water

MDPI

Article

Machine Learning-Based Water Level Prediction in Lake Erie

Qi Wang [1] and Song Wang [2,*]

[1] Department of Civil Engineering, Queen's University, Kingston, ON K7L 3N6, Canada; qi.wang@queensu.ca
[2] Lassonde School of Engineering, York University, Toronto, ON M3J 1P3, Canada
* Correspondence: wangsong@eecs.yorku.ca; Tel.: +1-416-736-2100 (ext. 33939)

Received: 15 August 2020; Accepted: 21 September 2020; Published: 23 September 2020

Abstract: Predicting water levels of Lake Erie is important in water resource management as well as navigation since water level significantly impacts cargo transport options as well as personal choices of recreational activities. In this paper, machine learning (ML) algorithms including Gaussian process (GP), multiple linear regression (MLR), multilayer perceptron (MLP), M5P model tree, random forest (RF), and k-nearest neighbor (KNN) are applied to predict the water level in Lake Erie. From 2002 to 2014, meteorological data and one-day-ahead observed water level are the independent variables, and the daily water level is the dependent variable. The predictive results show that MLR and M5P have the highest accuracy regarding root mean square error (RMSE) and mean absolute error (MAE). The performance of ML models has also been compared against the performance of the process-based advanced hydrologic prediction system (AHPS), and the results indicate that ML models are superior in predictive accuracy compared to AHPS. Together with their time-saving advantage, this study shows that ML models, especially MLR and M5P, can be used for forecasting Lake Erie water levels and informing future water resources management.

Keywords: machine learning; water levels; Lake Erie

1. Introduction

Water level plays an important part in the community's well-being and economic livelihoods. For example, water level changes can impact physical processes in lakes, such as circulation, resulting in changes in water mixing and bottom sediment resuspension, and thus could further affect water quality and aquatic ecosystems [1,2]. Hence, water level prediction attracts more and more attention [3,4]. For example, the International Joint Commission (IJC) suggests more efforts should be implemented to improve the methods of monitoring and predicting water level [5].

Water-level change is a complex hydrological phenomenon due to its various controlled factors, including meteorological conditions, as well as water exchange between the lake and its watersheds [6,7]. Thus, many tools used to forecast water levels, while considering influencing factors, have been developed, such as process-based models [8]. For example, Gronewold et al. showed that the advanced hydrologic prediction system (AHPS) can be used to capture seasonal and inter-annual patterns of Lake Erie water level from 1997 to 2009 [9]. However, the effectiveness of process-based models mainly depends on the accuracies of the models to represent the aquatic conditions and the abilities of the models in describing the variabilities in the observations [10,11]. In addition, process-based models are often time-consuming [12], so numerous studies have proposed using statistical models to predict water level, e.g., autoregressive integrated moving average model [13], artificial neural network [14], genetic programming [15], and support vector machine [16]. These studies mainly focused on leveraging historical water level without considering the physical process driven by meteorological conditions [6,17,18].

In this paper, six machine learning methodologies [19], with fast leaning speed, are applied to predict daily Lake Erie water level. And then the ML model predictive performance is compared to the process-based model (i.e., AHPS). ML models are established by considering not only the impacts of the historical water-level changes but also meteorological factors. Overall, there are three innovations of this work. First, it is the first work to take account of both historical water level and meteorological conditions in water level prediction for Lake Erie. Second, it is the first study to apply various ML models to forecast Lake Erie water level and compare the performance among various ML models. Third, it is the first work to compare the predictive performance between ML models and the process-based model AHPS.

The following parts of this paper are divided into four sections. Section 1 introduces the experimental area of this study. Section 2 describes various ML algorithms and the model performance assessment metrics. Section 3 presents the independent variables selection and performance comparison results for ML models. Section 4 discusses the forecasting results of various ML predicting models, and the performance comparisons between ML models and the process-based prediction system AHPS (Figure 1).

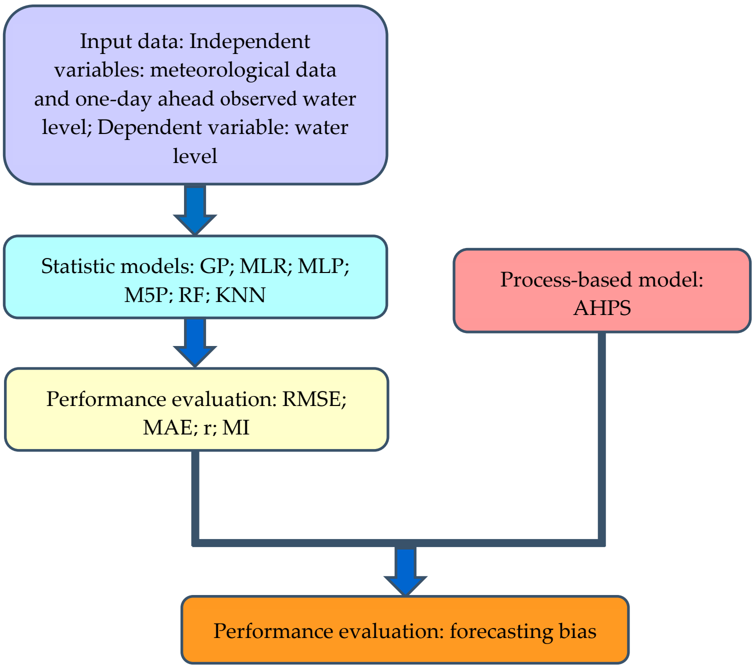
Input data: Independent variables: meteorological data and one-day ahead observed water level; Dependent variable: water level
Statistic models: GP; MLR; MLP; M5P; RF; KNN
Process-based model: AHPS
Performance evaluation: RMSE; MAE; r; MI
Performance evaluation: forecasting bias

Figure 1. Flow chart of this paper.

