Environmental Earth Sciences (2019) 78:202
https://doi.org/10.1007/s12665-019-8202-7

ORIGINAL ARTICLE

Extreme learning machine-based prediction of daily water temperature for rivers

Senlin Zhu[1] · Salim Heddam[2] · Shiqiang Wu[1] · Jiangyu Dai[1] · Benyou Jia[1]

Received: 30 August 2018 / Accepted: 1 March 2019 / Published online: 9 March 2019

Abstract

Water temperature impacts many processes in rivers, and it is determined by various environmental factors. This study proposed an extreme learning machine (ELM)-based model to predict daily water temperature for rivers. Air temperature (T_a), discharge (Q) and the day of the year (DOY) were used as predictors. Three rivers characterized by different hydrological conditions were investigated to test the modeling performances and the model results were compared with multilayer perceptron neural network (MLPNN) and simple multiple linear regression (MLR) models. Results showed that inclusion of three inputs as predictors (T_a, Q and the DOY) yielded the best modeling accuracy for all the developed models. It was also found that Q played a minor role and T_a and DOY are the most important explanatory variables for river water temperature predictions. Additionally, sigmoidal and radial basis activation functions within the ELM model performed the best for river water temperature forecasting. ELM and MLPNN models outperformed MLR model, and ELM model with sigmoidal and radial basis activation functions performed comparably to MLPNN model. Overall, results indicated that the ELM model developed in this study can be effectively used for river water temperature predictions.

Keywords River water temperature · Air temperature · Discharge · Extreme learning machine · Artificial neural network

Introduction

Water temperature (T_w) is one of the most important water quality variables and has received great attention during the last decades (Du et al. 2019). Beyond its either direct or indirect influence on several physical, chemical, and biological processes (Williamson et al. 2018), the variation of T_w directly affects the overall health of aquatic ecosystems. Because of its fundamental importance in freshwater ecosystems, the accurate estimation of T_w is necessary for controlling and monitoring river and stream water quality, and the quantification of warming rates of river is directly related to river temperature monitoring (Isaak et al. 2018; Pohle et al. 2018). In a recently published paper, Kim et al. (2018a, b) has demonstrated that T_w significantly affects the diatom growth in a freshwater ecosystem and the diatom concentration was strongly affected by T_w rather than the phosphorus concentration. Cha et al. (2017) found that cyanobacteria dominance is strongly correlated to T_w and warmer and stratified water have been considered highly favorable for reproducing cyanobacteria. Additionally, accurate estimates of T_w are important for modeling of dissolved oxygen (DO) and a strong negative correlation was found between T_w and DO (Heddam and Kisi 2018; Šiljić Tomić et al. 2018a, b).

Spatial and temporal variability of T_w are linked to several factors: (i) riparian vegetation (Garner et al. 2017), (ii) heat exchange between atmosphere and water column (Garner et al. 2017), (iii) climate and precipitation change, and (iv) energy transport processes in river and anthropogenic activities (Kędra and Wiejaczka 2018). Mathematical models to estimate river T_w from environmental factors have either been based on the use of empirical and statistical approaches, which relate T_w to several other variables (e.g., air temperature and flow discharge) or physically based deterministic models. Physically based deterministic models and statistical models have been proposed and successfully

✉ Senlin Zhu
slzhu@nhri.cn

[1] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China

[2] Hydraulics Division, Faculty of Science, Agronomy Department, Laboratory of Research in Biodiversity Interaction, Ecosystem and Biotechnology, University 20 Août 1955, Route El Hadaik, BP 26, Skikda, Algeria

applied for river T_w predictions in the past decades. Both models have their advantages and disadvantages and their potentials have been demonstrated in numerous studies (Benyahya et al. 2007; Marcé and Armengol 2010; Cole et al. 2014; Kwak et al. 2017).

Many kinds of mathematical models have been successfully implemented to model river water temperature. Typical examples are linear and non-linear regression models (Van Vliet et al. 2012; Krider et al. 2013), stochastic regression models (Ahmadi-Nedushan et al. 2007; Rabi et al. 2015), and hybrid statistical–physical-based models (Gallice et al. 2015; Toffolon and Piccolroaz 2015; Piccolroaz et al. 2016). In addition, with the development of artificial intelligence, machine learning models were gradually applied in river water temperature predictions. Artificial neural network (ANN), as one of the typical machine learning models, has been widely used in river water temperature predictions due to its abilities to efficiently solve non-linear and multivariate problems (Sahoo et al. 2009; Hadzima-Nyarko et al. 2014; DeWeber and Wagner 2014; Piotrowski et al. 2014, 2015; Temizyurek and Dadasercelik 2018; Zhu et al. 2018a, b).

Huang et al. (2006a, b) introduced the extreme learning machine (ELM) method, since ELM is capable of demonstrating a good modeling performance and speeding up the learning process (Huang et al. 2006a, b; Liang et al. 2006). In this regard, ELM model has been widely used in hydrology, such as stream flow forecasting (Yaseen et al. 2016; Rezaie-Balf and Kisi 2017), evaporative loss estimation (Deo et al. 2016), and water quality prediction (Heddam and Kisi 2017). However, it has never been used for river water temperature predictions. Despite the fact that various models have been developed to model river water temperature in recent years, modeling river water temperature using new robust techniques to offer increased accuracy and reliability in the predictions is still highly demanding owing to the significance of accessibility to the river water temperature in various scientific fields. Air temperature and discharge are generally the most available variables for modeling temperatures in rivers, and they have been shown to have the greatest impact on water temperature (Webb et al. 2003; Van Vliet et al. 2011). However, the role of discharge has not been comprehensively investigated for various river systems characterized by different hydrological regimes, and most of the previous studies employed only atmospheric factors as predictors (Sahoo et al. 2009; Marcé and Armengol 2010; Hadzima-Nyarko et al. 2014; Rabi et al. 2015; Zhu et al. 2018a; Temizyurek and Dadasercelik 2018). Therefore, the stimulus for this research is to develop ELM models for various river systems characterized by different hydrological conditions with air temperature (T_a), flow discharge (Q) and the day of the year (DOY) as predictors. The aim of this study is to test whether the widely used ELM model can be used for T_w forecasting by comparing modeling performances with traditional ANN models. The novelty of this study is the assessment of the capability of ELM models and testing of different activation functions for T_w forecasting, which has never been reported. The results in this study may contribute to river water temperature modeling and provide references for water resources management.

Materials and methods

Case studies

In this study, the performances of the proposed ELM models were tested using data from three rivers in US, which are characterized by different hydrological conditions. The rivers are briefly introduced below and summarized in Table 1, where the periods of data availability are also listed. The three rivers are: (i) the Cedar River is located in the east part of Seattle, Washington State, US. It is about 72 km in length. Its upper watershed is a protected area, which is mountainous, wilder and cold, and provides drinking water for the greater Seattle area. (ii) The Fanno Creek is located near Portland, Oregon, US. It is about 24 km in length. It is mainly hilly, and covered by agriculture and residential areas. (iii) The Irondequoit Creek is located at lowland areas near Rochester, New York, US. Its catchment is covered by agriculture and residential areas. Detailed information about these three rivers can be found in Piotrowski and Napiorkowski (2018). Figure 1 presents the time series of T_w, T_a and Q for the three rivers. It is shown that T_a and T_w present strong seasonal pattern with larger values in summer and lower values in winter, and Q mainly distributed in spring and winter periods.

Table 1 Characteristics of river and meteorological stations

River Name	Cedar	Fanno	Irondequoit
River station name/number	USGS 12119000	USGS 14206950	USGS 0423205010
Catchment area (km^2)	477	82	368
Calibration period	2001–2012	2003–2012	2005–2013
Validation period	2013–2017	2013–2017	2014–2017
Meteorological station name	USW00024233	USW00024229	USW00014768
Distance from river station (km)	9.4	24.1	13.7

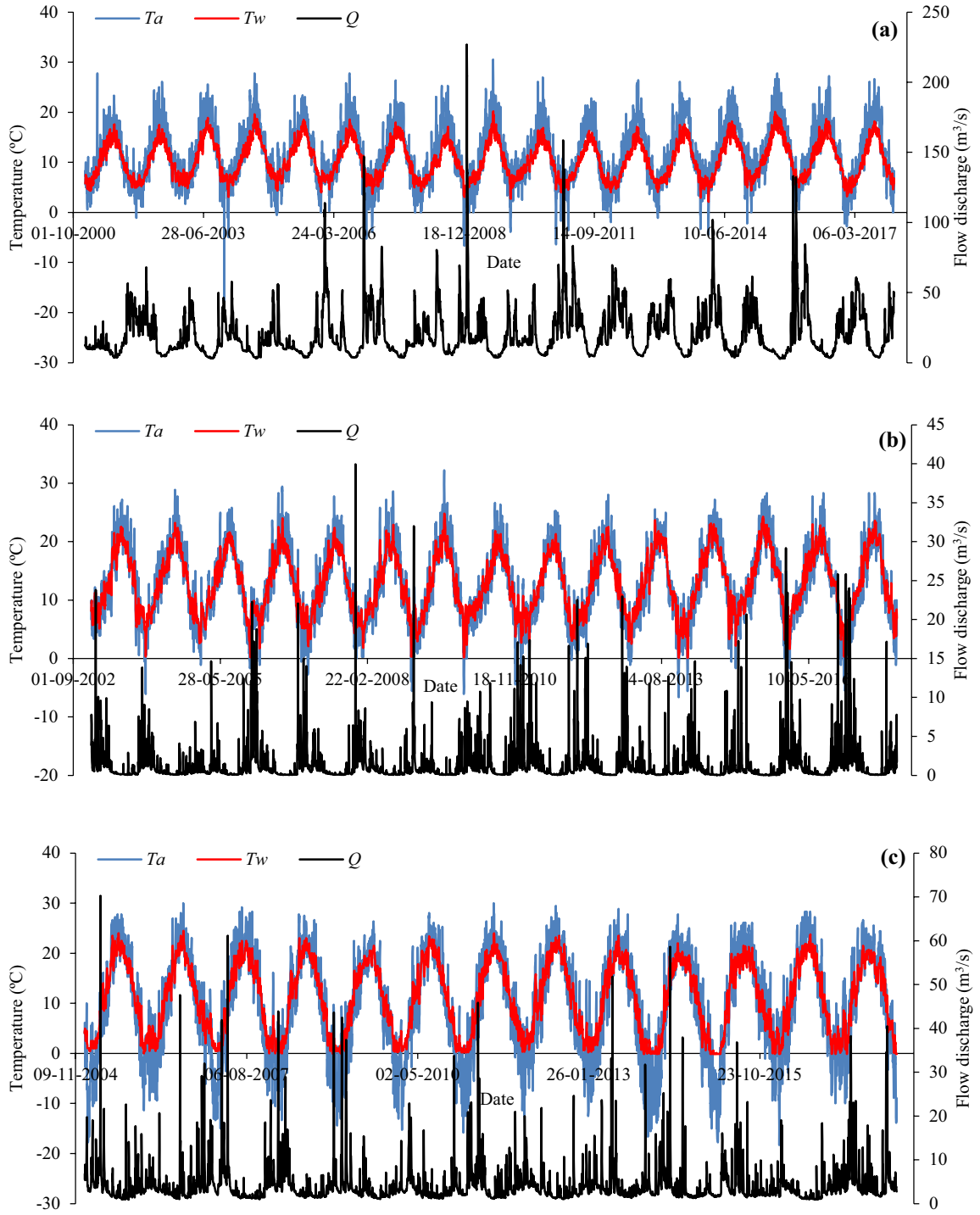

Fig. 1 Time series of air temperature (T_a), water temperature (T_w) and flow discharge (Q) for the three rivers: **a** Cedar, **b** Fanno, and **c** Irondequoit

By analysis of the time series data using Pearson correlation coefficients (R^2), it was found that T_w correlates well with T_a with R^2 equaling 0.878, 0.915 and 0.904 for Cedar, Fanno and Irondequoit, respectively. However, T_w shows no apparent correlation with Q.

