



Proceeding Paper

# Time-Series-Based Air Temperature Forecasting Based on the Outlier Robust Extreme Learning Machine <sup>†</sup>

Isa Ebtehaj <sup>1</sup>, Hossein Bonakdari <sup>2,\*</sup>, Bahram Gharabaghi <sup>3</sup> and Mohamed Khelifi <sup>1</sup>

<sup>1</sup> 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.)

<sup>2</sup> Department of Civil Engineering, University of Ottawa, 161 Louis Pasteur Private, Ottawa, ON K1N 6N5, Canada

<sup>3</sup> School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada; bgharaba@uoguelph.ca

\* Correspondence: hbonakda@uottawa.ca; Tel.: +1-613-562-5800

<sup>†</sup> 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 outlier robust extreme learning machine (IORELM) is introduced as a new method for multi-step-ahead hourly air temperature forecasting. The proposed method was calibrated and used to estimate the hourly air temperature for one to ten hours in advance after finding its most optimum values (i.e., orthogonality effect, activation function, regularization parameter, and the number of hidden neurons). The results showed that the proposed IORELM has an acceptable degree of accuracy in predicting hourly temperatures ten hours in advance ( $R = 0.95$ ;  $NSE = 0.89$ ;  $RMSE = 3.74$ ;  $MAE = 1.92$ ).

**Keywords:** extreme learning machine (ELM); hourly air temperature forecasting; machine learning; outlier robust extreme learning machine (ORELM); Quebec; real-time forecasting; water resource management



**Citation:** Ebtehaj, I.; Bonakdari, H.; Gharabaghi, B.; Khelifi, M.

Time-Series-Based Air Temperature Forecasting Based on the Outlier Robust Extreme Learning Machine.

*Environ. Sci. Proc.* **2023**, *25*, 51.

[https://doi.org/10.3390/](https://doi.org/10.3390/ECWS-7-14236)

[ECWS-7-14236](https://doi.org/10.3390/ECWS-7-14236)

Academic Editor: Athanasios Loukas

Published: 16 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

One of the most significant challenges facing humanity is climate change mitigation. Despite the existing challenges in forecasting climate change effects on Earth, there is scientific agreement on its detrimental consequences. The effects of climate change have been identified as adversely affecting ecosystems, soil erosion, reducing biodiversity, rising sea levels, extreme temperature changes, and global warming. Additionally, a significant impact is expected on food security, human health, energy consumption, and economy. Forecasting air temperatures, in particular, has become an increasingly crucial climatic aspect in various fields, including tourism, energy, agriculture, industry, and so on [1,2]. There are several applications for air temperature forecasting, including forecasting cooling and energy consumption for residential buildings [3], adaptively controlling greenhouse temperatures [4], and predicting natural hazards [5]. As a result, there is a need to reliably anticipate air temperature, because it would assist in a planning horizon for constructing an energy policy, an insurance policy, and infrastructure upgrades, as well as business development, when combined with the study analysis of additional elements in the topic of interest.

Machine Learning (ML) helps improve several types of systems so that their use is expanded. Although feed-forward neural network (FFNN) is the most well-known and widely-used MA-based technique in modeling complex systems, it has some weaknesses, such as overfitting, trapping in local minima, lengthy training process, and slow convergence [6–9]. Huang et al. [10] proposed a single-layer FFNN (SLFFNN) known as the extreme learning machine (ELM) to overcome the mentioned limitations. Compared with conventional FFNN methods, ELM requires minimal user involvement, provides