2. Materials and Methods

2.1. Study Area

Lake Erie is the southernmost, shallowest, and smallest part of the Great Lakes System and its mean depth, surface area, and volume are 18.9 m, 25,657 km^2, and 484 km^3 [20]. It has three distinguishing basins including western, central, and eastern basins, with average depths of 7.4 m, 19 m, and 28.5 m [21,22]. The western basin is separated from the central basin by the islands extending from Point Pelee in Ontario to Marblehead in Ohio, and the eastern basin is separated from the central basin by the Pennsylvania Ridge from Long Point extending to Erie Pennsylvania [21]. There are five major inflows (Detroit River, Maumee River, Sandusky River, Cuyahoga River, and Grand River), and one outflow (Niagara River) for Lake Erie. The water retention capacity of Lake Erie is 2.76 years [20,23]. Figure 2 is generated based on GLERL (Great Lakes Environmental Research Laboratory) data and the identifiers represent the meteorological stations (red stars) and water level measuring stations (black circles).

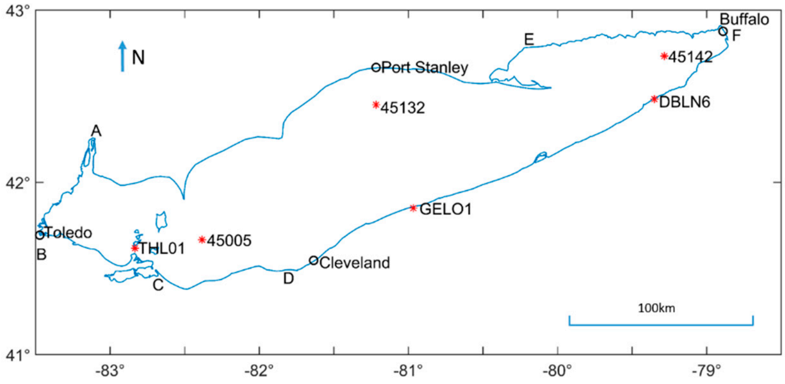

Figure 2. Lake Erie bathymetric plan view (generated from GLERL data); Meteorological stations (red stars); Water level stations (black circles): Toledo, Port Stanley, Buffalo, and Cleveland; A~F represent Detroit, Maumee, Sandusky, Cuyahoga, Grand, and Niagara Rivers.

2.2. Data Source

Following existing studies, e.g., Bucak et al. (2018), who use various meteorological variables to simulate the water level of Lake Beysehir in Turkey, this study also considers meteorological variables including precipitation, air temperature, shortwave radiation, longwave radiation, wind speed, and relative humidity. The average daily measured water level data at four stations including St. Toledo, Port Stanley, Buffalo, and Cleveland are regarded as the observed daily water level of Lake Erie (Table 1).

Table 1. Summary of sources of meteorological and water level data.

Parameter	Source and Location	Frequency

Air temperature	US National Data Buoy Center (NDBC) buoys (western basin, station 45005);
	Environment and Climate Change Canada (ECCC) lake buoy data (central basin, Port Stanley 45132; eastern basin, Port Colborne 45142);
	Great Lakes Environmental Research Laboratory (GLERL) land stations (station THLO1);
	US NDBC land stations (stationGELO1, DBLN6)
		Hourly

Wind speed		
Longwave radiation		
Shortwave radiation		
Relative humidity		
Precipitation	National Oceanic and Atmospheric Administration (NOAA)	Daily
Water level		

The independent variables are selected in terms of the relevance degree between the meteorological variables (daily, one-day-, two-day-, and three-day-ahead observed average, minimum, and maximum values of air temperature, shortwave radiation, longwave radiation, wind speed, and relative humidity; daily, one-day-, two-day-, and three-day-ahead observed average precipitation values) (Table 2) and the water level. The correlation function has been typically applied in previous studies [17,24], measuring the linear dependence between two variables. When the relatedness is not linear, it is not appropriate to apply this method. In this study, we use mutual information (MI) instead of correlation function since MI can measure the general dependence between two variables [25].

Table 2. Summary of variables.

Variable	Description	Variable	Description	Variable	Description	Variable	Description	Variable	Description	Variable	Description
X1	Wind speed(daily ave.)	X2	Air temperature (daily ave.)	X3	Relative humidity (daily ave.)	X4	Shortwave Radiation (daily ave.)	X5	Longwave radiation (daily ave.)	X6	Precipitation (daily ave.)
X7	Wind speed (daily max)	X8	Air temperature (daily max)	X9	Relative humidity (daily max)	X10	Shortwave Radiation (daily max)	X11	Longwave radiation (daily max)	X12	Wind speed (daily min)
X13	Air temperature (daily min)	X14	Relative humidity (daily min)	X15	Shortwave Radiation (daily min)	X16	Longwave radiation (daily min)	X17	Wind speed (2-day ave.)	X18	Air temperature (2-day ave.)
X19	Relative humidity (2-day ave.)	X20	Shortwave Radiation (2-day ave.)	X21	Longwave radiation (2-day ave.)	X22	Precipitation (2-day ave.)	X23	Wind speed (2-day max)	X24	Air temperature (2-day max)
X25	Relative humidity (2-day max)	X26	Shortwave Radiation (2-day max)	X27	Longwave radiation (2-day max)	X28	Wind speed (2-day min)	X29	Air temperature (2-day min)	X30	Relative humidity (2-day min)
X31	Shortwave Radiation (2-day min)	X32	Longwave radiation (2-day min)	X33	Wind speed (3-day ave.)	X34	Air temperature (3-day ave.)	X35	Relative humidity (3-day ave.)	X36	Shortwave Radiation (3-day ave.)
X37	Longwave radiation (3-day ave.)	X38	Precipitation (3-day ave.)	X39	Wind speed (3-day max)	X40	Air temperature (3-day max)	X41	Relative humidity (3-day max)	X42	Shortwave Radiation (3-day max)
X43	Longwave radiation (3-day max)	X44	Wind speed (3-day min)	X45	Air temperature (3-day min)	X46	Relative humidity (3-day min)	X47	Shortwave Radiation (3-day min)	X48	Longwave radiation (3-day min)

MI, i.e., Equation (1), measures the relevance degree between two variables according to the joint probability distribution function $p(x, y)$ [26].

$$\mathrm{MI}(Y;X) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)-p(y)}\right) \tag{1}$$

2.3. Machine Learning Algorithms

This study adopts six widely used ML methods, including Gaussian process, multiple linear regression, multilayer perceptron, M5P model tree, random forest, and k-nearest neighbors.