Multilayer perceptron neural network (MLPNN)

McCulloch and Pitts (1943) proposed the first mathematical neuron mainly inspired from the function of the biological neuron. By definition, an ANN model is a system composed

of interconnected nodes, and each node is responsible to perform linear and nonlinear operations (Atkinson and Tatnall 1997). The ANN models were proposed to solve nonlinear problems (Ghiassi and Nangoy 2009) and the most well-known kind of ANN is the multilayer perceptron neural network (MLPNN) architecture reported in the literature as universal approximators (Hornik et al. 1989; Hornik 1991). ANN has been widely applied in hydrology, such as modeling of the rainfall–runoff process (Karami and Dariane 2017), stream flow forecasting (Humphrey et al. 2016), groundwater management (Trichakis et al. 2011), and water quality simulation (Chaves and Kojiri 2007).

The basic architecture of the MLPNN model used in the present investigation is shown in Fig. 2. It consists of three layers: input, hidden and output layers, connected by an ensemble of weights (W_{ij} and W_{jk}) responsible for information transmission from the input to the output layer. The other components of this architecture are the biases (B_j and B_0) allocated to each hidden and output neuron. Using the MLPNN model, the available information stored in the input layer is passed to the output layer via a series of successive mathematical operation: multiplication, summation and transformation using an activation function. The MLPNN model uses a learning algorithm to minimize the predictive error between the desired and the calculated value of T_w. The learning algorithm is an iterative process, repeated several times until the model meets the desired criteria by maximizing the predictive generalization. The hidden neuron represents an important part of any MLPNN model, because they operate double tasks: each neuron in the hidden layer receiving the entire incoming input variable is multiplied by its weight (W_{ij}) and adding bias terms (B_j) to the results. In a second stage, the result is passed to the next layer via an activation function, generally the sigmoidal activation function (f_1) as adopted in the present study. The final result is calculated as a summation of the output of each hidden neuron multiplied by its weights (W_{jk}). During the supervised training phase, the patterns is composed of the three inputs (T_a, Q and DOY) and the desired output (T_w) is presented to the model to calculate a value of T_w, by making a link between the input variables (T_a, Q and DOY) and the output variable (T_w) via the optimal set of weights and biases. By several iterations, the desired and the calculated values of T_w are becoming increasingly closer and the mean squared error decreases significantly and close to zero. During the training process, the entire set of weights and biases (W_{ij}, W_{jk}, B_j and B_0) was updated using the backpropagation algorithm. Finally, the generalization capability of the model is checked during the validation stage, using a set of data not included during the training phase. More details about the ANN models can be found in Haykin (1999). The output of the MLPNN model is calculated as follows:

The input of each hidden neuron (H_i) is given by Eq. (1):

$$I_j = \sum_{j=1}^{n} x_i w_{ij} + B_j \tag{1}$$

where x_i is the input variable (T_a, Q and DOY), w_{ij} is the weight between the input i and the hidden neuron j, and B_j is the bias of the hidden neuron j. The result of Eq. (1) is used in a nonlinear sigmoidal activation function in Eq. (2):

$$E_j = f_1(I_j) = \frac{1}{1 + e^{-I_j}} \tag{2}$$

The outputs of the hidden neurons are used as input to the output layer:

$$O = \sum_{j=1}^{n} E_j w_{jk} + B_0 \tag{3}$$

where w_{jk} is the weight of connection of neuron j in the hidden layer to unique neuron in the output layer; O is the input of the unique output neuron, and B_0 is the bias of the output neuron. Finally, the output of the neuron k in the output layer is calculated using a linear activation.

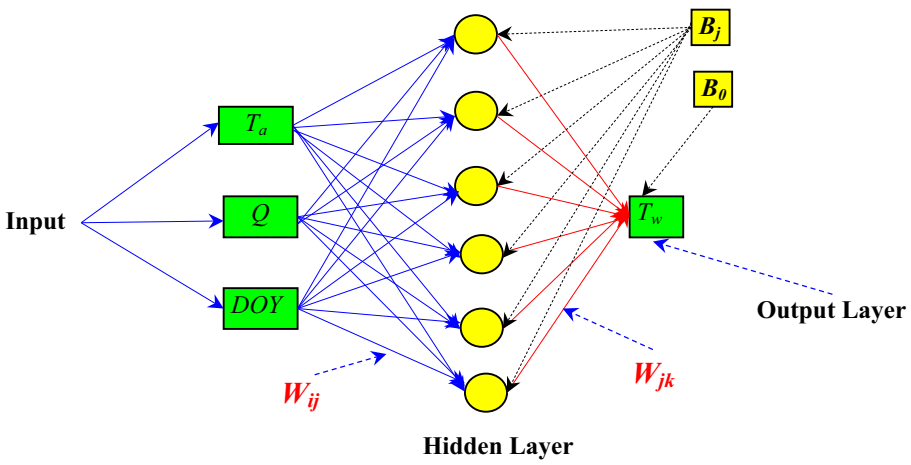

Fig. 2 Multilayer perceptron neural network (MLPNN) architecture (T_a air temperature; Q flow discharge; DOY the day of the year; T_w water temperature)

Extreme learning machine (ELM)

The extreme learning machine (ELM) algorithm was proposed by Huang et al. (2006a, b) as a training algorithm for the single hidden layer feedforward network (SLFN). For the standard ANN described in the previous section, all models parameters (weights and biases) from the input to the output layers were updated during the training process. In some cases, especially when dealing with a large dataset, the training process becomes a hard task, very time consuming and does not converge to the best optimal set of parameters, and consequently the generalization capacity of the model can be affected. Huang et al. (2006a, b) proposed a reformulation of the training process for which the input weights from the input layer to the hidden layer were randomly chosen and the second set of weights between the hidden layer and the output layer are calculated analytically by the simple generalized inverse operation. ELM has several advantages: faster learning, high generalization ability and a small number of epochs (Huang et al. 2015). Many activation functions such as sigmoidal, sine, radial basis function and hard-limiting function can be used, and the output neurons have linear activation function (Huang et al. 2006a, b). In the following, a brief description of the ELM model was provided.

Assuming that the architecture of the SLFN is composed of L hidden neurons, one input layer; each hidden neuron uses an activation function $g(x)$ and a unique neuron in the output layer. All neurons have a bias term (B). The SLFN is trained using a dataset composed by N arbitrary patterns (X_i, Y_i), where $x_i = [x_{i1}, x_{i2},\ldots, x_{iD}]^T \in R^D$ and $[y_{i1}, y_{i2},\ldots, y_{iD}]^T \in R^D$. SLFN is mathematically presented as (Huang et al. 2006a, b):

$$\sum_{i=1}^{L} \beta_i g_i\left(x_j\right) = \sum_{i=1}^{L} \beta_i g_i\left(w_i \cdot x_j + B_i\right) = o_k \tag{4}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector connecting the ith hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the ith hidden neuron and the output neurons, $o_k = [o_{k1}, o_{k2}, \ldots, o_{km}]^T$ is the kth output vector of the SLFN, and B_i is the bias of the ith hidden neuron. g_i (.) is the output of the jth hidden neuron, and g is the activation function of the hidden neuron (Huang et al. 2006a, b, 2015; Huang and Chen 2008). The nonlinear Eq. (4) for a SLFN can be written in matrix form as:

$$H\beta = T \tag{5}$$

where:

$$H = \begin{bmatrix} g_1\left(w_1 \cdot x_1 + b_1\right) & \cdots & g_L\left(w_L \cdot x_1 + b_L\right) \\ \vdots & \vdots & \vdots \\ g_1\left(w_1 \cdot x_N + b_1\right) & \cdots & g_L\left(w_L \cdot x_N + b_L\right) \end{bmatrix} \tag{6}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix} \tag{7}$$

$$\mathrm{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix} \tag{8}$$

where H is the hidden layer output matrix of the network.

The output weights between the hidden layer and the output layer are given by:

$$\hat{\beta} = H^+ T \tag{9}$$

where H^+ is the Moore–Penrose generalized inverse of the hidden layer output matrix H (Huang et al. 2015). According to Huang et al. (2006a, b), ELM is developed in three steps: (i) randomly generating hidden node parameters w_i and bias β_i depending on the number of hidden neurons and the number of input variables based on continuous probability density function (Vasu and Lee 2016), (ii) calculating the hidden layer output matrix H according to Eq. (6) and (iii) calculating the output weight $\hat{\beta}$ using the Moore–Penrose generalized inversion method.

Multiple linear regression (MLR)

In the present investigation, the multiple linear regression (MLR) is selected as a benchmarking model, to demonstrate the need and the utility of using ELM and MLPNN nonlinear models. MLR model linking the Q, T_a and DOY to T_w is presented as follows:

$$T_w = \alpha_0 + \left(\alpha_1 \times T_a\right) + \left(\alpha_2 \times Q\right) + \left(\alpha_3 \times \mathrm{DOY}\right) \tag{10}$$

where α_0 is the intercept and α_1, α_2 and α_3 are the regression coefficients.

Model evaluation

Several statistical indicators, such as the coefficient of correlation (R), the Willmott index of agreement (d), the root mean squared error (RMSE), and the mean absolute error (MAE) are used for evaluating the performances of different models (Deo et al. 2018; Ghorbani et al. 2019).

$$R = \left[\frac{\frac{1}{N}\sum\left(O_i - O_m\right)\left(P_i - P_m\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(O_i - O_m\right)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(P_i - P_m\right)^2}} \right] \tag{11}$$

$$d = 1 - \frac{\sum_{i=1}^{N}\left(P_i - O_i\right)}{\sum_{i=1}^{N}\left(\left|P_i - O_m\right| + \left|O_i - O_m\right|\right)^2} \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(O_i - P_i\right)^2} \quad (13)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|O_i - P_i\right| \quad (14)$$

where N is the number of data points, O_i is the measured and P_i is the corresponding model prediction of water temperatures. O_m and P_m are the average values of O_i and P_i.