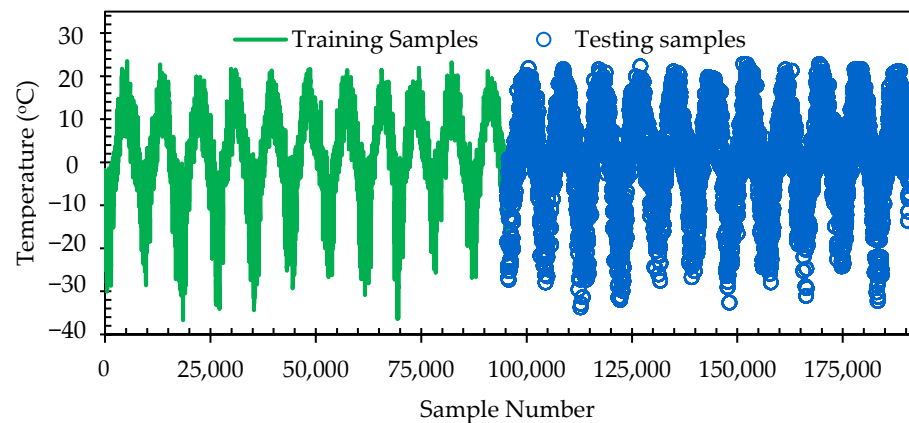
rapid training, and has high generalizability [10,11]. Although this method has many advantages, it has several limitations, one of which is its low performance in the presence of outliers. Zhang and Luo [12] proposed the outlier robust ELM (ORELM) to improve the model performance by increasing sparsity. Moreover, the second limitation of the ELM is the random determination of the two main matrices (i.e., input weights and bias of hidden neurons), which account for at least 66% of all tuned parameters of the final model [13]. It was suggested by Ebtehaj et al. [13] that, by introducing a simple iterative process, the random initialization of the two matrices would be reduced to a minimum. To the authors' best knowledge, there is no study on applying the ORELM in hourly air temperature forecasting.

This research aimed to develop a novel technique and explore the potential of new data intelligence models by integrating the ORELM and iterative process defined by Ebtehaj et al. [13] to forecast hourly air temperature in Quebec City, Canada. The parameters of the developed improved version of the ORELM (IORELM), including the activation function, regularization parameter, and the number of hidden neurons, were optimized by defining different models. Moreover, the performance of the IORELM was evaluated for multi-step-ahead forecasting of the hourly air temperature.

## 2. Materials and Methods

### 2.1. Study Area

A time series of hourly air temperature data was collected at the station of Sainte-Catherine-de-la-Jacques Cartier (latitude: 46.8378 and longitude = -71.6217) from 1 January 2001 to 30 November 2022 for the present study (Figure 1). A 50:50 ratio was used to split the collected data into training and testing stages (i.e., 95,376 samples were collected for each stage). It has a minimum, an average, a standard deviation, and a maximum of -36.37, 1.6, 11.455, and 23.5 °C, respectively. Measurements of the dataset were conducted by the "Ministère de l'Environnement et de la Lutte contre les changements climatiques, de la Faune et des Parcs" [14] of Québec, Canada.



**Figure 1.** Time series of historical temperature for both the training and testing stages.

### 2.2. Improved Outlier Robust Extreme Learning Machine (IORELM)

Considering several  $k$  random samples as  $(x_i, t_i)$  and  $A(x)$  as the activation function, the general form of the SLFFNN is defined as follows:

$$\sum_{j=1}^L W_j f(\mathbf{In}W_j \cdot x_i + S_j) = y_i, \quad i = 1, 2, \dots, k \quad (1)$$

where  $k$  is the number of samples;  $x_i$  and  $y_i$  are the input and output variables, respectively;  $S_j$  is the bias of hidden neurons;  $\mathbf{In}W_j$  is the input weight matrix;  $A(x)$  is the activation function;  $W_j$  is the vector of the output weight; and  $L$  is the number hidden neurons.

