Research papers

River water temperature forecasting using a deep learning method

Rujian Qiu [a], Yuankun Wang [a,*], Bruce Rhoads [b], Dong Wang [a], Wenjie Qiu [a], Yuwei Tao [a], Jichun Wu [a]

[a] Key Laboratory of Surficial Geochemistry, Ministry of Education, Department of Hydrosciences, School of Earth Sciences and Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing, PR China
[b] Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Champaign, IL, USA

ARTICLE INFO

ABSTRACT

Accurate water temperature forecasting is essential for understanding thermal regimes of rivers in the context of climate change and anthropogenic disturbances, such as dam construction. Machine-learning models proffer an empirically based approach to predicting water temperatures with a high degree of accuracy. This study explores the potential of long short-term neural network (LSTM), a type of deep learning method, to forecast daily river water temperatures and quantify temporal variations in thermal regime induced by changes in climate and by dam construction. The performance of LSTM is compared with that of several other models using daily water-temperature data for nine river gauges around the world. In a detailed analysis, the models are evaluated for the Yichang gauge on the Yangtze River to reconstruct the natural thermal conditions and help to assess daily water temperature variations induced by operation of the Three Gorges Reservoir (TGR). The collective results show that LSTM outperforms other methods for predicting mean daily water temperature in rivers, capturing accurately mean daily variations in thermal regime. The construction of the TGR strongly influenced water temperature variations at Yichang, producing the strongest cooling effect from mid-April to mid-May and the greatest warming effect in late December and early January. These marked effects are most prominent at the highest water levels in the TGR. The enhanced predictive capabilities of the LSTM model provide a powerful tool for water temperature forecasting and ecological management of rivers in the Anthropocene.

1. Introduction

Because water temperature (WT) strongly affects many physical, chemical and biological processes in rivers, it plays a crucial role in determining the quality of aquatic ecosystems (Caissie, 2006). Changes in WT affect amounts of dissolved oxygen (Marzadria et al., 2013; Matthews and Berg, 1997; Vega et al., 1998), the distribution and metabolic rate of fish species (Lessard and Hayes, 2003; Buentello et al., 2000; Durance and Ormerod, 2009) and other aquatic organisms (Ward and Stanford, 1982; Cha et al., 2017), the evaporation rate at the water surface (Maheu, et al., 2014; Wanders and Wada, 2015) and the formation of ice (Wanders et al., 2019). River temperature variations mainly depend on the heat flux between air and water at the water surface and on temperatures of different runoff components that contribute water to the river (e.g. surface runoff, groundwater inputs, and snowmelt inputs) (Mohseni and Stefan, 1999; Ficklin et al., 2013). Previous studies have shown that river WT is strongly linked to climate change (e.g., Kaushal et al., 2010; Isaak et al., 2012; Du et al., 2019) and anthropogenic perturbations, such as dams (e.g., Olden and Naiman, 2010; Poole and Berman, 2001; Cai et al., 2018). Given the importance of temperature as a factor influencing the environmental quality of rivers as well as the potential for change in thermal regimes related to human impacts, a critical need exists to forecast accurately river temperatures in an era of increasing human influence now referred to as the Anthropocene.

In the past several decades, two classes of water temperature models, physical-based and data-driven models have been developed. Physical-based models are based on heat budget equations and require site-specific data that are unavailable for many river systems; thus, the geographic extent over which such models can be implemented is limited (Grabowski et al., 2016; Jackson et al., 2018; Piotrowski and Napiorkowski, 2019). Data-driven models mainly use statistical or data-mining techniques to model WT as a function of other explanatory variables, such as air temperature and discharge. Air temperature (AT) is

commonly assumed to be the essential predictor for WT because variations in AT can be viewed as a surrogate for net changes in heat flux at the air–water interface (Van Vliet et al., 2013; Ficklin et al., 2013). Discharge (Q) acts as an additional explanatory variable that can strongly influence WT (Webb et al., 2003; Van Vliet et al., 2011).