Results and discussion

Daily water temperatures (T_w) at the three reference stations were predicted using the ELM, MLPNN and MLR models according to the following three different scenarios: (i) scenario 1 (ELM1, MLPNN1 and MLR1) using only the air temperature (T_a) as input variable, (ii) scenario 2 (ELM2, MLPNN2 and MLR2) using two input variables, flow discharge and air temperature (Q and T_a) and (iii) scenario 3 (ELM3, MLPNN3 and MLR3) with three input variables (T_a, Q and DOY). The MLPNN models were applied using the sigmoidal activation function for the hidden neurons and the identity (linear) function for the output neuron. Additionally, for the ELM models, and for each scenario described above, different activation functions, including sigmoidal (SIG), sine (SIN), radial basis function (RAD), Hardlim (HAL), and triangular function (TRI) were compared. The estimated values of the performance indices in the training and validation phases are shown in Tables 2, 3 and 4, respectively. According to the results obtained, several conclusions can be drawn. First, the comparative results between the three input combinations, from one input to three input variables, reveal that the best accuracy was obtained by the ELM3 and MLPNN3 at the three stations. Hence, the addition of the DOY variable to the T_a and Q consistently increased the accuracy of both ELM and MLPNN models. ELM3 and MLPNN3 yielded the best accuracy among all the developed models, and outperformed all the other two models in term of higher R, d, and lower RMSE and MAE, at the three river stations. Second, regarding the importance of the input variables, it is clear from the obtained results that inclusion of the Q variable does not necessarily lead to better model performance; in some case the performances of the models were slightly decreased. For example, at Irondequoit station all the ELM1 models with all five activation functions outperform slightly the ELM2 models. Similarly, at the same station, MLPNN1 outperforms slightly the MLPNN2 model. Third, globally speaking, it can be observed from

Table 2 Performances of different models in modeling water temperature (T_w °C) for Cedar station

Activation function	Model version	Training (calibration)				Validation			
		R	*d*	RMSE (°C)	MAE (°C)	*R*	*d*	RMSE (°C)	MAE (°C)
SIG	ELM3	0.986	0.993	0.588	0.463	0.981	0.987	0.977	0.735
	ELM2	0.966	0.982	0.925	0.719	0.957	0.973	1.387	1.010
	ELM1	0.945	0.971	1.168	0.906	0.952	0.975	1.211	0.933
SIN	ELM3	0.986	0.993	0.592	0.468	0.957	0.973	1.415	0.777
	ELM2	0.966	0.982	0.926	0.719	0.958	0.975	1.318	0.961
	ELM1	0.945	0.971	1.169	0.906	0.952	0.975	1.211	0.932
RAD	ELM3	0.986	0.993	0.587	0.462	0.983	0.988	0.861	0.670
	ELM2	0.966	0.982	0.924	0.717	0.962	0.977	1.262	0.950
	ELM1	0.945	0.971	1.165	0.906	0.952	0.975	1.213	0.934
HAL	ELM3	0.974	0.987	0.813	0.634	0.977	0.988	0.914	0.713
	ELM2	0.963	0.981	0.968	0.748	0.966	0.982	1.071	0.833
	ELM1	0.945	0.971	1.166	0.905	0.952	0.975	1.220	0.943
TRI	ELM3	0.981	0.991	0.687	0.532	0.969	0.982	1.099	0.727
	ELM2	0.965	0.982	0.936	0.727	0.969	0.983	1.068	0.828
	ELM1	0.947	0.972	1.153	0.898	0.947	0.972	1.277	0.983
SIG	MLPNN3	0.985	0.992	0.618	0.488	0.983	0.991	0.768	0.593
	MLPNN2	0.963	0.981	0.961	0.745	0.970	0.984	1.019	0.801
	MLPNN1	0.945	0.971	1.168	0.905	0.952	0.975	1.211	0.933
–	MLR3	0.951	0.974	1.109	0.843	0.963	0.978	1.105	0.755
–	MLR2	0.947	0.972	1.153	0.876	0.958	0.976	1.157	0.923
–	MLR1	0.934	0.965	1.274	0.984	0.943	0.967	1.344	1.071

Table 3 Performances of different models in modeling water temperature (T_w °C) for Fanno station

Activation function	Model version	Training (calibration)				Validation			
		R	*d*	RMSE (°C)	MAE (°C)	*R*	*d*	RMSE (°C)	MAE (°C)
SIG	ELM3	0.981	0.991	0.971	0.759	0.981	0.991	1.078	0.759
	ELM2	0.959	0.979	1.441	1.138	0.969	0.985	1.356	1.077
	ELM1	0.957	0.978	1.461	1.157	0.969	0.984	1.375	1.096
SIN	ELM3	0.969	0.984	1.242	0.980	0.979	0.990	1.125	0.885
	ELM2	0.958	0.979	1.445	1.144	0.969	0.985	1.356	1.082
	ELM1	0.957	0.978	1.461	1.158	0.969	0.984	1.375	1.096
RAD	ELM3	0.981	0.990	0.981	0.765	0.986	0.993	0.928	0.735
	ELM2	0.958	0.978	1.449	1.143	0.968	0.984	1.385	1.096
	ELM1	0.958	0.978	1.460	1.154	0.968	0.984	1.381	1.100
HAL	ELM3	0.972	0.986	1.181	0.914	0.978	0.990	1.143	0.892
	ELM2	0.958	0.979	1.446	1.141	0.967	0.984	1.404	1.109
	ELM1	0.957	0.978	1.665	1.160	0.968	0.984	1.392	1.110
TRI	ELM3	0.978	0.989	1.048	0.823	0.982	0.991	1.041	0.819
	ELM2	0.960	0.980	1.411	1.111	0.969	0.985	1.358	1.082
	ELM1	0.958	0.979	1.445	1.141	0.966	0.983	1.439	1.151
SIG	MLPNN3	0.982	0.991	0.962	0.751	0.985	0.993	0.948	0.754
	MLPNN2	0.960	0.979	1.417	1.112	0.970	0.985	1.355	1.076
	MLPNN1	0.958	0.978	1.458	1.153	0.968	0.984	1.383	1.101
–	MLR3	0.954	0.976	1.516	1.212	0.966	0.978	1.566	1.277
–	MLR2	0.954	0.976	1.518	1.213	0.966	0.978	1.564	1.277
–	MLR1	0.954	0.976	1.523	1.218	0.965	0.977	1.575	1.285

Table 4 Performances of different models in modeling water temperature (T_w °C) for Irondequoit station

Activation function	Model version	Training (calibration)				Validation			
		R	*d*	RMSE (°C)	MAE (°C)	*R*	*d*	RMSE (°C)	MAE (°C)
SIG	ELM3	0.989	0.994	1.067	0.830	0.986	0.993	1.213	0.919
	ELM2	0.963	0.981	1.930	1.444	0.964	0.981	1.910	1.419
	ELM1	0.960	0.980	1.994	1.486	0.966	0.982	1.861	1.385
SIN	ELM3	0.979	0.988	1.455	1.112	0.983	0.991	1.326	1.022
	ELM2	0.962	0.980	1.967	1.482	0.963	0.980	1.951	1.456
	ELM1	0.960	0.979	1.995	1.486	0.966	0.982	1.860	1.385
RAD	ELM3	0.989	0.994	1.075	0.838	0.985	0.992	1.241	0.950
	ELM2	0.962	0.980	1.956	1.456	0.964	0.981	1.914	1.401
	ELM1	0.961	0.980	1.992	1.487	0.966	0.982	1.863	1.386
HAL	ELM3	0.978	0.989	1.480	1.137	0.973	0.986	1.677	1.269
	ELM2	0.962	0.980	1.970	1.470	0.964	0.981	1.942	1.415
	ELM1	0.961	0.980	1.972	1.474	0.965	0.982	1.893	1.409
TRI	ELM3	0.984	0.992	1.293	0.971	0.981	0.990	1.416	1.066
	ELM2	0.964	0.982	1.894	1.415	0.963	0.980	1.955	1.435
	ELM1	0.960	0.979	1.995	1.488	0.966	0.982	1.862	1.386
SIG	MLPNN3	0.989	0.994	1.078	0.840	0.985	0.992	1.242	0.933
	MLPNN2	0.964	0.981	1.900	1.420	0.963	0.981	1.943	1.419
	MLPNN1	0.961	0.980	1.983	1.480	0.966	0.982	1.867	1.387
–	MLR3	0.953	0.975	2.167	1.686	0.951	0.974	2.270	1.713
–	MLR2	0.953	0.975	2.168	1.686	0.951	0.974	2.271	1.713
–	MLR1	0.950	0.974	2.226	1.730	0.952	0.975	2.253	1.693

the obtained results that ELM3 provided the best accuracy at the Fanno and Irondequoit stations, while the MLPNN3 is the best accurate model at the Cedar station. The following sections discuss the results of the ELM and MLPNN methods for water temperature estimation at the three selected stations. Additionally, the two methods with three different input combinations were analyzed to determine the optimal approach and compared to the simple MLR approach as a benchmarking model.

At Cedar station (Table 2), for the two modeling approaches (ELM and MLPNN), it can be concluded that, both during the training and the validation phase, results showed that the selection of the DOY as input variable has a marked effect on the performances of the models. During the training phase, it is clear from the obtained results that, when air temperature (T_a) is used individually in the models, the performances of both ELM1 and MLPNN1 models are very similar. An analysis of the statistical indices shows that the R and d values are in the range of 0.945–0.947, and 0.971–0.972. Similarly, the RMSE and MAE range from 1.153 (°C) to 1.169 (°C), and from 0.898 (°C) to 0.906 (°C), respectively. In the training phase, based on the metric in Table 2, it is expected that using a combination of input variables improves the performances of the models. For both ELM2 and MLPNN2, the performances of the models are better compared to the ELM1 and MLPNN1, due to the inclusion of Q as input variable. Using T_a and Q as input variables, the best ELM2 model using the radial basis function as activation function (RAD) had RMSE = 0.924 (°C), MAE = 0.717(°C), $R = 0.966$ and $d = 0.982$, while the MLPNN2 model showed marginally higher RMSE (0.961 °C) and MAE values (0.745 °C), and R and d values slightly less than ELM2 model ($R = 0.963$, $d = 0.981$). Finally, for the ELM3 and MLPNN3 models using the three input variables as inputs, the increase in the performances with increasing in R and d values is much clearer than both ELM2 and MLPNN2. For the ELM3, the best accuracy was obtained using three activation functions (sigmoidal, sin and radial basis) with relatively similar accuracy. Using the DOY as input, the R and d values of the ELM3 models have increased by 0.02 and 0.011, whereas the RMSE and MAE are reduced by 0.337 and 0.255 °C, respectively. In addition, using the DOY as input, the R and d values of the ELM3 models have increased by 0.013 and 0.007, whereas the RMSE and MAE are reduced by 0.251 and 0.208 °C, respectively. Also, it is clear that MLPNN3 and ELM3 with radial basis function were similar in accordance with the four statistical indices and the differences are very marginal.