A matrix representation of the above-mentioned equation, which is composed of  $L$  equations, is given below:

$$\mathbf{HW} = \mathbf{y} \tag{2}$$

$$\mathbf{W} = [W_1, \dots, W_k]^T \tag{3}$$

$$\mathbf{y} = [y_1, \dots, y_k]^T \tag{4}$$

$$\mathbf{H} = \begin{bmatrix} A(\mathbf{InW}_1 \cdot \mathbf{x}_1 + S_{A1}) & \cdots & A(\mathbf{InW}_h \cdot \mathbf{x}_1 + S_L) \\ \vdots & \ddots & \vdots \\ A(\mathbf{InW}_1 \cdot \mathbf{x}_k + S_1) & \cdots & A(\mathbf{InW}_h \cdot \mathbf{x}_k + S_L) \end{bmatrix}_{k \times L} \tag{5}$$

The matrix  $\mathbf{H}$  could be calculated easily owing to the random definition of bias of hidden neurons (i.e.,  $S$ ) and the output weight (i.e.,  $\mathbf{InW}$ ). Consequently, the only unknown variable in the matrix form of the equations provided in Equation (2) is  $\mathbf{W}$  (i.e., output weight matrix). Ebtehaj and Bonakdari [15] mentioned that the number of tuned parameters through modeling must be less than the number of training samples. Therefore, Equation (2) is not square in most cases and solving it to find the unknown variable is not simple. Thus, the lost function defined for IORELM is as follows:

$$E_{\text{ORELM}} = \min_W \|\mathbf{e}\|_1 + \frac{1}{C} \|\mathbf{W}\|_2^2 \text{ subjected to } \mathbf{y} - \mathbf{HW} = \mathbf{e} \tag{6}$$

where  $C$  is the regularization parameter. Because solving this equation results in a constrained convex optimization problem, the augmented Lagrange of Equation (6) is defined as follows:

$$L(\mathbf{W}, \mathbf{e}, \lambda) = \|\mathbf{e}\|_1 + \frac{1}{C} \|\mathbf{W}\|_2^2 + \lambda^T (\mathbf{y} - \mathbf{HW} - \mathbf{e}) + \frac{\mu}{2} \|\mathbf{y} - \mathbf{HW} - \mathbf{e}\|_2^2 \tag{7}$$

where  $\mu$  is a penalty parameter, defined as  $\mu = 2k / \|\mathbf{y}\|_1$ , and  $\lambda$  is the Lagrange multiplier vector. The explicit solution of Equation (6) using the defined augmented Lagrange vector in Equation (7) is as follows:

$$\beta_{k+1} = (\mathbf{H}^T \mathbf{H} + 2/C\mu \mathbf{I})^{-1} \mathbf{H}^T (\mathbf{y} - \mathbf{e}_k + \lambda_k / \mu) \tag{8}$$

where

$$\begin{aligned} \mathbf{e}_{k+1} &= \text{shrink}(\mathbf{y} - \mathbf{HW}\beta_{k+1} + \lambda_k / \mu, 1/\mu) \\ &\cong \max\{|\mathbf{y} - \mathbf{HW}_{k+1} + \lambda_k / \mu| - 1/\mu, 0\} \circ \text{sign}(\mathbf{y} - \mathbf{HW}_{k+1} + \lambda_k / \mu) \end{aligned} \tag{9}$$

where “ $\circ$ ” is the element-wise multiplication and  $\mathbf{I}$  is the identity matrix.

For all defined iteration numbers by the user, the output weight was calculated and the value of the samples in both training and testing was computed using the three matrices (i.e., input weights, bias of hidden neurons, and output weights) and the activation function. Finally, the optimized values of the mentioned matrices related to the model with the lowest testing error were stored as the final model.

### 3. Results and Discussion

In this section, the IORELM-based model is discussed in detail in terms of finding the optimal parameters, including the regularization parameter, the activation function, and the number of hidden neurons (NHN). Using the most optimum values of the NHN, the regularization parameter, and the best activation function, the performance of the IORELM in multi-step-ahead forecasting of the hourly air temperature was assessed.

The performance of the IORELM is evaluated in Figure 2 by considering NHN in the range of [2, 20]. The performance of NHN = 1 was also evaluated, but its use in a figure prevented the observation of the differences between other values, so it was not included in this figure. Figure 2 provides information on some statistical indices, including corrected Akaike information criteria, root mean square errors (RMSEs), mean absolute errors (MAEs), Nash–Sutcliffe efficiency (NSE), and correlation coefficients (R). The mathematical definition of these indices can be found in recent publications [16,17]. To forecast hourly air temperatures, the following inputs were taken into consideration:

$$T(t) = f(T(t-1), T(t-2), T(t-3)) \tag{10}$$

where  $T(t)$ ,  $T(t-1)$ ,  $T(t-2)$ , and  $T(t-3)$  are the air temperature at time  $t$ ,  $t-1$ ,  $t-2$ , and  $t-3$ , respectively.

