère de l'Environnement et de la Lutte contre les changements climatiques, de la Faune et des Parcs" of Quebec, Canada.

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

References

1. Ebtehaj, I.; Bonakdari, H. A comprehensive comparison of the fifth and sixth phases of the coupled model intercomparison project based on the Canadian earth system models in spatio-temporal variability of long-term flood susceptibility using remote sensing and flood frequency analysis. *J. Hydrol*. **2023**, *617*, 128851.
2. Zaghloul, M.S.; Ghaderpour, E.; Dastour, H.; Farjad, B.; Gupta, A.; Eum, H.; Axhari, G.; Hassan, Q.K. Long Term Trend Analysis of River Flow and Climate in Northern Canada. *Hydrology* **2022**, *9*, 197.

3. 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.
4. 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.
5. Bonakdari, H.; Qasem, S.N.; Ebtehaj, I.; Zaji, A.H.; Gharabaghi, B.; Moazamnia, M. An expert system for predicting the velocity field in narrow open channel flows using self-adaptive extreme learning machines. *Measurement* **2020**, *151*, 107202.
6. 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. *Coastal Eng.* **2018**, *135*, 1–15.
7. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501.
8. Ebtehaj, I.; Bonakdari, H. A reliable hybrid outlier robust non-tuned rapid machine learning model for multi-step ahead flood forecasting in Quebec, Canada. *J. Hydrol*. **2022**, *614*, 128592.
9. 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.
10. Deng, W.; Zheng, Q.; Chen, L. Regularized extreme learning machine. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Data Mining, Nashville, TN, USA, 30 March–2 April 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 389–395.
11. Ministère de l'Environnement et de la Lutte contre les changements climatiques, de la Faune et des Parcs. Données du Réseau de surveillance du climat du Québec, Direction de la qualité de l'air et du climat, Québec. 2022.
12. Walton, R.; Binns, A.; Bonakdari, H.; Ebtehaj, I.; Gharabaghi, B. Estimating 2-year flood flows using the generalized structure of the Group Method of Data Handling. *J. Hydrol.* **2019**, *575*, 671–689.
13. Zeynoddin, M.; Ebtehaj, I.; Bonakdari, H. Development of a linear based stochastic model for daily soil temperature prediction: One step forward to sustainable agriculture. *Comput. Electron. Agric.* **2020**, *176*, 105636.