The validation results are shown in Table 2. The scatterplot of measured and calculated water temperature (T_w °C) is shown in Fig. 3. According to Table 2, inclusion of the DOY variable as inputs resulted in overall higher accuracy. The best accuracy with R and d values equal to 0.983 and 0.988 was achieved using ELM3 with radial basis function, whereas the MLPNN3 model provided R and d values with 0.983 and 0.991, respectively. Although the R and d values were relatively similar with negligible difference, it is clear from Table 2 that the MLPNN3 model decreased the RMSE and MAE of the ELM3-RAD by 10.80% and 11.49%, respectively. The benefit of using the DOY as input variable to reduce the RMSE and MAE was more pronounced when comparing the models with and without it, especially using the ELM3 model. MLPNN3 decreased the RMSE and MAE of the MLPNN2 by 24.63% and 25.96%, respectively. In addition, using the DOY as input variable, the ELM3 model decreased the RMSE and MAE of the ELM2 by 31.77% and 29.47%, respectively. Regarding the five ELM3 models, it is clear that the lowest accuracy was achieved using the ELM3 with RAD function. Finally, the results achieved using the simple MLR model are also presented in Table 2. According to Table 2, the best accuracy was obtained using the MLR3 slightly and marginally better than the MLR2. In addition, the MLR3 model provided less accuracy compared to the MLPNN3 and possesses the lowest accuracy compared to all the ELM models except the ELM3 with sine (SIN) activation function, for which the MLR3 is slightly better.

Figures 4 and 5 show the box plots and the violin plots for measured and calculated water temperature (T_w °C) values using the MLPNN3, MLR3 and the five ELM3 models for Cedar station. Each box plot displays the median that correspond to the central mark, the 25th and 75th percentiles, and the "whiskers" of the box plot extend to minimum and maximum water temperature (T_w °C) values excluding outliers. Outliers are depicted with red crosses. For the violin plot, the two lines with black and red color display the mean and the median values of T_w. According to Fig. 4, ELM3-SIN with sine activation function possesses the lowest accuracy and is unable to correctly capture the outlier's values, in addition the minimum and the maximum values calculated using the ELM3-SIN are more superior to the measured values. There is a clear resemblance between the measured T_w and the calculated values using the ELM3-SIG and ELM3-RAD computed based on the entire median, the 25th and 75th percentiles and the whiskers. Regarding the violin plot reported in Fig. 5, the high resemblance between measured and calculated T_w values was achieved using MLPNN3, ELM3-RAD and ELM3-SIG, and the MLR3 and ELM3-SIN performed worse for T_w modeling.

At Fanno station (Table 3), in the training phase, the best accuracy was obtained with the models using the three input variables (T_a, Q and DOY). MLPNN3 showed relatively higher R (0.982) and d (0.991), lower RMSE (0.962 °C) and MAE (0.751 °C) than ELM3 models. The best accuracy was obtained using the ELM3 with sigmoidal activation function, marginally higher than the ELM3 with radial basis, Hardlim and triangular activation functions. However,

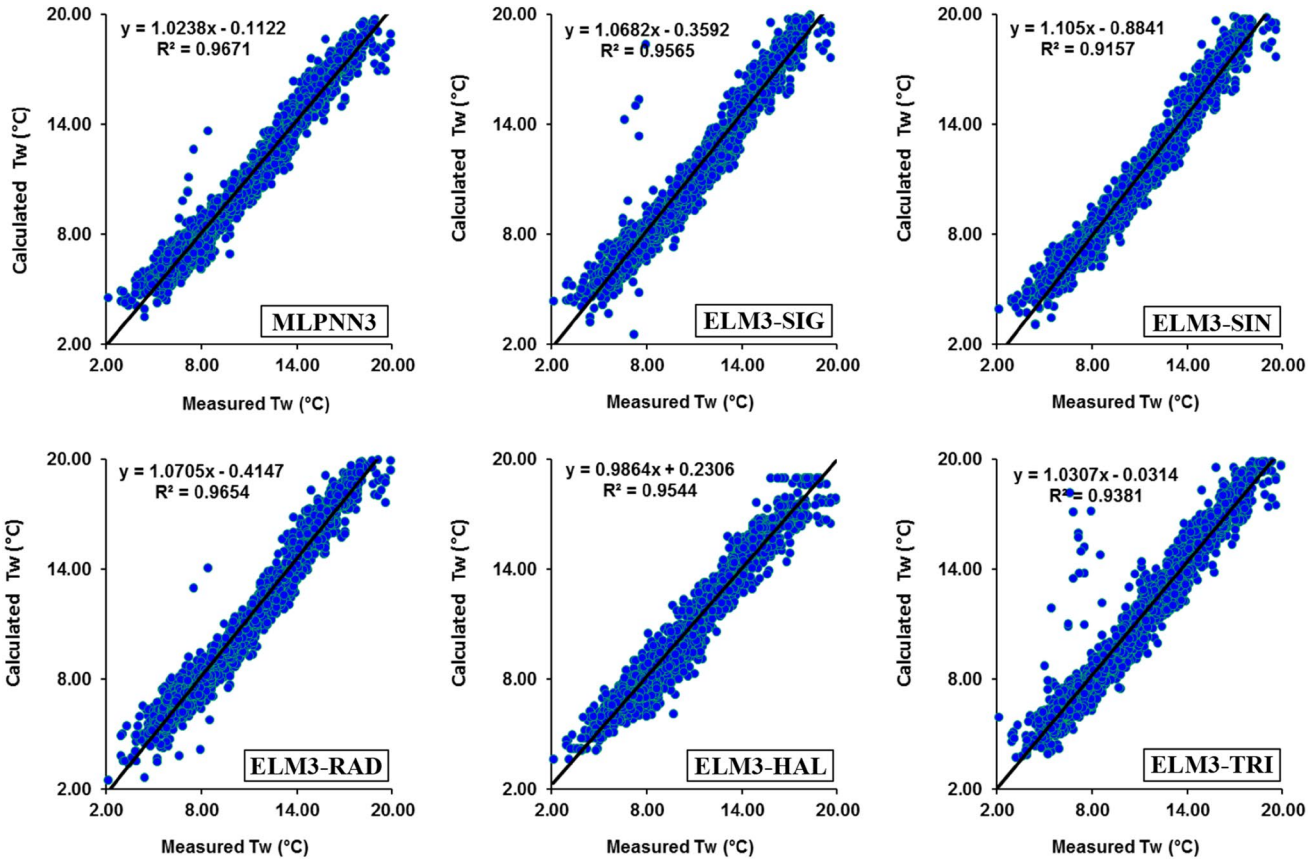

Fig. 3 Scatterplots showing the relation between the measured and calculated values of water temperature (T_w °C) in the validation phase of the Cedar station, using the five ELM3 and the MLPNN3 models

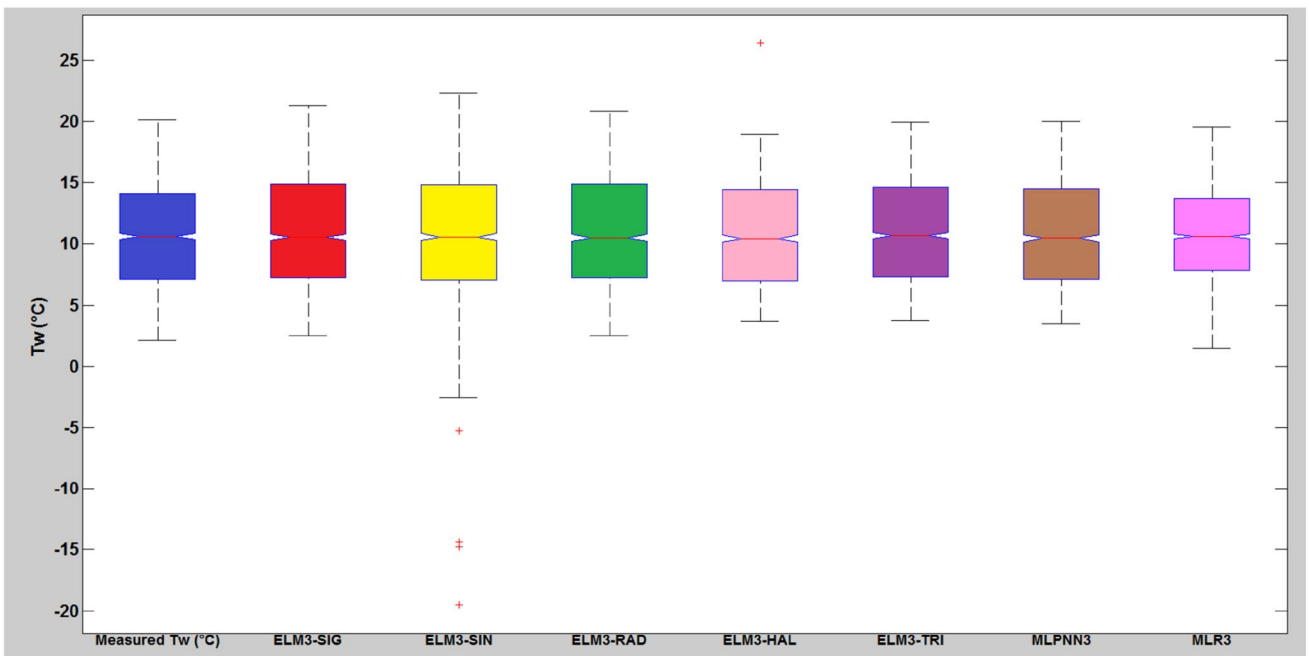

Fig. 4 Box plots of measured and calculated values of water temperature (T_w °C) in the validation phase of the Cedar station, using the five ELM3 and the MLPNN3 models. The central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers correspond to the most extreme data points