Recent studies have compared and assessed the performances of machine-learning methods, a type of data-mining technique, for WT forecasting (e.g., Sahoo et al., 2009; St-Hilaire et al., 2012; Hadzima-Nyarko et al., 2014; DeWeber and Wagner, 2014; Cole et al., 2014; Piotrowski et al., 2015; Sohrabi et al., 2017; Piotrowski and Napiorkowski, 2019; Zhu et al., 2019b). Other work has coupled machine-learning models with data-processing techniques (Zhu et al., 2019a; Graf et al., 2019) to enhance predictive accuracy. However, data-driven methods have not explored fully the extent to which river WT forecasting can be improved through application of deep learning methods. Recently, application of deep learning methods (e.g., long short-term memory neural network, LSTM) to other time series forecasting tasks has demonstrated the advantages of these methods compared to conventional data-driven models (Shen. 2018).

River temperature has been significantly altered by the constructions of large dams, which in turn can influence aquatic ecosystems (Olden and Naiman, 2010). The Yangtze River, the longest river in China, has attracted considerable attention since the construction of the Three Gorge Reservoir (TGR), which is the world' largest and possibly most controversial water conservancy project (Liu et al., 2018). Over the past several decades, a large number of studies have examined the impacts of TGR on river flow regime, extreme drought events, sediment deposition, and habitat suitability of aquatic animals (e.g., Chai et al., 2019; Li et al., 2011; Yi et al., 2010; Wang et al., 2016, 2018). A few studies have explored the impact of the TGR on variations in WT on the annual and seasonal using either a linear regression approach (Tao et al., 2020) or a semi-physical model (air2stream, Cai et al., 2018). However, little effort has been made to daily variations in WT. Whether deep learning models can meet the challenge to predict accurately daily variations in thermal regime variations under the influence of anthropogenic perturbations requires further investigation.

The goals of this paper are twofold: 1) to evaluate the forecasting performance of an LSTM model to predict mean daily river temperatures compared to that of several other river-temperature predictive models; and 2) to use the LSTM to reconstruct the expected daily natural thermal regime of a regulated river that would occur in the absence of a dam and to identify factors influencing daily variations in the thermal regime of this river. The results improve the proficiency to predict an important factor, mean daily WT, related to the environmental quality of river systems, and also provide a new tool in evaluating the thermal regime variations in rivers regulated by dams.

2. Data for model calibration and verification

To assess the performance of LSTM in daily WT forecasting relative to other methods, a comparison between LSTM and several benchmark models (air2stream, RF, BPNN) was conducted using daily WT data for nine gauges on seven rivers in a variety of physiographic settings (Table 1). Data for gauges on rivers in Switzerland (Yvonand, Sion, and Davos) and the United States (Cedar, Fanno, Irondequoit) were selected on the basis that these data were included in previous studies (Piotrowski and Napiorkowski, 2018; Zhu et al., 2019b). The three gauges on the Yangtze River, Cuntan, Yichang, and Datong all have long-term daily records of water temperature required for model calibration. Daily AT and Q, two important controls of daily WT, were used as input variables to predict the WT for the same day (no lag time). For the purpose of model calibration and validation, data for each gauge were divided chronologically into calibration and validation periods by the ratio of 7:3, which is a commonly used ratio for model training and testing, although there is no formal guideline (Khosravi et al., 2018) (Table 1).

In addition to providing information for evaluating model performance, daily data on AT, Q, and WT for the Yichang gauge, located 44 km downstream from the TGR, provided an opportunity to evaluate the capability of the different models to accurately reconstruct daily water temperatures that would occur in the absence of a large reservoir. Past work has indicated that water temperatures at Yichang are strongly affected by the TGR (Liu et al., 2018). To quantify the potential effects of TGR on the Yangtze River thermal regime, data on Yichang were divided into a pre-TGR calibration period (1983–2002) and a post-TGR validation period (2003–2013). The LSTM and benchmark models were compared and used to reconstruct the natural thermal conditions at Yichang that would have occurred during the validation period in the absence of TGR –a necessary reconstruction for identifying the impacts of climate change and the TGR on WT. The TGR began to fill in June 2003 to raise the water level of the reservoir to 135 m. During the second filling stage of TGR, from October 2006 to September 2008, the water level reached 156 m. Subsequently, it rose to 173 m in November 2008, and by the end of October 2010 (the third filling) the reservoir water level was raised to its designed normal pool level of 175 m (Zhang et al., 2012). Since that time, the water level of TGR has been lowered to 145 m for flood control during mid-June to late September and then raised to 175 m for power generation from January to March (Liu et al., 2015; Zheng, 2016).