2.3.1. Gaussian Process

Gaussian process (GP) [27] can be applied to settle two categories of problems: (1) Regression, where the data are continuous, and optimization of study can provide a close prediction for most of the time; (2) Classification, where the datasets are discrete and the predictions end up with a discrete set of classes [28]. In this work, we adopt the same algorithm of GP from existing studies [27–29]. Taking account of the noise of observation y, the Gaussian process is shown below:

$$y = f(x) + \varepsilon \tag{2}$$

$$f(x) = g(m(x), k(x)) \tag{3}$$

where x and y represent the input and output, respectively, f is a multi-dimensional Gaussian distribution determined by m(x) (mean function) and k(x,x′) (covariance matrix), x and x′ are the input values at two points, and $\varepsilon\ (\sim N(0, \sigma^2))$ is the independent white Gaussian noise. To make choosing an appropriate noise level easier, this study applies normalization to the target attribute as well as the other attributes. If the data can be scaled properly, the mean could be zero. A function producing a positive semi-definite covariance matrix with elements $[K]_{i,j} = k(x_i, x_j)$ is applied as the covariance function, so f is expressed as $p(f|x) = N(0, K)$. Based on previous noise assumptions, $p(y|f) = N(f, \sigma^2 I)$ is obtained, and I is the unit matrix.

$$p(y|X) = \int p(y|f)p(f|x)df = N(0, K + \sigma^2 I) \tag{4}$$

The prediction y_* is the output corresponding to the input x_*:

$$p\left(y_* \middle| x, y, x_*\right) = N(\left(y_* | \mu_*, \sigma^2_{y_*}\right) \tag{5}$$

$$\mu_* = k_*^T (K + \sigma^2 I)^{-1} y \tag{6}$$

$$\sigma^2_{y_*} = k_{**} - k_*^T (K + \sigma^2 I)^{-1} k_* \tag{7}$$

where k_* can be obtained from x and x_* through $[k_*] = k(x, x_*)$, and k_{**} can be obtained from x_* through $[k_{**}] = k(x_*, x_*)$. μ_* and $\sigma^2_{y_*}$ represent expectation and variance. The Gaussian covariance function is:

$$k\left(x_i, x_j\right) = \upsilon e^{[-\frac{1}{2} \sum_{d=1}^{m} \omega_d (x_i^d - x_j^d)^2]} \tag{8}$$

where $[\upsilon, \omega_1, \omega_2, \ldots, \omega_m]$ represent a parameter vector, called hyperparameters. The simplest expression can be shown as follows:

$$logp\left(y \middle| x\right) = \frac{1}{2} y^T (K + \sigma^2 I)^{-1} y - \frac{1}{2} \log\left(\left|K + \sigma^2 I\right|\right) - \frac{n}{2} \log(2\pi) \tag{9}$$

2.3.2. Multiple Linear Regression

The multiple linear regression (MLR) method describes the linear relationship between the dependent and independent variables. In this work, we adopt the same MLR algorithm as existing studies [30–32].

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + X_{k,i} + \varepsilon_i \tag{10}$$

where $X_{1,i}$, $X_{2,i}, \ldots,$ $X_{k,i}$ and Y_i are the ith observations of the independent and dependent variables, respectively; β_0, β_1, …, β_k are regression coefficients; and ε_i is the residual for the ith observations.

2.3.3. Multilayer Perceptron

An artificial neural network (ANN) mimics human brain functioning, processing information through a single neuron. In this work, we adopt the same MLP algorithm as existing studies [33–36]. Multiplayer perceptron (MLP), a class of feedforward ANN, has three layers including input, hidden, and output layers. The hidden layers receive the signals from the nodes of the input layer and transform them into signals that are sent to all output nodes, transforming them into the last layer of outputs. The weights between the adjacent layers are adjusted automatically while minimizing the errors between the simulations and observations by using the backpropagation algorithm.

2.3.4. M5P Model Tree

There is a collection of training samples (T); each training sample is featured by a fixed set of attributes, which are input values, and has a corresponding target, which is the output value. In this work, we adopt the same M5P algorithm from existing studies [37–39]. In the beginning, T is associated with a leaf storing linear regression function or split into subsets accounting for the outputs, and then the same process is recursive for the subsets.

In this process, choosing appropriate attributes to split T, in order to minimize the expected error at a particular node, requires a splitting criterion, and standard deviation reduction is the expected error reduction, calculated by Equation (11).

$$SDR = s\,d(T) - \sum_{I} \frac{|T_i|}{|T|} sd(T_i) \tag{11}$$

where T is a group of samples that achieves the node; T_i is the result of splitting the node in terms of the chosen attribute. The model finally chooses the split that maximizes the expected error reduction.

2.3.5. Random Forest

In this work, we adopt the same random forest (RF) algorithm from existing studies [40,41]. Unlike random tree (RT), instead of choosing features, RF can determine the most important ones among all the features after training. It combines many binary decision trees and each tree learns from a random sample of the observations without pruning. In RF, bootstrap samples are drawn to build multiple trees, which means that some samples will be used multiple times in a single tree, and each tree grows with a randomized subset of features from the total number of features. Finally, the output representing the average of each single-tree methodology is generated. Since RF contains a lot of trees, handling a larger number of data, limited generalization errors occur, avoiding overfitting.

2.3.6. K-Nearest Neighbor

Generally, the above machine learning methods can be regarded as eager methods, which means these methods start learning when they receive the data and create a global approximation. On the other hand, k-nearest neighbor (KNN) belongs to lazy learning, which creates a local approximation. In this work, we adopt the same KNN algorithm from existing studies [42–44]. It simply stores the training data without learning until it is given a test set, and calculates in terms of the current data set instead of coming up with an algorithm based on historic data. KNN is a method that can be used for either classification or regression, while distributing the attribute values for a target, which is the average of the values of its k nearest neighbors.

2.4. Model Performance Evaluation

The outputs of each ML model are daily water levels of Lake Erie. Then, these predictions are compared against the observations, which are the average measured water levels at the four stations mentioned above. The model performance evaluation criteria selected include root means square error (RMSE), mean absolute error (MAE), correlation coefficient (r), and mutual information (MI) mentioned in Section 2.2.