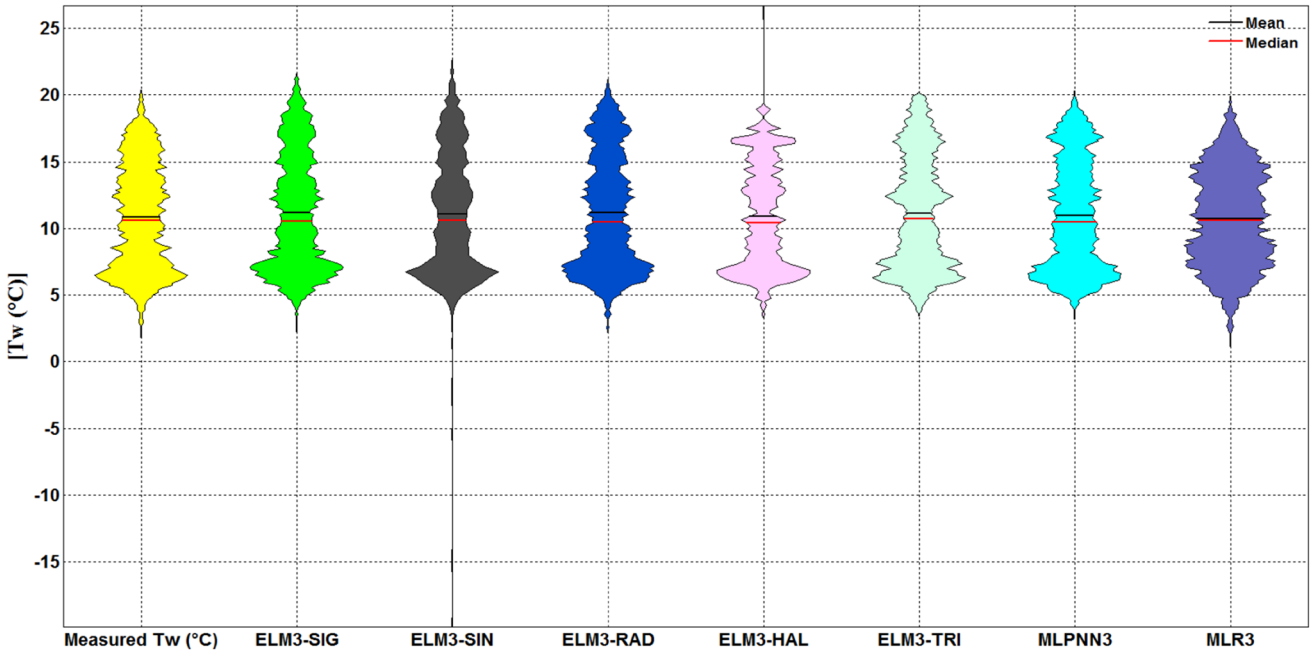

Fig. 5 Violin plots of measured and calculated daily water temperature (T_w °C) in the validation phase of the Cedar station

when compared to the ELM3 with sine activation function, the ELM3 sigmoidal activation functions is more accurate. When using only the T_a as input variable, all developed models showed similar results with negligible differences. Also it is clear from Table 3, that with the inclusion of the discharge Q as input variable, the improvement accuracy was marginal. This statement indicated that Q played a minor role and T_a and DOY are the most important explanatory variables. In the validation phase, as shown in Table 3, the best fit model was found when using the three input variables as inputs (T_a, Q and DOY) with ELM3 using radial basis function ($R=0.986$, $d=0.993$), while the worst fit was observed for the ELM3 using the Hardlim activation function with marginal difference in accuracy. The MLPNN3 provided relatively similar accuracy. Furthermore, both models (ELM3 and MLPNN3) showed a relatively high accuracy. Also, by analyzing the obtained results and by comparing all ELM2 and ELM1 models, it can be concluded that by inclusion of the Q to the input variables, the performance of the models was not improved. However, using the DOY combined with T_a and Q correspond to an improvement in MLPNN3 prediction accuracy of 31.45% and 29.47% regarding the RMSE and MAE, compared with MLPNN1. ELM1 was less accurate than ELM3, and the latter yielded improvement of 32.80% and 33.18% against ELM1. The gain in accuracy achieved through the inclusion of the DOY as input, in comparison to the simple models using only T_a as input, was comparatively large. The scatterplot of measured and calculated water temperature (T_w °C) is shown in Fig. 6. For comparison, the results using the simple MLR model are also presented in Table 3. According to Table 3, it is clear that the three MLR models provided relatively similar accuracy and with the inclusion of the Q and DOY in addition to the T_a, the performance of the models was not improved. In addition, the MLR model provided less accuracy compared to the MLPNN3 and all the ELM models with different activation functions.

The box plots and the violin plots for measured and calculated water temperature (T_w °C) values using the MLPNN3, MLR3 and the five ELM3 models for Fanno station are shown in Figs. 7 and 8. The lowest precisions within the proposed models correspond to the MLR3, while the highest precisions were generally related to the MLPNN3 and ELM3-RAD models (Fig. 7). Regarding the violin plot reported in Fig. 8, the high resemblance between measured and calculated T_w values was achieved using MLPNN3, ELM3-RAD and ELM3-SIG, and the MLR3, ELM3-TRI and ELM3-HAL performed worse for T_w modeling.

At Irondequoit station, statistical indices of the ELM and MLPNN models are listed in Table 4. During the training phase, results showed that high correlation between measured and calculated T_w existed for both ELM and MLPNN models. Using only T_a as input variable, R and d ranged from 0.960 to 0.961, and 0.979 to 0.980 for ELM1 models, and values equal to 0.961 and 0.980 for the MLPNN1 model. The MAE values observed from training data in the five ELM1 models tended to be of a similar magnitude, roughly from 1.474 to 1.488 °C. Similarly, RMSE were fairly constant across models, roughly from 1.972 to 1.995 °C. Inclusion of the Q variable slightly improves the performance of

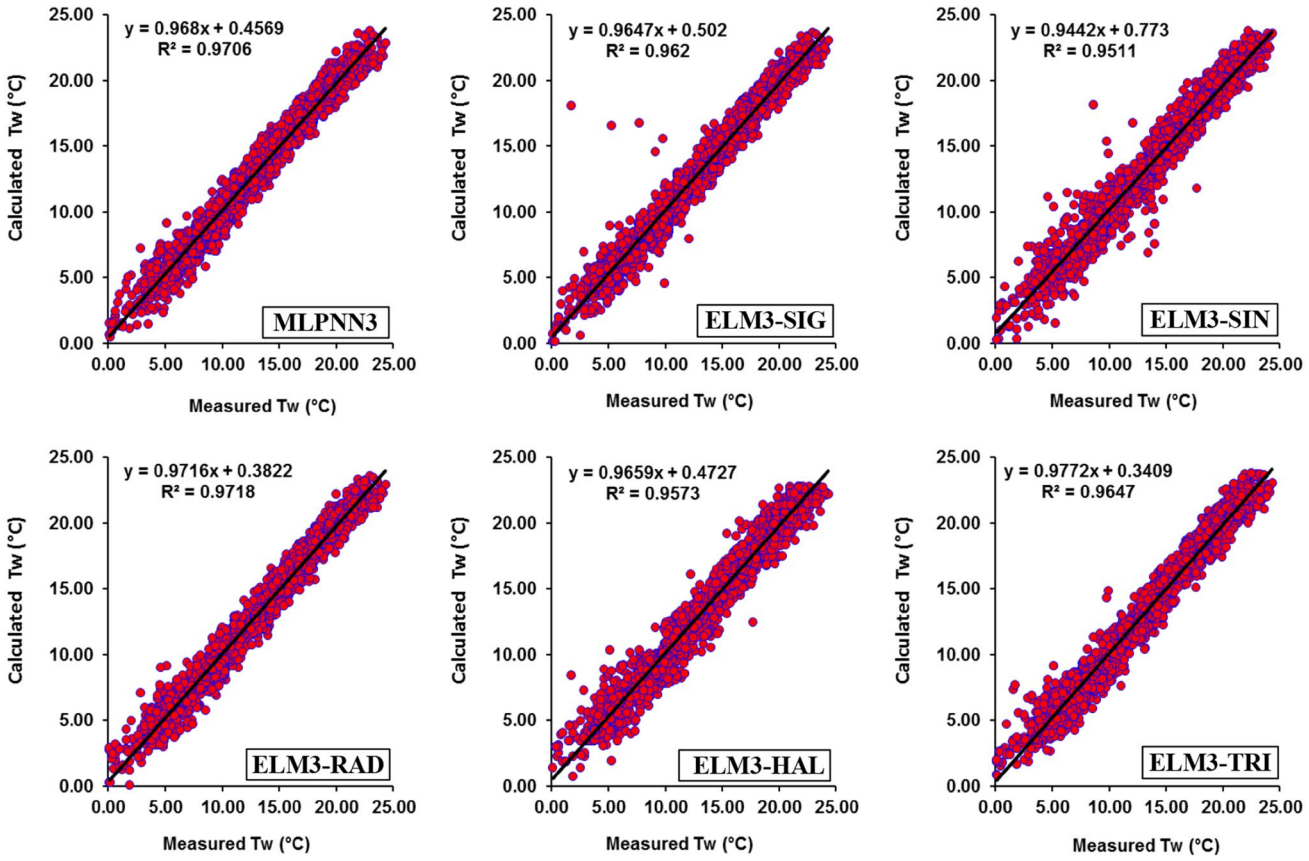

Fig. 6 Scatterplots showing the relation between the measured and calculated values of water temperature (T_w °C) in the validation phase of the Fanno station, using the five ELM3 and the MLPNN3 models

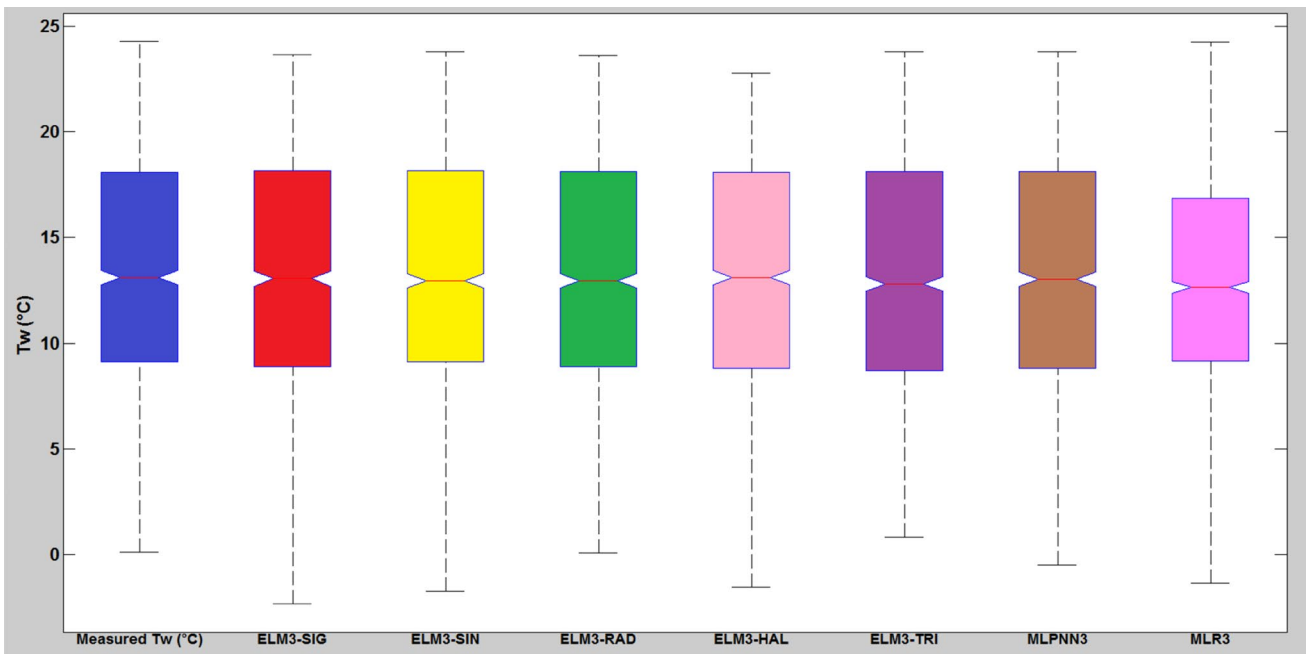

Fig. 7 Box plots of measured and calculated values of water temperature (T_w °C) in the validation phase of the Fanno station, using the five ELM3 and the MLPNN3 models. The central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers correspond to the most extreme data points