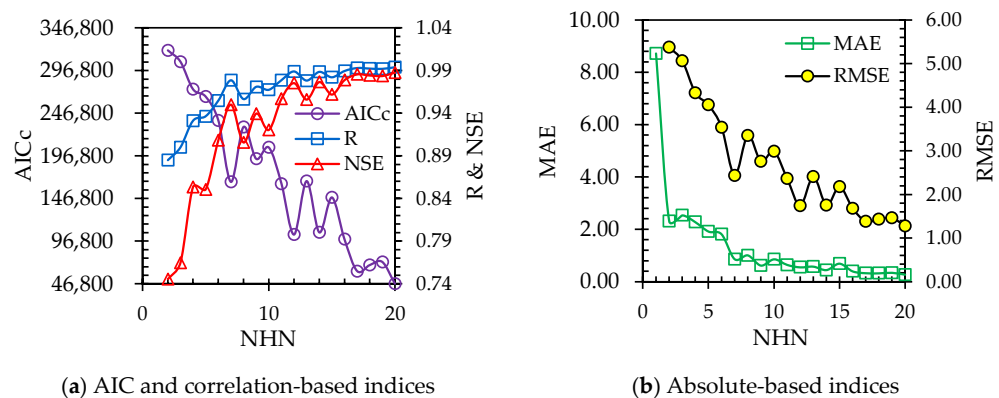
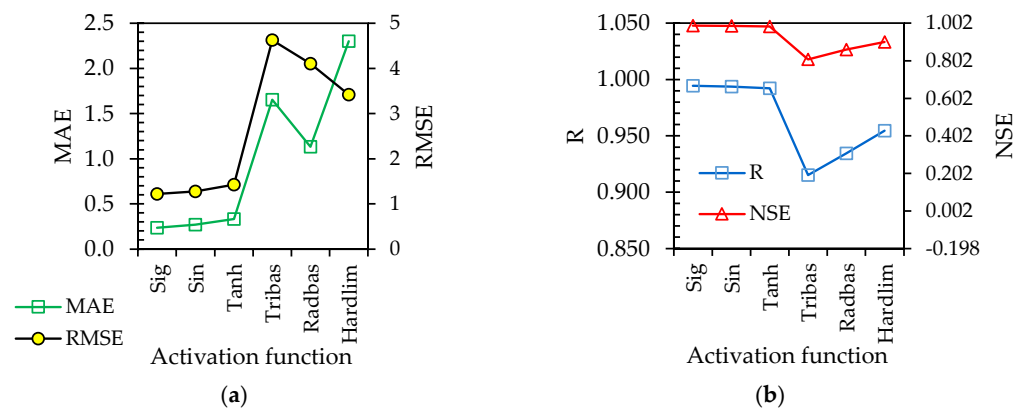


Figure 2. Performance evaluation of the IORELM with the different number of hidden neurons.

The general trend presented in Figure 2 shows that the increase in the number of neurons in the hidden layer has a direct relationship with the modeling accuracy. However, there are some exceptions for some indices. For example, all indices for the models with  $NHN = 8, 12, 14, 17,$  and  $18$  have higher values than those for  $NHN = 9, 13, 15, 18,$  and  $19$ , respectively. It should be noted that, as  $NHN$  increases, the differences between the presented models decrease. There could be explained by the fact that the input weights and biases of hidden neurons are determined randomly, which accounts for more than 66% of the total number of optimized values [13]. Based on this knowledge, the number of iterations increased to 100,000, and the best solution is presented in Figure 2. It can be seen that increasing  $NHN$  by more than 17 does not significantly change the values of  $R, NSE, RMSE,$  and  $MARE$  indices. Nevertheless, the  $AICc$  index, which considers the accuracy and simplicity of the model, should also be evaluated. Considering that the lowest value of  $AICc$  was observed in  $NHN = 20$  and that there is an insignificant difference between other indices in the high value of  $NHN$ ,  $NHN = 20$  was chosen as the optimal number of neurons.