RMSE shows the residual value between the predictions and the observations, thus a smaller RMSE value represents a better fit between observations and predictions.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \tag{12}$$

The mean absolute error (MAE) can be also applied to access the accuracy of fitting time-series data. Similar to RMSE, a smaller MAE value represents a better fit.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \tag{13}$$

The correlation coefficient (r) describes the weight of the relationship between observations and predictions and ranges from −1 to 1. The closer to 0, the weaker linear relationship between observations and predictions. For example, a value of zero represents no linear relationship, −1 represents a strong negative linear relationship, and 1 represents a strong positive linear relationship.

$$r = \frac{\sum_{i=1}^{n}(O_i - \overline{O_i})(P_i - \overline{P_i})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O_i})^2 \sum_{i=1}^{n}(P_i - \overline{P_i})^2}} \tag{14}$$

where O_i and P_i represent the observations and predictions at the ith time step, $\overline{O_i}$ and $\overline{P_i}$ are the means of observations and predictions, and n represents total time step numbers.

RMSE and MAPE describe the overall accuracy of the models, while r and MI describe the differences between the observations and the predictions [19].

3. Results

3.1. Input Selection

A MI value of zero represents zero relatedness between two variables, and the larger value represents stronger relatedness [45]. All meteorological variables except one-day-, two-day-, and three-day-ahead observed minimum shortwave radiation (15×, 31×, and 47×) show a strong relevance to water level (Figure 3), so this study uses daily, one-day-, two-day-, and three-day-ahead observed average, minimum, and maximum values of air temperature, longwave radiation, relative humidity, and wind speed; daily, one-day-, two-day-, and three-day-ahead observed average and maximum values of shortwave radiation; daily, one-day-, two-day-, and three-day-ahead observed average values of precipitation as independent variables. One-day-ahead observed water level is also considered as an independent variable since it can also impact water level changes the next day [46].

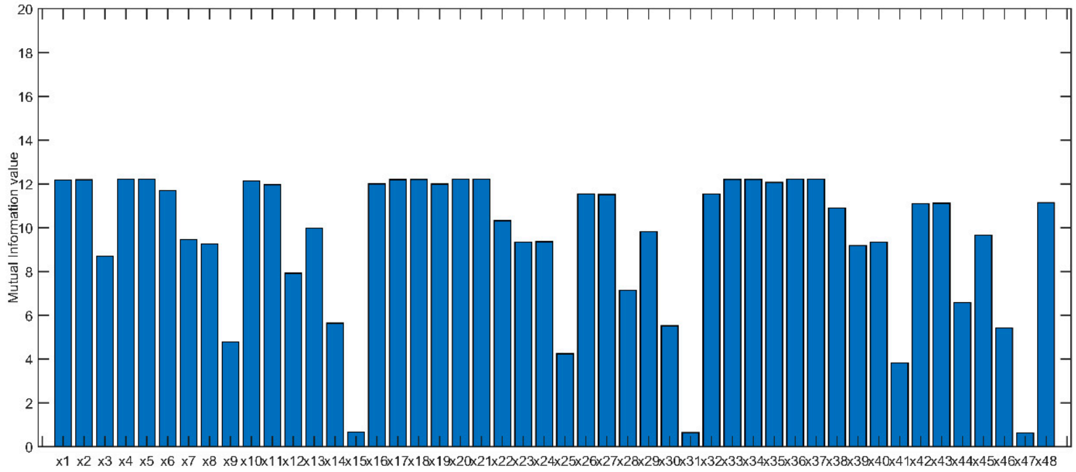

Figure 3. Mutual information of all variables.

Following a previous study [47], we split the whole dataset from 2002 to 2014 into two sections, including the training set (approximately 84% observations) from 2002 to 2012, aiming to establish the model and adjust the weights, as well as the testing set (approximately 16% observations) from 2013 to 2014, aiming to assess the model performance.

3.2. Model Performance Comparison

Table 3 shows the performance of ML models for predicting water level by RMSE, MAE, r, and MI. MLR and M5P provide the most reliable predictions of water level during the testing period from 2013 to 2014 with RMSE and MAE values of 0.02 and 0.01, respectively.

The RMSE values for different ML models in this study are also comparable to previous studies on water-level prediction, e.g., Khan et al. (2006) found RMSE values for the time horizon of 1 to 12 months of 0.057–0.239, 0.085–0.246, and 0.068–0.304 by applying SVM, MLP, and SAR models to Lake Erie [16].

Table 3. Summary of performance for multiple ML models.

Model	RMSE	MAE	r	MI
Gaussian process	0.07	0.06	0.94	8.45
Multiple linear regression	0.02	0.01	0.99	8.54
Multiple Perceptron	0.03	0.02	0.99	8.57
M5P Model Tree	0.02	0.01	0.99	8.53
Random Forest model tree	0.05	0.04	0.97	8.43
KNN	0.10	0.09	0.83	8.25

4. Discussion

4.1. Multiple ML Models Comparison

Based on RMSE and MAE (Table 3), the predictions of ML models are acceptable compared to previous studies. For example, Bazartseren et al. predicted water levels with RMSE values of 1.39 and 3.40 for the Oder and Rhein River in Germany by using artificial neural networks [48]. Yoon et al. predicted groundwater levels in the coastal aquifer in Donghae City, Korea, through two wells with average RMSE values of 0.13 and 0.136 by applying ANN and SVM (support vector machine), respectively [49].

The predicted water levels agree with the observations qualitatively (Figure 4). Due to the long period of testing time (2013–2014), we will divide it into three periods for each year to discuss the predictions in detail. In Figure 5a,b, both observations and predictions show the increasing trends, from January to March in 2013 and 2014, due to low precipitation frequency, the water levels are relatively stable and become lowest during the whole year. Beginning from April, the water level rises with frequent rainfall. From May to August (Figure 5c,d), the water level continues increasing to reach a peak in July in 2013 and 2014, showing the significant effects of precipitation on water level changes. From September to December, with precipitation reducing, the water level decreases.