3. Methodology

3.1. Long short-term memory neural network (LSTM)

LSTM is a special kind of recurrent neural network (RNN) (Hochreiter and Schmidhuber, 1997). RNN has a recurrent hidden unit to deal with sequential data, and the output of each time step serves as the input to the next time step. A single RNN model updates only a single past state, and it is trained by the backpropagation-through-time algorithm (Mozer, 1989; Werbos, 1990), by which the loss function is propagated backward to determine updates to weights. With simple recurrent networks, backpropagation encounters the vanishing gradient problem in which the training signal becomes exponentially small as it propagates into the network, making backpropagation ineffective for

Table 1
Brief descriptions of the studied river gauges.

River name	Gauge name	Calibration period	Validation period	Meteorological station	Catchment size (km^2)	Statistics of daily WTMean/max/min/range (°C)	Description
The Yangtze River	Yichang	1983–2002	2003–2013	Yichang	1.01×10^6	18.47/29.70/7.40/22.30	plain
	Cuntan	1993–2006	2007–2011	Shapingba	0.87×10^6	18.69/28.80/7.60/21.20	mountainous
	Datong	1977–1984	1985–1987	Tongling	1.71×10^6	17.64/30.80/2.00/28.80	plain
Cedar	USGS12119000	2001–2012	2013–2017	USW00024233	477	10.45/20.20/2.10/18.10	mountainous
Fanno	USGS14206950	2003–2012	2013–2017	USW00024229	82	12.77/24.80/0.10/24.70	hilly
Irondequoit	USGS0423205010	2005–2013	2014–2017	USW00014768	368	10.85/24.50/-0.10/24.60	lowland
Mentue	Yvonand	2002–2009	2010–2012	USW00024233	105	9.72/21.70/0.00/21.70	lowland
Rhône	Sion	1984–2004	2005–2013	USW00024229	3373	7.00/12.16/0.31/11.86	valley
Dischmabach	Davos	2004–2009	2010–2012	USW00014768	43.3	4.30/11.02/0.17/10.85	mountainous

deep networks (Shen, 2018). The LSTM is a special variation of RNN which can effectively avoid the vanishing gradient problem of RNN.

LSTM consists of an input layer, one or more LSTM layers, one or more fully connected layers and an output layer. The key to LSTM is the hidden neurons in the LSTM layer called memory cells (LSTM cells), which not only receive information from the input layer, but also perceive the information at the previous moment (Olah, 2015). The LSTM layer consists of two kinds of state, the hidden state and the cell state. At each time step, the hidden state contains the output of the LSTM layer for this time step, and the cell state contains the information learned from the previous time steps. The LSTM cells use three gates, the forget gate f_t, the input gate i_t, and the output gate o_t, to control the level of cell state reset (forget), the level of cell state update, and the level of cell state added to the hidden state, respectively (Fig. 1 (a)). The LSTM layer uses the tanh as the activation function to update the cell and hidden state, and the sigmoid function are used in the calculations for the gates. The working mechanism of LSTM layer can be expressed as following (Olah, 2015; Xiao et al., 2019):

$$f_t = \sigma\left(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i\right) \tag{2}$$

$$o_t = \sigma\left(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o\right) \tag{3}$$

$$\widetilde{c}_t = tanh\left(W_{\widetilde{c},x}x_t + W_{\widetilde{c},h}h_{t-1} + b_{\widetilde{c}}\right) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t \tag{5}$$

$$h_t = tanh(c_t) \odot o_t \tag{6}$$

where x_t is the input; h_{t-1} is the hidden state at time t-1; $W_{f,x}$, $W_{f,h}$, $W_{i,x}$, $W_{i,h}$, $W_{o,x}$, $W_{o,h}$, $W_{\widetilde{c},x}$, and $W_{\widetilde{c},h}$ are corresponding weight matrices of the three gates and the candidate value; b_f, b_i, b_o, and $b_{\widetilde{c}}$ denote corresponding bias vectors; $\odot$ represents the element-wise multiplication between two vectors; $\sigma(\cdot)$ is the logistic sigmoid activation function; $tanh(\cdot)$ is the hyperbolic tangent function; and h_t is the output of LSTM layer at time step t.