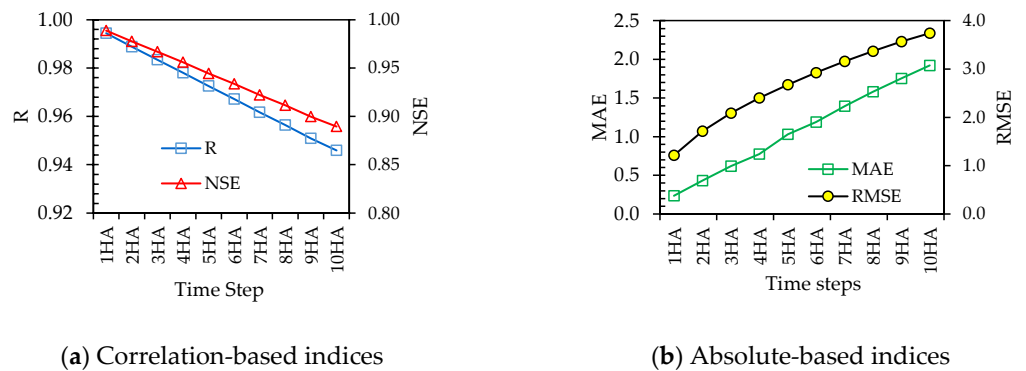
The optimal activation function was found using the optimal number of hidden layer neurons and the functional equation presented in Equation (10). The results presented in Figure 3a show that, among the six functions given in this figure (i.e., Sigmoid (Sig), sine (Sin), tangent hyperbolic (Tanh), radial basis function (Radb), triangular basis function (Tribas), and hard limit (Hardlim)), the performance of the tribas, radbas, and hardlim functions is much weaker than that of the others. The statistical indices indicated that the Sigmoid function ( $R = 0.994; NSE = 0.989; RMSE = 1.217; MAE = 0.236$ ) outperformed the others. The RMSEs of the IORELM modeling using Sig are more than 4.9%, 17%, 280%, 236%, and 180% lower than those for Sin, Tanh, Tribas, Radbas, and Hardlim, respectively. For MARE, these ratios are 14%, 41%, 601%, 380%, and 876%, respectively. Therefore, the Sigmoid function was selected as the most optimum activation function.



**Figure 3.** Investigation of the effects of the (a) activation function and (b) regularization parameter on the IORELM performance.

Finding the optimum regularization parameters (i.e., 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, and 0.5) is considered in Figure 3b. The trend of changes presented in this figure shows that there is no clear relationship between the regularization parameter and the model’s accuracy, so it can be seen that MAE at  $C = 0.01$  is better than the value of this index at  $C = 0.005$  and  $0.5$ . The RMSE and MAE ranges for all regularization parameter values are [1.217, 1.235] and [0.231, 0.275], respectively. Although the difference between the presented values is significant, the best performance of IORELM was obtained at the lowest value of the regularization parameter (i.e.,  $C = 0.0001$ ).

Figure 4 illustrates the statistical indices of IORELM in forecasting hourly air temperatures one to ten hours ahead. Three lags were considered as input variables for developing these models, as described in Equation (10) for one hour ahead (1HA). Two to ten hours ahead, the first inputs are  $T(t-2)$  to  $T(t-10)$ , the second inputs are  $T(t-3)$  to  $T(t-11)$ , and the third inputs are  $T(t-4)$  to  $T(t-12)$ , respectively. The increase in forecast time (i.e., from 1HA to 10HA) decreases the values of two correlation-based indices (R and NSE) while increasing the values of RMSE and MAE. The relative differences of the value of one index for 2HA to 10HA compared with the 1HA indicated that this ratio for R, NSE, MAE, and RMSE is [0.56%, 4.48%], [1.13%, 10.04%], [82.84%, 714.7%], and [40.87%, 207.53%], respectively. However, it can be seen that the results of hourly air temperature for 10HA ( $R = 0.95$ ;  $NSE = 0.89$ ;  $RMSE = 3.74$ ;  $MAE = 1.92$ ) are acceptable. It should be noted that this model has predicted 95,376 different samples with only 20 neurons in the hidden layer, which confirms the model’s simplicity.