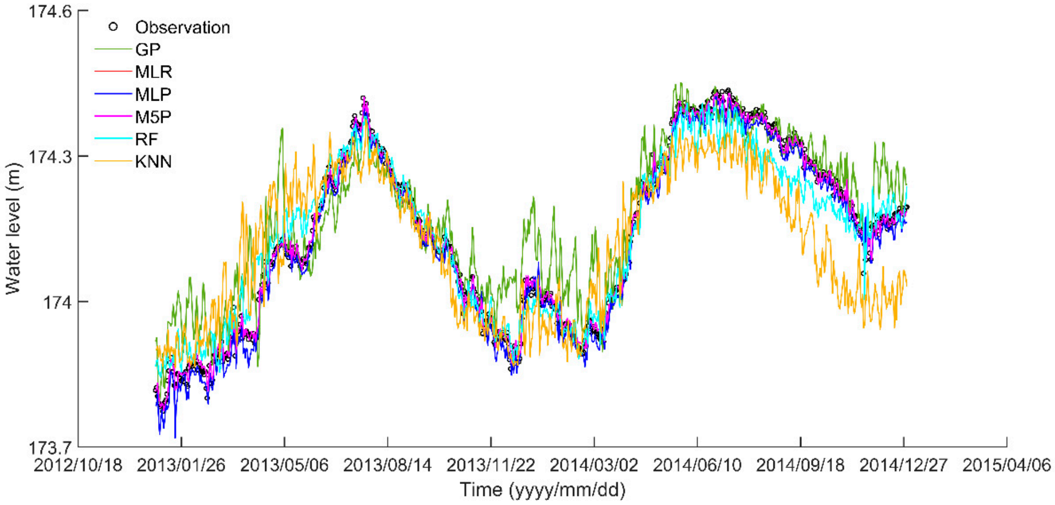

Figure 4. Time series of predicted and observed water level of Lake Erie during the testing period from January 1st 2013 to December 31st 2014.

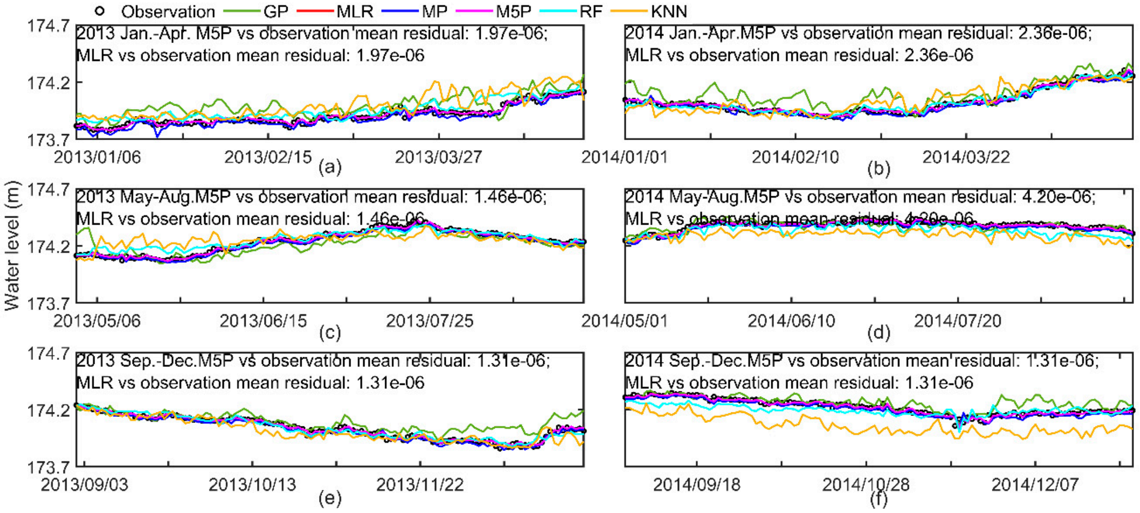

Figure 5. Comparisons between the predictions of six ML models and observations for water level from: (**a**) January to April 2013; (**b**) January to April 2014; (**c**) May to August 2013; (**d**) May to August 2014; (**e**) September to December 2013; (**f**) September to December 2014.

Figure 6 shows that the predicted water levels agree with the observations quantitatively. The slopes are close to 1, indicating that the predictions from ML models show a strong correlation with the observed water levels even though there is a small difference between them. The dots in the KNN comparison figure seem scattered, showing that the predictions are smaller/larger than the observations.

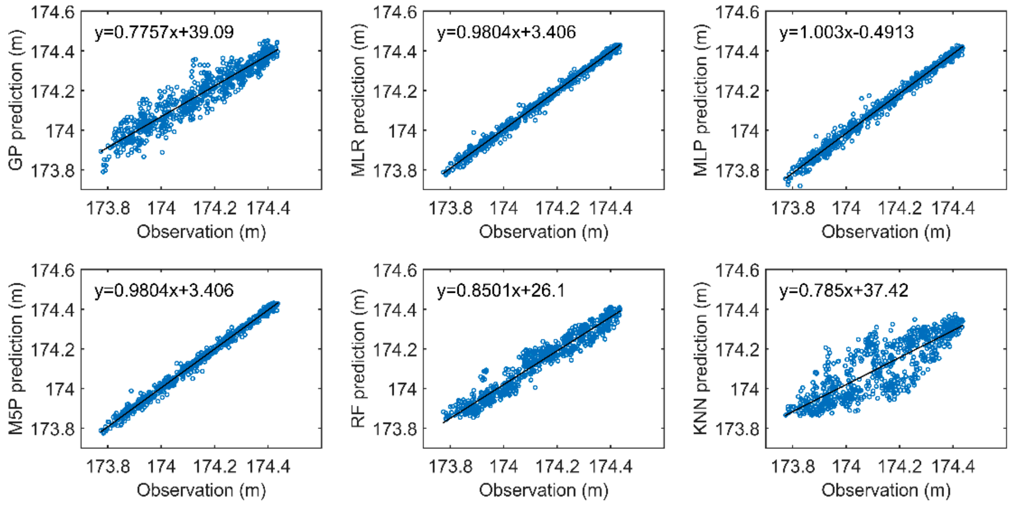

Figure 6. Scatter plots of the observations and predictions from GP, MLR, MLP, M5P, RF, and KNN models during the testing period from January 1st 2013 to December 31st 2014.