(a) Data flow at time step t

(b) Data flow in the LSTM model using time series of AT and Q as the inputs to predict WT
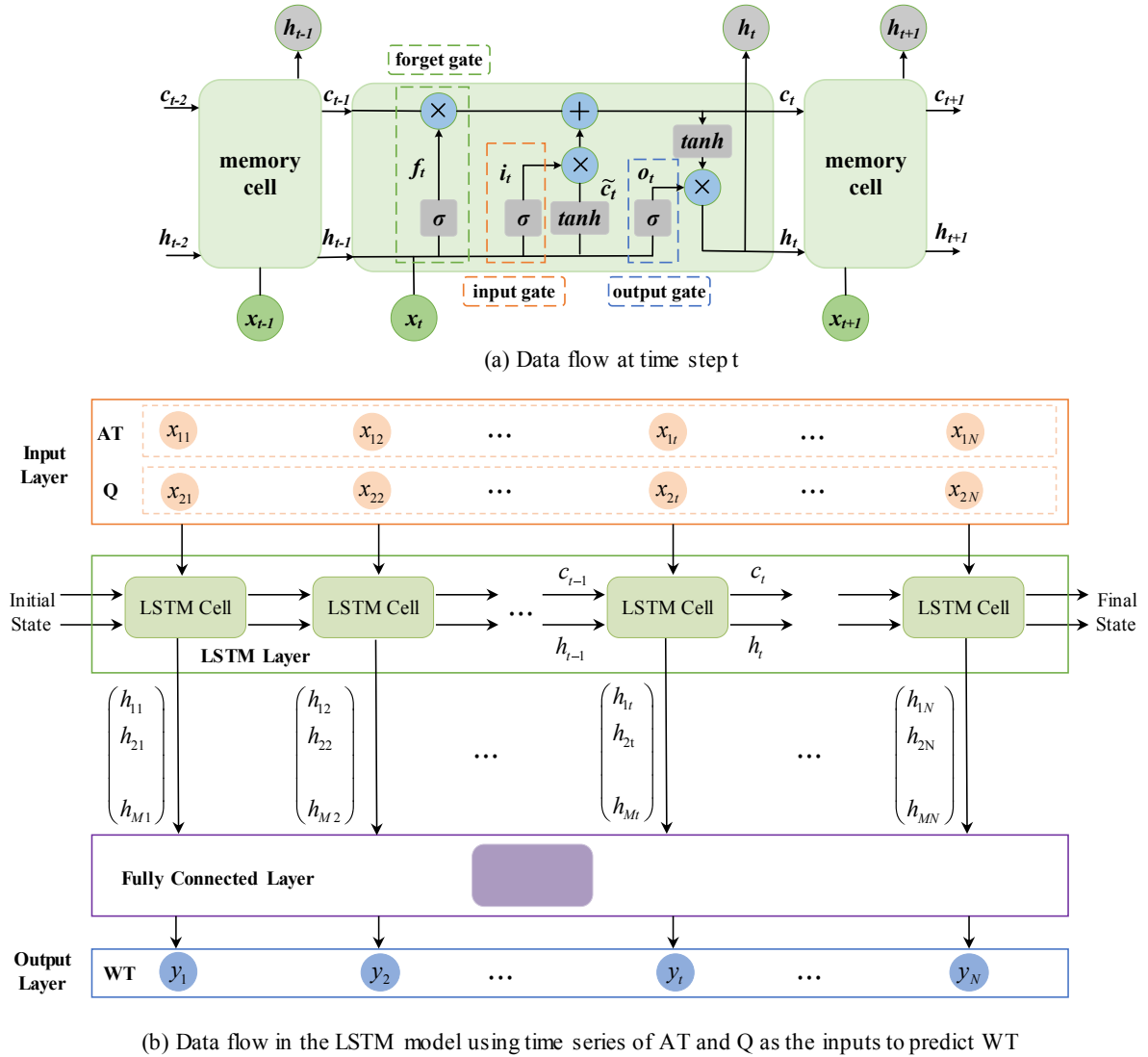
Fig. 1. Architecture of LSTM for WT forecasting*. Note*: Times series of AT and Q $\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \end{pmatrix}$ were used as the input variables to predict the WT ($y_1 \ y_2 \ \cdots \ y_N$) for the same period (no lag time), where N denotes the length of time series. ($h_{1t} \ h_{2t} \ \cdots \ h_{Mt}$) represents the output of LSTM layer at time step t, where M denotes the number of LSTM cells (hidden units) in the LSTM layer.