**Figure 4.** Assessment of the IORELM capacity in multi-step-ahead precipitation forecasting.

#### 4. Conclusions

This study suggests the use of the improved outlier robust extreme learning machine (IORELM) for forecasting air temperature hourly up to ten hours in advance. The most optimal values of hidden neurons (e.g., 20), the most efficient activation function (e.g.,

Sigmoid function), and the regularization parameters (e.g.,  $C = 0.0001$ ) were determined by calibrating this model for one hour in advance ( $R = 0.994$ ;  $NSE = 0.989$ ;  $RMSE = 1.217$ ;  $MAE = 0.236$ ). Model performance was assessed for one to ten hours in advance using the most optimal values of the different parameters. The more the time ahead increases, the more the accuracy of the model decreases. Nevertheless, the model's performance in terms of ten hours ahead of precipitation forecasting was acceptable ( $R = 0.95$ ;  $NSE = 0.89$ ;  $RMSE = 3.74$ ;  $MAE = 1.92$ ). As the effect of input variables has not been assessed, it is recommended that future studies use feature selection and compare the model's performance with optimal inputs.

**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 agreed to the published version of the manuscript.

**Funding:** This research was funded by 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).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**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 would like to thank the “Ministère de l'Environnement et de la Lutte contre les changements climatiques, de la Faune et des Parcs” of Québec, Canada.

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

## References

1. Tol, R.S. Estimates of the damage costs of climate change. Part 1: Benchmark estimates. *Environ. Resour. Econ.* **2002**, *21*, 47–73. [\[CrossRef\]](#)
2. Pachauri, R.K.; Allen, M.R.; Barros, V.R.; Broome, J.; Cramer, W.; Christ, R.; Church, J.A.; Clarke, L.; Dahe, Q.; Dasgupta, P.; et al. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2014.
3. Chen, G.; Hua, J.; Shi, Y.; Ren, C. Constructing air temperature and relative humidity-based hourly thermal comfort dataset for a high-density city using machine learning. *Urban Clim.* **2023**, *47*, 101400. [\[CrossRef\]](#)
4. Dombaycı, Ö.A.; Gölcü, M. Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey. *Renew. Energy* **2009**, *34*, 1158–1161. [\[CrossRef\]](#)
5. Yu, X.; Hu, X.; Wang, G.; Wang, K.; Chen, X. Machine-Learning Estimation of Snow Depth in 2021 Texas Statewide Winter Storm Using SAR Imagery. *Geophys. Res. Lett.* **2022**, *49*, e2022GL099119. [\[CrossRef\]](#)
6. 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.
7. 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. [\[CrossRef\]](#)
8. 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. [\[CrossRef\]](#)
9. 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. [\[CrossRef\]](#)
10. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#)
12. Zhang, K.; Luo, M. Outlier-robust extreme learning machine for regression problems. *Neurocomputing* **2015**, *151*, 1519–1527. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)



14. 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. Available online: <https://www.environnement.gouv.qc.ca/> (accessed on 23 November 2022).
15. Ebtehaj, I.; Bonakdari, H. Discussion of "Comparative Study of Time Series Models, Support Vector Machines, and GMDH in Forecasting Long-Term Evapotranspiration Rates in Northern Iran" by Afshin Ashrafzadeh, Ozgur Kişi, Pouya Aghelpour, Seyed Mostafa Biazar, and Mohammadreza Askarizad Masouleh. *J. Irrig. Drain. Eng.* **2021**, *147*, 07021005.
16. 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. [[CrossRef](#)]
17. 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. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.