Among these six models, the MLR and M5P models show better performance than others with only a 0.003% difference in capturing the peak values. The KNN model underestimates the peaks in 2013 and 2014 (Figure 4) and overestimates the lowest level in winter 2014, which may be due to the k value selection. Choosing small values for k can be noisy and impact the predictions, while the large values can lead to smooth decision boundaries, resulting in low variance but increased bias and expensive computation. In this study, the value of k is set to 69, which is the square root of the total number of samples based on a previous study [50].

4.2. Comparison between ML Models and AHPS

An advanced hydrologic prediction system (AHPS) is a web-based suite of forecast products, and it has been widely used in predicting the water levels and hydrometeorological variables of the Laurentian Great Lakes [51]. In AHPS, the spatial meteorological forcing data obtained from the National Climatic Data Center (NCDC) for sub-basins and over-lakes are averaged by Thiessen Polygon Software, and the results are regarded as the inputs of the lake thermodynamics model, calculating heat storage, and the large basin runoff model, calculating moisture storage, and the net basin supply is calculated based on water balance. Finally, the net basin supplies are translated to water levels and joint flows by the lake routing and regulation model [52]. Gronewold et al. predicted the water level of Lake Erie by using AHPS. He compared 3-month and 6-month mean predicted water level values to the monthly averaged water level observations for 13 years (1997–2009) and assessed AHPS predictive performance by forecasting bias (the averaged differences between the median value of predictions and monthly averaged observations) [9]. In this part, we also compare ML models and AHPS predictive performance in terms of forecasting bias.

All absolute values of forecasting bias based on ML models during the testing period (2013–2014) are smaller than those in AHPS from 1997 to 2009 (Table 4), showing that the predicted median water level values in ML models are much closer to the monthly average observations than AHPS, even though both AHPS and the five ML models (MLR, MLP, M5P, RF, and KNN) underestimate observations with negative forecasting bias. Table 5 indicates the time consumed during training and testing periods in ML models, which are all less than one minute except for the GP model. This is much faster than AHPS. ML models are more accurate than process-based AHPS in forecasting water level, and can also save computational time and expense.

Table 4. Performance comparisons between AHPS and multiple ML models.

Method	Bias in Median Water Level Forecast (cm)	
	3-Month Forecast	6-Month Forecast
AHPS	−5.50	−5.40
Gaussian Process	1.81	2.58
Multiple Linear Regression	−2.05	−3.31
Multilayer Perceptron	−3.64	−4.57
M5P Model Tree	−2.05	−3.31
Random Forest Model Tree	−1.97	−4.25
KNN	−2.42	−3.60

Table 5. Time to train and test multiple ML models.

Model	Train Time	Test Time
Gaussian Process	156.0 s	35.39 s
Multiple Linear Regression	0.28 s	0.61 s
Multilayer Perceptron	53.78 s	0.49 s
M5P Model Tree	2.48 s	0.51 s
Random Forest Model Tree	5.83 s	0.73 s
KNN	<0.1 s	2.35 s

4.3. Impact of Training and Testing Data Selection

In this study, following existing studies [18,53], we used 84% of the total dataset (years 2002–2012) as the training data and the remaining 16% (years 2013–2014) as the testing data. The performance of ML models can vary with different sizes of training data. To explore the impact of the training and testing data selection, we further examined these ML models with different sizes of training data. Specifically, we built ML-based prediction models with different training data and tested their performance on the same set of testing data (years 2013–2014). The detailed results are shown in

Tables 6–11 (training data are from years 2008–2012, 2007–2012, 2006–2012, 2005–2012, 2004–2012, and 2003–2012, respectively). We observed consistent performance of these ML models with different training data, i.e., MLR and M5P are the best ML models for predicting the water level of Lake Erie.

Table 6. Summary of performance of multiple ML models with training data from 2008 to 2012.

Model	RMSE	MAE	r	MI
Gaussian process	0.09	0.07	0.91	8.43
Multiple linear regression	0.02	0.01	0.99	8.52
Multiple Perceptron	0.03	0.02	0.99	8.52
M5P Model Tree	0.02	0.01	0.99	8.54
Random Forest model tree	0.06	0.05	0.95	8.42
KNN	0.10	0.08	0.83	8.17

Table 7. Summary of performance of multiple ML models with training data from 2007 to 2012.

Model	RMSE	MAE	r	MI
Gaussian process	0.09	0.07	0.91	8.45
Multiple linear regression	0.02	0.01	0.99	8.42
Multiple Perceptron	0.02	0.01	0.99	8.59
M5P Model Tree	0.02	0.01	0.99	8.50
Random Forest model tree	0.06	0.05	0.95	8.41
KNN	0.11	0.09	0.80	8.21

Table 8. Summary of performance of multiple ML models with training data from 2006 to 2012.

Model	RMSE	MAE	r	MI
Gaussian process	0.08	0.07	0.92	8.50
Multiple linear regression	0.02	0.01	0.99	8.52
Multiple Perceptron	0.02	0.02	0.99	8.56
M5P Model Tree	0.02	0.01	0.99	8.50
Random Forest model tree	0.05	0.04	0.96	8.44
KNN	0.11	0.09	0.81	8.19

Table 9. Summary of performance of multiple ML models with training data from 2005 to 2012.

Model	RMSE	MAE	r	MI
Gaussian process	0.08	0.06	0.93	8.45
Multiple linear regression	0.02	0.01	0.99	8.53
Multiple Perceptron	0.03	0.02	0.99	8.57
M5P Model Tree	0.02	0.01	0.99	8.53
Random Forest model tree	0.06	0.05	0.95	8.39
KNN	0.11	0.09	0.79	8.21

Table 10. Summary of performance of multiple ML models with training data from 2004 to 2012.

Model	RMSE	MAE	r	MI
Gaussian process	0.08	0.06	0.94	8.44
Multiple linear regression	0.02	0.01	0.99	8.54
Multiple Perceptron	0.03	0.02	0.99	8.57
M5P Model Tree	0.02	0.01	0.99	8.54
Random Forest model tree	0.06	0.04	0.96	8.39
KNN	0.10	0.09	0.82	8.19

Table 11. Summary of performance of multiple ML models with training data from 2003 to 2012.