The working mechanism and data flows of LSTM (Fig. 1) involve several steps. First, the input data x_t at time step t and the hidden state of the previous moment h_{t-1} flow to the forget gate to determine how much previous information should be removed from the cell state. The data are transformed to a range from 0 to 1 by the sigmoid activation function $\sigma(\cdot)$ with a value of one indicating "completely keep it" and a value of zero signaling "completely remove it". The retained information in the hidden state flows to the input gate, where it serves as new input information to the memory cell for updating the cell state. Then the new candidate value $\widetilde{c}_t$ is generated, which is stored in the cell state c_t. Finally, the updated cell state (c_t) flows to the output gate, which controls the information within the cell state that is propagated to the hidden variable state at the next time step.

Based on the LSTM, a 4-layer deep neural network model is constructed for WT forecasting. This model consists of an input layer, a LSTM layer, a fully connected layer, and an output layer (Fig. 1(b)). The number of LSTM cells is set to a range of 30 to 50 by the trial and error method (Table S1), and a dropout value of 0.5 is applied empirically to help prevent overfitting (Yan et al., 2020; Zhang et al., 2018).

3.2. Benchmark models

3.2.1. Air2stream model

The air2stream model is based on lumped heat budget exchange between an unknown water volume, tributaries and atmosphere in the river (Toffolon and Piccolroaz, 2015). It has a high degree of accuracy for WT forecasting (Cai et al., 2018; Piotrowski and Napiorkowski, 2018). In the model, AT is assumed to be the main proxy for the heat flux exchange at the air–water interface, and Q is introduced to account for the water inflows and river thermal inertia. In this study, the most complete eight parameter a_1-a_8 formulation was used, namely the a2s-8 model. The equations for the simple form of the air2stream model with eight parameters version can be expressed as follows:

$$\frac{dT_w}{dt} = \frac{1}{\theta^{a_4}}\left(a_1 + a_2 T_a - a_3 T_w + \theta\left(a_5 + a_6 cos\left(2\pi\left(\frac{t}{t_y} - a_7\right)\right) - a_8 T_w\right)\right) \tag{7}$$

$$\theta = \frac{Q}{\overline{Q}} = \frac{Q}{(t_2 - t_1)^{-1}\int_{t_1}^{t_2} Q(t)dt} \tag{8}$$

where T_w is the water temperature; T_a is the air temperature; t_y denotes the duration of one year in the units used for time, θ is the dimensionless discharge, and $\overline{Q}$ represents a reference value averaged over the time series from t_1 to t_2. More details about the air2stream model are reported in Toffolon and Piccolroaz (2015), and Piotrowski and Napiorkowski (2018).