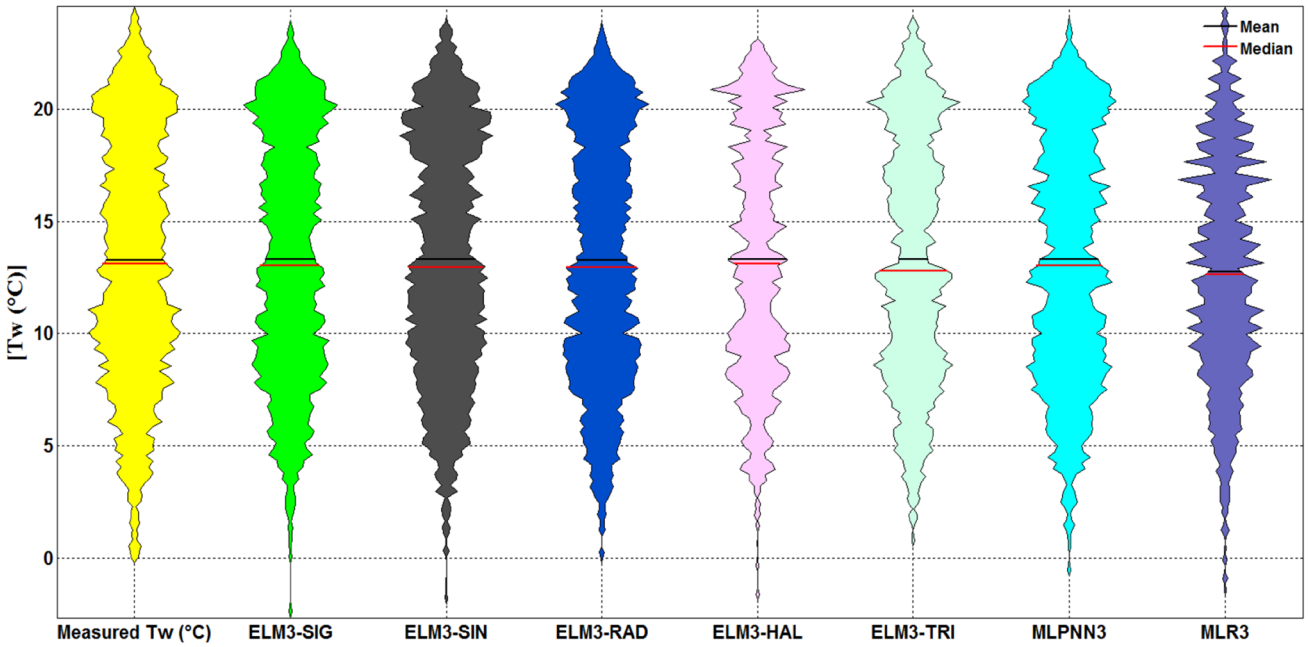

Fig. 8 Violin plots of measured and calculated daily water temperature (T_w °C) in the validation phase of the Fanno station

the models and in the major cases the improvement accuracy is marginal and overall negligible. Despite the change in accuracy as stated in Table 4, it is clear that the inclusion of Q did not lead to a significant increase of the R and d values and to a significant decrease in the RMSE and MAE error indices. For ELM2 models, the R and d values switched from 0.960 to 0.964 and from 0.979 to 0.982, respectively. Similarity, the value of the RMSE and MAE dropped from 1.995 to 1.930 °C and from 1.488 to 1.415 °C. It was noted that for the MLPNN2 model the improvement was greater than the ELM2 model, the RMSE and MAE dropped from 1.983 to 1.900 °C and from 1.480 to 1.420 °C, while the improvement regarding the R and d was marginal. Finally, the best accuracy was obtained using the ELM3 and MLPNN3 models, and with the inclusion of the DOY variable as inputs, the performance of the models was significantly improved. Analysis shows that the RMSE and MAE of the MLPNN1 are reduced by 45.63% and 43.24%, respectively, which is quite a significant improvement. Using the DOY as the input variable, the performances of the ELM3 model were significantly improved. Analysis shows that the RMSE and MAE of the ELM1 are reduced by 46.50% and 44.14%, respectively, which is also a significant improvement.

During the validation phase, the five ELM1 and MLPNN1 models had overall RMSE error ranging from 1.860 to 1.983 °C, and a value equal to 1.867 °C for MLPNN1, and the lowest RMSE was found to be 1.860 °C. Regarding the R and d values, ELM1 with sigmoidal activation function possess the elevated values ($R = 0.986$, $d = 0.991$), followed by ELM1 with radial basis function and the MLPNN1 ranged in the third place. By analyzing the obtained results at Irondequoit and by comparing the performances of the models, it is clear from Table 4, that with inclusion of the Q to the input variables, the performances of the models have slightly decreased. Hence, increasing the number of input variables did not necessarily lead to better model performance, and these results reveal the importance of input variables selection. Finally, using all the three variables as inputs (T_a, Q and DOY), the performances of the models were significantly improved. The most noticeable remark is the marked reduction in RMSE and MAE values. ELM3 reduced the RMSE and MAE of the ELM1 by 34.82% and 33.64%, respectively. Similarly, MLPNN3 reduced the RMSE and MAE of the MLPNN1 by 33.47% and 32.73%, respectively. By comparison, the ELM3 with sigmoidal activation function is slightly higher than the MLPNN3 and the difference in performances is marginal. It is clear from the results reported in Table 4, that the lowest accuracy was obtained using the MLR models compared to all the MLPNN and ELM models, with high RMSE and MAE values, and lower values of R and d indices. Additionally, including Q and DOY as input variables did not help to improve the performances of the MLR models. The scatterplot of measured and calculated water temperature (T_w °C) is shown in Fig. 9.

The box plots and the violin plots for measured and calculated water temperature (T_w °C) values using the MLPNN3, MLR3 and the five ELM3 models for Irondequoit station are shown in Figs. 10 and 11. The box plots in Fig. 10 give an idea of the accuracy of the developed models. For MLPNN3, ELM3-RAD and ELM3-SIN, the estimated and measured

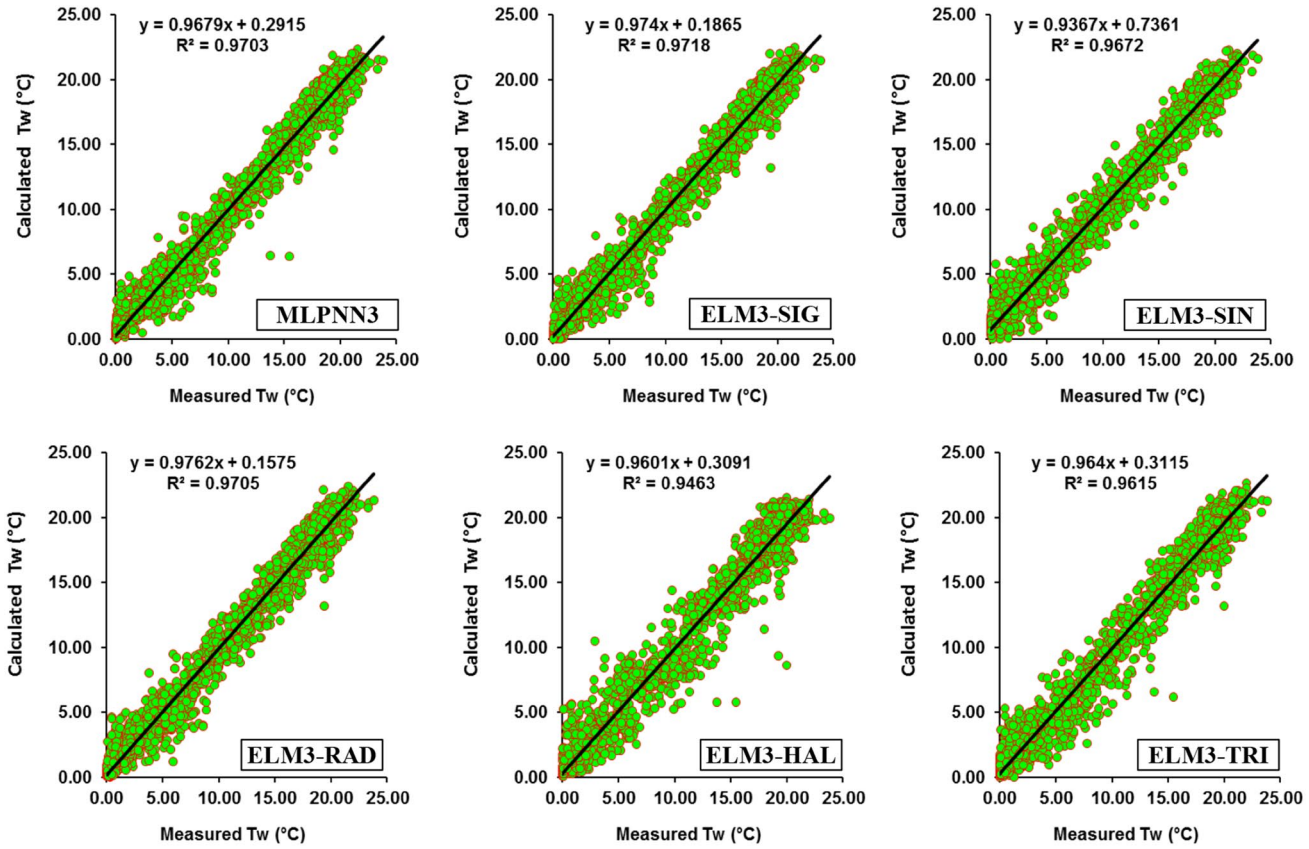

Fig. 9 Scatterplots showing the relation between the measured and calculated values of water temperature (T_w °C) in the validation phase of the Irondequoit station, using the five ELM3 and the MLPNN3 models