Model	RMSE	MAE	r	MI
Gaussian process	0.08	0.06	0.94	8.47
Multiple linear regression	0.02	0.01	0.99	8.53
Multiple Perceptron	0.02	0.01	0.99	8.56
M5P Model Tree	0.02	0.01	0.99	8.53
Random Forest model tree	0.05	0.04	0.97	8.44
KNN	0.10	0.08	0.84	8.26

5. Conclusions

Considering multiple independent variables including meteorological data and one-day-ahead observed water level, this study developed multiple ML models to forecast Lake Erie water level and compared the performance among ML models by RMSE and MAE. MLR and M5P, among all ML models, had the best performance in capturing the variations of water level with the smallest RMSE and MAE, which were 0.02 and 0.01, respectively. Furthermore, this paper also compared the performance of process-based AHPS and ML models, showing that ML models have higher accuracy than AHPS in predicting water level. Based on the advantages of high accuracy and short computational cost, ML methods could be used for informing future water resources management in Lake Erie.

Author Contributions: Conceptualization, Q.W. and S.W.; methodology, Q.W. and S.W.; software, Q.W. and S.W.; validation, Q.W. and S.W.; formal analysis, Q.W. and S.W.; investigation, Q.W.; resources, Q.W.; data curation, Q.W.; writing—original draft preparation, Q.W.; writing—review and editing, S.W.; visualization, Q.W.; supervision, S.W.; project administration, S.W.; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research is partially supported by Natural Sciences and Engineering Research Council of Canada.

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

References

1. Håkanson, L.; Parparov, A.; Hambright, K. Modelling the impact of water level fluctuations on water quality (suspended particulate matter) in Lake Kinneret, Israel. *Ecol. Model.* **2000**, *128*, 101–125. [CrossRef]
2. Oganesian, R.; Parparov, A. The problems of Lake Sevan and ways of solution. In Proceedings of the Symposium Biologica Hungarica, Conservation and Management of Lakes, Budapest, Hungary, 11–17 September 1989; Akademiai Kiado: Budapest, Hungary, 1989.
3. Grima, A.; Wilson-Hodges, C. Regulation of Great Lakes water levels: The public speaks out. *J. Great Lakes Res.* **1977**, *3*, 240–257. [CrossRef]
4. International Joint Commission. *Living with the Lakes: Challenges and Opportunities-Annex G Public Information Program*; International Joint Commission: Windsor, ON, Canada, 1989.
5. International Joint Commission. *Levels Reference Study: Great Lakes-St. Lawrence River Basin*; The Board: Windsor, ON, Canada, 1993.
6. Altunkaynak, A. Forecasting surface water level fluctuations of Lake Van by artificial neural networks. *Water Resour. Manag.* **2007**, *21*, 399–408. [CrossRef]
7. Karimi, S.; Shiri, J.; Kisi, O.; Makarynskyy, O. Forecasting water level fluctuations of Urmieh Lake using gene expression programming and adaptive neuro-fuzzy inference system. *IJOCS* **2012**, *3*, 109–125. [CrossRef]
8. Marchand, D.; Sanderson, M.; Howe, D.; Alpaugh, C. Climatic change and great lakes levels the impact on shipping. *Clim. Chang.* **1988**, *12*, 107–133. [CrossRef]
9. Gronewold, A.D.; Clites, A.H.; Hunter, T.S.; Stow, C.A. An appraisal of the Great Lakes advanced hydrologic prediction system. *J. Great Lakes Res.* **2011**, *37*, 577–583. [CrossRef]
10. Arhonditsis, G.B.; Brett, M.T. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Mar. Ecol. Prog. Ser.* **2004**, *271*, 13–26. [CrossRef]
11. Beck, M.B.; Ravetz, J.R.; Mulkey, L.; Barnwell, T.O. Hydraulics. On the problem of model validation for predictive exposure assessments. *Stoch. Hydrol. Hydraul.* **1997**, *11*, 229–254. [CrossRef]