3.2.2. Random forest (RF)

Random forest (RF) (Breiman, 2001) is one of the most successful predictors based on bagging, a type of ensemble method applied to decision trees. RF integrates a series of independently selected and identically distributed decision trees, and obtains the final single prediction by averaging the outputs of all decision trees in the forest (Svetnik et al., 2003). In RF, each node has d features of the base decision tree, $\sqrt{d}$ features are selected randomly from the feature set (d features), and then the optimal feature is chosen by the Gini impurity index (IG) for the best split threshold (Strobl et al., 2008). The IG can be expressed as follows (Breiman, 2001; Heddam et al., 2020):

$$I_G(f) = 1 - \sum_{i=1}^{m} f_i^2 \tag{9}$$

where f_i denotes the probability of class i at node m, and the best split threshold is indicated by the lowest value of IG. The number of decision trees in the RF model in this study was determined from trial and error by varying the volume of trees from 100 to 300 (Table S2).

3.2.3. Back propagation neural network (BPNN)

Back propagation neural network (BPNN) is a kind of multilayer feedforward neural network (Rumelhart et al., 1988), which consists of an input layer, one or more hidden layers, and an output layer. A three-layer BPNN was used in this study. The main characteristics of BPNN are forward transmission of signals and back propagation of errors. The data transmit layer by layer in the forward transmission process, and the errors generated by the comparisons between the predicted and observed values are back propagated to each layer, thereby adjusting the weights and thresholds of the neurons until the error meets the termination condition. The number of hidden neurons is the primary requirement for the design of BPNN. If there are too many hidden neurons, there is too much flexibility, which usually leads to an overfitting phenomenon, in which a neural network is good at dealing with one dataset at the expense of being very bad at assessing other datasets (Tetko et al., 1995). However, too few hidden neurons restrict the learning capability of the model (Haykin, 1998). Thus, the number of neurons in the hidden layer were determined by the trial and error approach in the range of [1,16] (Yan et al., 2020). The BPNN model was trained over this range at an interval of one, and the number of hidden layer neurons in the hidden layer that had the highest performance (validation period) was retained for the calibrated model (Table S3). Considering both the prediction accuracy and the running time of model, when the prediction accuracy can be achieved with fewer neurons as well as with more neurons, the value of fewer neurons is used.

3.3. Data pre-processing

Because differences in absolute values of the training/testing time-series have a negative effect on the model learning ability, the data on AT, Q, and WT were standardized before training as:

$$x_{norm} = \frac{x_{raw} - \overline{x}}{\sigma_x} \tag{10}$$

where x_{raw} is the original data; x_{norm} is the standardized value of variable x; $\overline{x}$ and σ_x are the average value and standard deviation of $\times$, respectively.

3.4. Model performance indicators

Model performance is evaluated using four different and widely employed metrics for water temperature forecasting (Graf et al., 2019; Piotrowski and Napiorkowski, 2018; Moriasi et al., 2015): the mean absolute error (MAE), the root mean square error (RMSE), Nash-Sutcliffe efficiency coefficient (NSE), the coefficient of determination (R^2), and they are given by:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_p - y_o| \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_p - y_o)^2} \tag{12}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_p - y_o)^2}{\sum_{i=1}^{n}(y_o - \overline{y_o})^2} \tag{13}$$

$$R^2 = \left[\frac{\frac{1}{n}\sum_{i=1}^{n}(y_o - \overline{y_o})(y_p - \overline{y_p})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_o - \overline{y_o})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_p - \overline{y_p})^2}}\right]^2 \tag{14}$$

where n is the number of samples, and y_p and $y_o y_O$ are the predicted and observed values, respectively. The smallest values of MAE and RMSE and values of NSE and R^2 closest to 1 indicate the highest prediction accuracy.