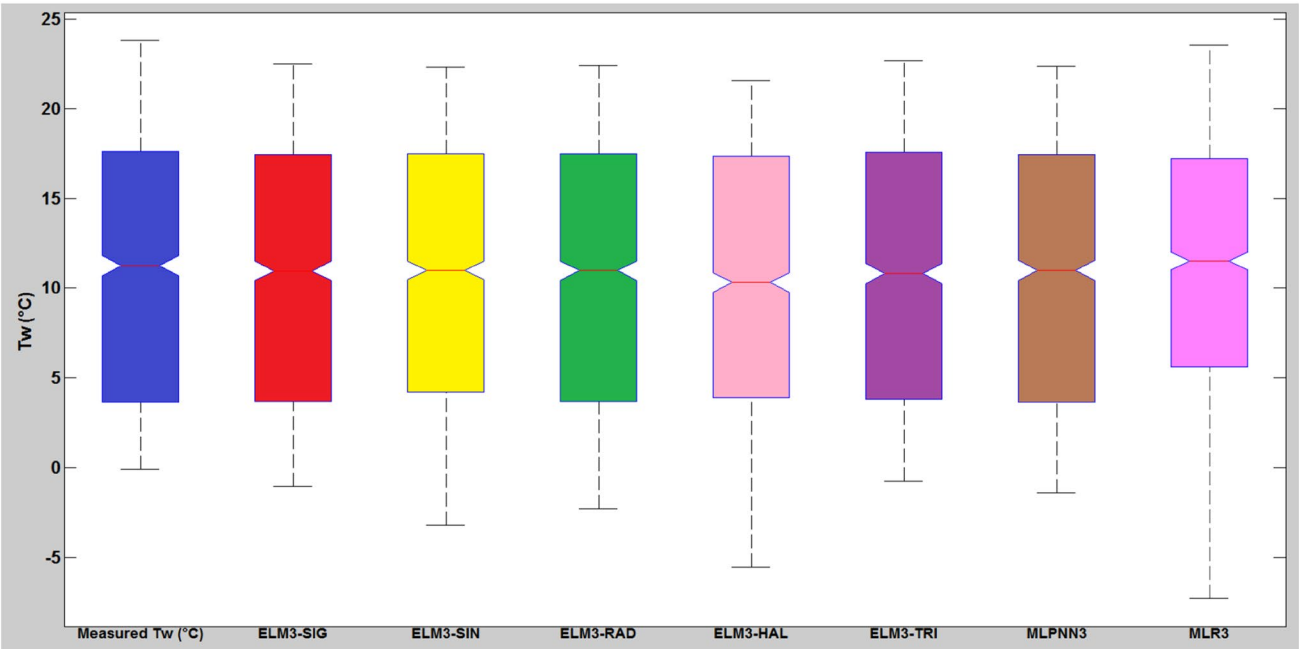

Fig. 10 Box plots of measured and calculated values of water temperature (T_w °C) in the validation phase of the Irondequoit station, using the five ELM3 and the MLPNN3 models. The central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers correspond to the most extreme data points

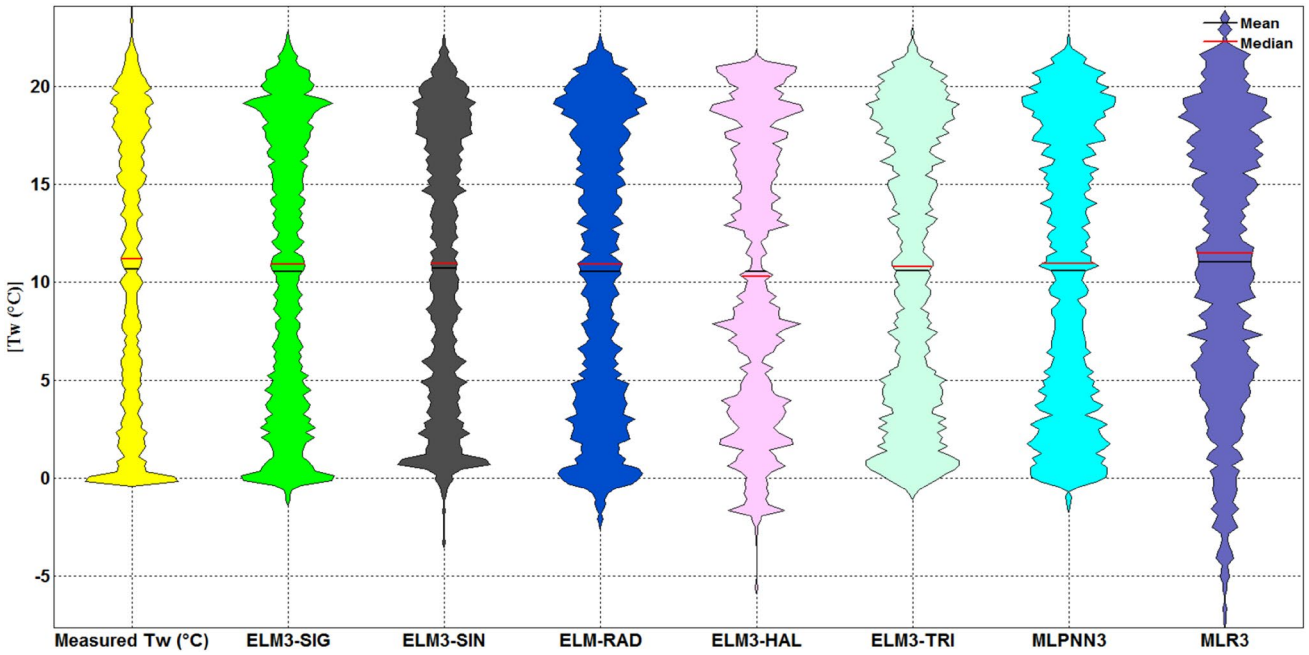

Fig. 11 Violin plots of measured and calculated daily water temperature (T_w °C) in the validation phase of the Irondequoit station

values were relatively similar with high resemblance, while the extreme values tend to remain large for the negative and low for the positive values. However, the extent of variation is smaller for ELM3-SIG than the other models, signifying that the retrieval accuracy of ELM3-SIG has lower sensitivity for varying water temperature values. In contrast, the MLR3 model shows a large extent of variation with high negative and positive values. This statement suggests that MLR3 has a large prediction error and the lowest accuracy among the developed models.

Finally in Fig. 12, Taylor diagram plot was used to compare the accuracy of the proposed ELM3, MLPNN3 and MLR3 models during the validation phase, using the correlation coefficient (R) and the standard deviation (SD). Taylor diagram is one of the most and highly recommended diagrams for performance comparison of machine learning techniques (Samadianfard et al. 2018; Khatibi et al. 2017; Kim et al. 2018a, b; Ghorbani et al. 2018a, b). It is clear from the diagram that the observed value (the reference) is plotted using a red circle along the X-axis. A different color point was drawn based on the calculated values of the T_w using the five ELM3, MLPNN3 and the MLR3 models. We can conclude that if the SD of the calculated value is higher than the SD of the measured value, it will result in

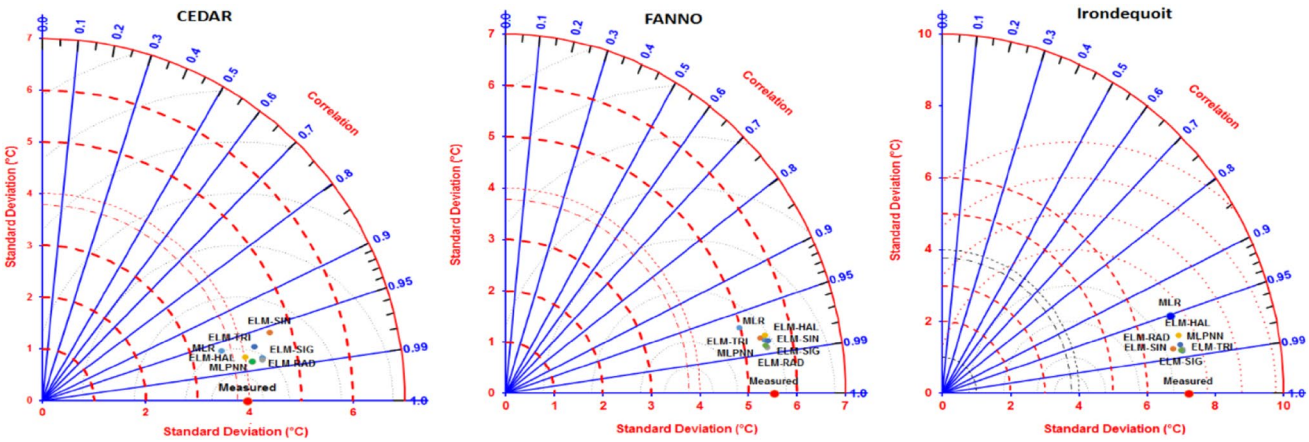

Fig. 12 Taylor diagram showing the performance of different ELM3, MLPNN3 and MLR3 models in terms of correlation coefficient and standard deviation between measured and calculated T_w during the validation phase for the three stations

overestimation and vice versa. It is shown in Fig. 12 that ELM-SIG and ELM-RAD performed better in general, since these points are closer to the reference points.

Conclusions

In this study, ELM models were developed to model daily water temperature for rivers. The proposed models were tested in three river stations characterized by different hydrological conditions. Modeling results showed that inclusion of three inputs as predictors (T_a, Q and the DOY) yielded the best modeling accuracy for all the developed models. The use of DOY is complementary to the use of T_a and Q, and provides additional relevant information on the seasonality of the river thermal dynamics, possibly mimicking the effect of lateral and upstream water and heat inputs. In addition, the results indicated that sigmoidal and radial basis functions within the ELM model performed the best for river water temperature forecasting. ELM and MLPNN models outperformed the MLR models with smaller RMSE and MAE values, and larger R and d values. ELM with sigmoidal and radial basis functions performed comparably with the traditional MLPNN models for T_w forecasting. The overall modeling performances indicated that ELM model developed in this study can be effectively used for river water temperature prediction.

Acknowledgements This work was jointly funded by the National Key R&D Program of China (2018YFC0407200), China Postdoctoral Science Foundation (2018M640499), and the research project from Nanjing Hydraulic Research Institute (Y118009).

References

Ahmadi-Nedushan B, St Hilaire A, Ouarda TBMJ, Bilodeau L, Robichaud É, Thiémonge N, Bobée B (2007) Predicting river water temperatures using stochastic models: case study of the Moisie River Quebec, Canada. Hydrol Process 21:21–34. https://doi.org/10.1002/hyp.6353

Atkinson PM, Tatnall ARL (1997) Introduction neural networks in remote sensing. Int J Remote Sens 18(4):699–709. https://doi.org/10.1080/014311697218700

Benyahya L, Caissie D, St-Hilaire A, Ouarda TBMJ, Bobée B (2007) A review of statistical water temperature models. Can Water Resour J 32:179–192. https://doi.org/10.4296/cwrj3203179

Cha Y, Cho KH, Lee H, Kang T, Kim JH (2017) The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers. Water Res 124:11–19. https://doi.org/10.1016/j.watres.2017.07.040

Chaves P, Kojiri T (2007) Conceptual fuzzy neural network model for water quality simulation. Hydrol Process 21:634–646. https://doi.org/10.1002/hyp

Cole JC, Maloney KO, Schmid M, McKenna JE (2014) Developing and testing temperature models for regulated systems: a case study on the Upper Delaware River. J Hydrol 519:588–598. https://doi.org/10.1016/j.jhydrol.2014.07.058

Deo RC, Samui P, Kim D (2016) Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. Stoch Environ Res Risk Assess 30:1769–1784. https://doi.org/10.1007/s00477-015-1153-y

Deo RC, Ghorbani MA, Samadianfard S, Maraseni T, Bilgili M, Biazar M (2018) Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for wind speed prediction of target site using a limited set of neighboring reference station data. Renew Energy 116:309–323. https://doi.org/10.1016/j.renene.2017.09.078

DeWeber JT, Wagner T (2014) A regional neural network ensemble for predicting mean daily river water temperature. J Hydrol 517:187–200. https://doi.org/10.1016/j.jhydrol.2014.05.035

Du X, Shrestha NK, Wang J (2019) Assessing climate change impacts on stream temperature in the Athabasca River Basin using SWAT equilibrium temperature model and its potential impacts on stream ecosystem. Sci Total Environ 650:1872–1881. https://doi.org/10.1016/j.scitotenv.2018.09.344