12. Yu, P.S.; Chen, S.T.; Chang, I.F. Support vector regression for real-time flood stage forecasting. *J. Hydrol.* **2006**, *328*, 704–716. [CrossRef]
13. Yu, Z.; Lei, G.; Jiang, Z.; Liu, F. ARIMA modelling and forecasting of water level in the middle reach of the Yangtze River. In Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS), Banff, AB, Canada, 8–10 August 2017.
14. Alvisi, S.; Mascellani, G.; Franchini, M.; Bardossy, A. Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrol. Earth Syst. Sci.* **2006**, *10*, 1–17. [CrossRef]
15. Ghorbani, M.A.; Khatibi, R.; Aytek, A.; Makarynskyy, O.; Shiri, J. Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks. *Comput. Geosci.* **2010**, *36*, 620–627. [CrossRef]
16. Khan, M.S.; Coulibaly, P. Application of support vector machine in lake water level prediction. *J. Hydrol. Eng.* **2006**, *11*, 199–205. [CrossRef]
17. Buyukyildiz, M.; Tezel, G.; Yilmaz, V. Estimation of the change in lake water level by artificial intelligence methods. *Water Resour. Manag.* **2014**, *28*, 4747–4763. [CrossRef]
18. Coulibaly, P. Reservoir computing approach to Great Lakes water level forecasting. *J. Hydrol.* **2010**, *381*, 76–88. [CrossRef]
19. Meinshausen, M.; Smith, S.J.; Calvin, K.; Daniel, J.S.; Kainuma, M.L.T.; Lamarque, J.F.; Matsumoto, K.; Montzka, S.A.; Raper, S.C.B.; Riahi, K.; et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Chang.* **2011**, *109*, 213. [CrossRef]
20. Bolsenga, S.J.; Herdendorf, C.E. *Lake Erie and Lake St. Clair Handbook*; Wayne State University Press: Detroit, MI, UAS, 1993.
21. Boyce, F.M.; Chiocchio, F.; Eid, B.; Penicka, F.; Rosa, F. Hypolimnion flow between the central and eastern basins of Lake Erie during 1977 (interbasin hypolimnion flows). *J. Great Lakes Res.* **1980**, *6*, 290–306. [CrossRef]
22. Conroy, J.D.; Boegman, L.; Zhang, H.; Edwards, W.J.; Culver, D.A. "Dead Zone" dynamics in Lake Erie: The importance of weather and sampling intensity for calculated hypolimnetic oxygen depletion rates. *Aquat. Sci.* **2011**, *73*, 289–304. [CrossRef]
23. Bocaniov, S.A.; Scavia, D. Temporal and spatial dynamics of large lake hypoxia: Integrating statistical and three-dimensional dynamic models to enhance lake management criteria. *Water Resour. Res.* **2016**, *52*, 4247–4263. [CrossRef]
24. Kisi, O.; Shiri, J.; Nikoofar, B. Forecasting daily lake levels using artificial intelligence approaches. *Comput. Geosci.* **2012**, *41*, 169–180. [CrossRef]
25. Li, W. Mutual information functions versus correlation functions. *J. Stat. Phys.* **1990**, *60*, 823–837. [CrossRef]
26. Ross, B.C. Mutual information between discrete and continuous data sets. *PLoS ONE* **2014**, *9*, e87357. [CrossRef] [PubMed]
27. Grbić, R.; Slišković, D.; Kadlec, P. Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models. *Comput Chem Eng.* **2013**, *58*, 84–97. [CrossRef]
28. Rasmussen, C.E. Gaussian processes in machine learning. In Proceedings of the Summer School on Machine Learning, Canberra, Australia, 2–14 February 2003; Springer: Berlin/Heidelberg, Germany, 2003.
29. Sun, A.Y.; Wang, D.; Xu, X. Monthly streamflow forecasting using Gaussian process regression. *J. Hydrol.* **2014**, *511*, 72–81. [CrossRef]
30. Makridakis, S.; Wheelwright, S.C.; Hyndman, R.J. *Forecasting Methods and Applications*; John Wiley & Sons: Toronto, ON, Canada, 2008.
31. Piasecki, A.; Jurasz, J.; Skowron, R. Forecasting surface water level fluctuations of lake Serwy (Northeastern Poland) by artificial neural networks and multiple linear regression. *J. Environ. Eng. Landsc. Manag.* **2017**, *25*, 379–388. [CrossRef]
32. Kadam, A.K.; Wagh, V.M.; Muley, A.A.; Umrikar, B.N.; Sankhua, R.N. Environment. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Model Earth Syst. Environ.* **2019**, *5*, 951–962. [CrossRef]
33. Singh, A.; Imtiyaz, M.; Isaac, R.K.; Denis, D.M. Comparison of soil and water assessment tool (SWAT) and multilayer perceptron (MLP) artificial neural network for predicting sediment yield in the Nagwa agricultural watershed in Jharkhand, India. *Agric. Water Manag.* **2012**, *104*, 113–120. [CrossRef]
34. Hertz, J.A. *Introduction to the Theory of Neural Computation*; CRC Press: Boca Raton, FL, USA, 2018.

35. Lekkas, D.F.; Onof, C.; Lee, M.J.; Baltas, E.A. Application of artificial neural networks for flood forecasting. *Glob. Nest J.* **2004**, *6*, 205–211.
36. Ghorbani, M.A.; Deo, R.C.; Karimi, V.; Kashani, M.H.; Ghorbani, S. Design and implementation of a hybrid MLP-GSA model with multi-layer perceptron-gravitational search algorithm for monthly lake water level forecasting. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 125–147. [CrossRef]
37. Quinlan, J.R. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Tasmania, 16–18 November 1992; World Scientific: Singapore, 1992.
38. Solomatine, D.P.; Siek, M.B.L. Flexible and optimal M5 model trees with applications to flow predictions. In Proceedings of the 6th International Conference on Hydroinformatics, Singapore, 21–24 June 2004; World Scientific: Singapore, 2004.
39. Solomatine, D.P.; Dulal, K.N. Model trees as an alternative to neural networks in rainfall—Runoff modelling. *Hydrol. Sci. J.* **2003**, *48*, 399–411. [CrossRef]
40. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
41. Wang, X.; Liu, T.; Zheng, X.; Peng, H.; Xin, J.; Zhang, B. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Appl. Water Sci.* **2018**, *8*, 125. [CrossRef]
42. Bremner, D.; Demaine, E.; Erickson, J.; Iacono, J.; Langerman, S.; Morin, P.; Toussaint, G. Output-sensitive algorithms for computing nearest-neighbour decision boundaries. *Discrete. Comput. Geom.* **2005**, *33*, 593–604. [CrossRef]
43. Atkeson, C.G.; Moore, A.W.; Schaal, S. *Locally Weighted Learning*; Springer: Dordrecht, The Netherlands, 1997; pp. 11–73.
44. Poul, A.K.; Shourian, M.; Ebrahimi, H. A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. *Water Resour. Manag.* **2019**, *33*, 2907–2923. [CrossRef]
45. Kinney, J.B.; Atwal, G.S. Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3354–3359. [CrossRef]
46. Choi, C.; Kim, J.; Han, H.; Han, D.; Kim, H.S. Development of Water Level Prediction Models Using Machine Learning in Wetlands: A Case Study of Upo Wetland in South Korea. *Water* **2020**, *12*, 93. [CrossRef]
47. Goodfellow, I.; Bengio, Y.; Courville, A. Machine learning basics. In *Deep Learning*; Kaiser, D.I., Weck, O.L.D., Eds.; MIT Press: Cambridge, MA, USA, 2016; Volume 1, pp. 98–164.
48. Bazartseren, B.; Hildebrandt, G.; Holz, K.P. Short-term water level prediction using neural networks and neuro-fuzzy approach. *Neurocomputing* **2003**, *55*, 439–450. [CrossRef]
49. Yoon, H.; Jun, S.C.; Hyun, Y.; Bae, G.O.; Lee, K.K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [CrossRef]
50. Nadkarni, P. *Clinical Research Computing: A Practitioner's Handbook*; Academic Press: Cambridge, MA, USA, 2016.
51. Croley II, T. Using climate predictions in Great Lakes hydrologic forecasts. In *Climate Variations, Climate Change, and Water Resources Engineering*; Garbrecht, J.D., Piechota, T.C., Eds.; American Society of Civil Engineers: Arlington, VA, USA, 2006; pp. 166–187.
52. Croley II, T.E.; Hartmann, H.C. Resolving thiessen polygons. *J. Hydrol.* **1985**, *76*, 363–379. [CrossRef]
53. Deo, R.C.; Şahin, M. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. *Environ. Monit. Assess.* **2016**, *188*, 90. [CrossRef]