Gallice A, Schaefli B, Lehning M, Parlange MB, Huwald H (2015) Stream temperature prediction in ungauged basins: review of recent approaches and description of a new physics-derived statistical model. Hydrol Earth Syst Sci 19:3727–3753. https://doi.org/10.5194/hess-19-3727-2015

Garner G, Malcolm IA, Sadler JP, Hannah DM (2017) The role of riparian vegetation density, channel orientation and water velocity in determining river temperature dynamics. J Hydrol 553:471–485. https://doi.org/10.1016/j.jhydrol.2017.03.024

Ghiassi M, Nangoy S (2009) A dynamic artificial neural network model for forecasting nonlinear processes. Comput Ind Eng 57:287–297. https://doi.org/10.1016/j.cie.2008.11.027

Ghorbani MA, Deo RC, Yaseen ZM, Kashani MH, Mohammadi B (2018a) Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. Theoret Appl Climatol 133:1119–1131. https://doi.org/10.1007/s00704-017-2244-0

Ghorbani MA, Deo RC, Karimi V, Yaseen ZM, Terzi O (2018b) Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. Stoch Environ Res Risk Assess 32:1683–1697. https://doi.org/10.1007/s00477-017-1474-0

Ghorbani MA, Deo RC, Kashani MH, Shahabi M, Ghorbani S (2019) Artificial intelligence-based fast and efficient hybrid approach for spatial modelling of soil electrical conductivity. Soil Tillage Res 186:152–164. https://doi.org/10.1016/j.still.2018.09.012

Hadzima-Nyarko M, Rabi A, Šperac M (2014) Implementation of artificial neural networks in modeling the water-air temperature relationship of the River Drava. Water Resour Manage 28:1379–1394. https://doi.org/10.1007/s11269-014-0557-7

Haykin S (1999) Neural networks a comprehensive foundation. Prentice Hall, Upper Saddle River

Heddam S, Kisi O (2017) Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. Environ Sci Pollut Res 24:1–23. https://doi.org/10.1007/s11356-017-9283-z

Heddam S, Kisi O (2018) Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. J Hydrol 559:499–509. https://doi.org/10.1016/j.jhydrol.2018.02.061

Hornik K (1991) Approximation capabilities of multilayer Feedforward networks. Neural Netw 4(2):251–257. https://doi.org/10.1016/0893-6080(91)90009-T

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Netw 2(89):359–366. https://doi.org/10.1016/0893-6080 90020–8

Huang GB, Chen L (2008) Enhanced random search based incremental extreme learning machine. Neurocomputing 71:3460–3468. https://doi.org/10.1016/j.neucom.2007.10.008

Huang GB, Chen L, Siew CK (2006a) Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 17(4):879–892. https://doi.org/10.1109/TNN.2006.875977

Huang GB, Zhu QY, Siew CK (2006b) Extreme learning machine: theory and applications. Neurocomputing 70:489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Huang G, Huang GB, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48. https://doi.org/10.1016/j.neunet.2014.10.001

Humphrey GB, Gibbs MS, Dandy GC, Maier HR (2016) A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. J Hydrol 540:623–640. https://doi.org/10.1016/j.jhydrol.2016.06.026

Isaak DJ, Luce CH, Horan DL, Chandler GL, Wollrab SP, Nagel DE (2018) Global Warming of Salmon and Trout Rivers in the Northwestern U.S.: road to ruin or path through purgatory? Trans Am Fish Soc 147: 566–585. https://doi.org/10.1002/tafs.10059

Karami F, Dariane AB (2017) Optimizing signal decomposition techniques in artificial neural network-based rainfall-runoff model. Int J River Basin Manag 15:1–8. https://doi.org/10.1080/15715124.2016.1203331

Kędra M, Wiejaczka Ł (2018) Climatic and dam-induced impacts on river water temperature: assessment and management implications. Sci Total Environ 626:1474–1483. https://doi.org/10.1016/j.scitotenv.2017.10.044

Khatibi R, Ghorbani MA, Pourhosseini FA (2017) Stream flow predictions using nature-inspired firefly algorithms and a multiple model strategy-directions of innovation towards next generation practices. Adv Eng Inform 34:80–89. https://doi.org/10.1016/j.aei.2017.10.002

Kim JS, Seo IW, Lyu S, Kwak S (2018a) Modeling water temperature effect in diatom (Stephanodiscus hantzschii) prediction in eutrophic rivers using a 2D contaminant transport model. J Hydro Environ Res 19:41–55. https://doi.org/10.1016/j.jher.2018.01.003

Kim S, Seo Y, Rezaie-Balf M, Kisi O, Ghorbani MA, Singh VP (2018b) Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: peninsula vs continent. Theor Appl Climatol. https://doi.org/10.1007/s00704-018-2627-x

Krider LA, Magner JA, Perry J, Vondracek B, Ferrington LC (2013) Air-water temperature relationships in the trout streams of southeastern Minnesota's carbonate-sandstone landscape. J Am Water Resour Assoc 49:896–907. https://doi.org/10.1111/jawr.12046

Kwak J, St-Hilaire A, Chebana F (2017) A comparative study for water temperature modelling in a small basin, the Fourchue River, Quebec, Canada. Hydrol Sci J 62:64–75. https://doi.org/10.1080/02626667.2016.1174334

Liang NY, Huang GB, Rong HJ, Saratchandran P, Sundararajan N (2006) A fast and accurate on-line sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw 17:1411–1423. https://doi.org/10.1109/TNN.2006.880583

Marcé R, Armengol J (2010) Modelling river water temperature using deterministic, empirical, and hybrid formulations in a Mediterranean stream. Hydrol Process 22:3418–3430. https://doi.org/10.1002/hyp.6955

McCulloch WS, Pitts W (1943) A logical calculus of the ideas imminent in nervous activity. Bull Math Biophys 5:115–133. https://doi.org/10.1007/BF02478259

Piccolroaz S, Calamita E, Majone B, Gallice A, Siviglia A, Toffolon M (2016) Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. Hydrol Process 30:3901–3917. https://doi.org/10.1002/hyp.10913

Piotrowski AP, Napiorkowski JJ (2018) Performance of the air2stream model that relates air and stream water temperatures depends on the calibration method. J Hydrol 561:395–412. https://doi.org/10.1016/j.jhydrol.2018.04.016

Piotrowski AP, Osuch M, Napiorkowski MJ, Rowinski PM, Napiorkowski JJ (2014) Comparing large number of metaheuristics for artificial neural networks training to predict water temperature in a natural river. Comput Geosci 64:136–151. https://doi.org/10.1016/j.cageo.2013.12.013

Piotrowski AP, Napiorkowski MJ, Napiorkowski JJ, Osuch M (2015) Comparing various artificial neural network types for water temperature prediction in rivers. J Hydrol 529:302–315. https://doi.org/10.1016/j.jhydrol.2015.07.044

Pohle I, Helliwell R, Aube C, Gibbs S, Spencer M, Spezia L (2018) Citizen science evidence from the past century shows that Scottish rivers are warming. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2018.12.325

Rabi A, Hadzima-Nyarko M, Sperac M (2015) Modelling river temperature from air temperature in the River Drava (Croatia). Hydrol Sci J 60:1490–1507. https://doi.org/10.1080/02626667.2014.914215

Rezaie-Balf M, Kisi O (2017) New formulation for forecasting streamflow: evolutionary polynomial regression vs. extreme learning machine. Hydrol Res 49:939–953. https://doi.org/10.2166/nh.2017.283

Sahoo GB, Schladow SG, Reuter JE (2009) Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. J Hydrol 378:325–342. https://doi.org/10.1016/j.jhydrol.2009.09.037

Samadianfard S, Ghorbani MA, Mohammadi B (2018) Forecasting soil temperature at multiple-depth with a hybrid artificial neural network model coupled-hybrid firefly optimizer algorithm. Inform Process Agric 5:465–476. https://doi.org/10.1016/j.inpa.2018.06.005

Šiljić Tomić A, Antanasijević D, Ristić M, Perić-Grujić A, Pocajt V (2018a) A linear and non-linear polynomial neural network modeling of dissolved oxygen content in surface water: inter- and extrapolation performance with inputs significance analysis. Sci Total Environ 610–611:1038–1046. https://doi.org/10.1016/j.scitotenv.2017.08.192

Šiljić Tomić A, Antanasijević D, Ristić M, Perić-Grujić A, Pocajt V (2018b) Application of experimental design for the optimization of artificial neural network-based water quality model: a case study of dissolved oxygen prediction. Environ Sci Pollut Res 25:9360–9370. https://doi.org/10.1007/s11356-018-1246-5

Temizyurek M, Dadasercelik F (2018) Modelling the effects of meteorological parameters on water temperature using artificial neural networks. Water Sci Technol 77:1724–1733. https://doi.org/10.2166/wst.2018.058

Toffolon M, Piccolroaz S (2015) A hybrid model for river water temperature as a function of air temperature and discharge. Environ Res Lett 10:114011. https://doi.org/10.1088/1748-9326/10/11/114011

Trichakis IC, Nikolos IK, Karatzas GP (2011) Artificial neural network (ANN) based modeling for Karstic groundwater level simulation. Water Resour Manage 25:1143–1152. https://doi.org/10.1007/s11269-010-9628-6

Van Vliet MTH, Ludwig F, Zwolsman JJG, Weedon GP, Kabat P (2011) Global river temperatures and sensitivity to atmospheric warming and changes in river flow. Water Resour Res 47:247–255. https://doi.org/10.1029/2010WR009198

Van Vliet MTH, Yearsley JR, Franssen WHP, Ludwig F, Haddeland I, Lettenmaier DP, Kabat P (2012) Coupled daily streamflow and water temperature modeling in large river basins.

Hydrol Earth Syst Sci 16:4303–4321. https://doi.org/10.5194/hess-16-4303-2012

Vasu NN, Lee SR (2016) A hybrid feature selection algorithm integrating an extreme learning machine for landslide susceptibility modeling of Mt. Woomyeon, South Korea. Geomorphology 263:50–70. https://doi.org/10.1016/j.geomorph.2016.03.023

Webb BW, Clack PD, Walling DE (2003) Water-air temperature relationships in a Devon river system and the role of flow. Hydrol Process 17:3069–3084. https://doi.org/10.1002/hyp.1280

Williamson RJ, Entwistle NS, Collins DN (2018) Meltwater temperature in streams draining Alpine glaciers. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2018.12.215

Yaseen ZM, Jaafar O, Deo RC, Kisi O, Adamowski J, Quilty J, El-Shafie A (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. J Hydrol 542:603–614. https://doi.org/10.1016/j.jhydrol.2016.09.035

Zhu S, Nyarko EK, Nyarko MH (2018a) Modelling daily water temperature from air temperature for the Missouri River. Peer J 6:e4894. https://doi.org/10.7717/peerj.4894

Zhu S, Heddam S, Nyarko EK, Hadzima-Nyarko M, Piccolroaz S, Wu S (2018b) Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-018-3650-2