4. Results

4.1. Model performance for daily WT forecasting

Comparison of the performances of the four models in predicting the mean annual daily WT for the calibration and validation periods for the nine gauges indicates that LSTM outperformed the other models for both periods (Table 2). In particular, for the validation periods for all gauges but Yichang, which is influenced strongly by the TGR, the LSTM model has values of MAE $\leq$ 0.83 °C, values of RMSE $\leq$ 1.00 °C, and values of NSE and $R^2 \geq 0.90$. These results are clearly better than those obtained from the a2s-8 (air2stream in 8-parameter version), RF and BPNN models (Fig. 2), which have lower NSE and R^2 values, and higher MAE and RMSE values than the LSTM model. Averaged for the eight gauges (excluding Yichang), the LSTM model lowered mean MAE and RMSE values for the validation period by 13.96% and 13.68%, respectively, compared to a2s-8, by 44.52% and 44.46%, respectively, compared to RF and by 42.66% and 42.32%, respectively, compared to BPNN. Noticeably, the LSTM model performed much better than the a2s-8 model at the Datong, Irondequoit, and Davos stations, and performed slightly better than all other models for the Cedar and Sion stations (Table 2). With regard to the three benchmark models, the a2s-8 model produced higher accuracy with substantially lower values of mean MAE and RMSE during the validation period than the RF model (36.03% and 36.53%, respectively) and the BPNN model (33.81% and 33.94%, respectively). Furthermore, the performance of the RF model was slightly worse than the BPNN model, although mean values of the four metrics showed no significant differences, given the wide distributions

of the metrics for these two models (Fig. 2).

For Yichang, the MAE, RMSE, NSE and R^2 of the LSTM for the calibration period (pre-TGR period 1983–2002) were 0.61 °C, 0.78 °C, 0.98, and 0.98, respectively (Table 2), indicating the LSTM had a superior training effect compared to the other three models. Compared to the other eight gauges with no reservoir effects, the performances of all four models deteriorated substantially for the Yichang gauge in the validation period (post-TGR 2003–2013). This result is not surprising given the strong influence of the TGR on water temperatures in the post-TGR period. The average annual WT at Yichang in post-TGR period increased by 0.67 °C relative to pre-TGR period. This change was accompanied by changes in AT and Q. The average annual AT in the pre-TGR period (1983–2002) was 17.00 °C, and increased to 17.50 °C in the post-TGR period (2003–2013), whereas the average annual Q in the post-TGR decreased by 8.83% compared to the pre-TGR value. These changes challenge the capacity of the calibrated models to forecast conditions following reservoir construction. Nevertheless, the LSTM model still performed better than the other three models in the validation period. Overall, the forecasting performances for all nine gauges are ranked as LSTM > a2s-8 > BPNN > RF.

4.2. Determining impacts of climate and dam construction on WT at Yichang

Because the LSTM model performed best for predicting mean daily water temperatures in the post-TGR (validation period) for the Yichang gauge, this model was used to determine the impacts of climate, as reflected in changes in air temperature and discharge, and dam

Table 2
Performances of LSTM, a2s-8 (air2stream in 8-parameter version), RF and BPNN models in the calibration and validation periods*.

River station	Model version	Calibration (Training)				Validation (Testing)			
		MAE(°C)	RMSE(°C)	NSE	R^2	MAE(°C)	RMSE(°C)	NSE	R^2
Cuntan	LSTM	**0.59**	**0.80**	**0.98**	**0.98**	**0.51**	**0.66**	**0.98**	**0.98**
	a2s-8	0.69	0.90	0.97	0.97	0.58	0.75	0.98	0.98
	RF	0.91	1.17	0.95	0.95	1.23	1.59	0.90	0.90
	BPNN	1.29	1.63	0.90	0.89	1.19	1.55	0.91	0.91
Datong	LSTM	**0.49**	**0.63**	**0.99**	**0.99**	**0.57**	**0.79**	**0.99**	**0.99**
	a2s-8	0.71	0.88	0.99	0.99	0.79	1.03	0.98	0.98
	RF	1.16	1.52	0.96	0.96	1.84	2.38	0.91	0.91
	BPNN	1.62	2.07	0.93	0.93	1.71	2.22	0.92	0.92
Cedar	LSTM	**0.43**	**0.55**	**0.98**	**0.98**	**0.59**	**0.74**	**0.97**	**0.98**
	a2s-8	0.52	0.66	0.97	0.97	0.60	0.75	0.96	0.97
	RF	0.60	0.78	0.95	0.95	0.82	1.04	0.93	0.94
	BPNN	0.76	0.98	0.92	0.93	0.80	1.01	0.94	0.94
Fanno	LSTM	**0.46**	**0.59**	**0.99**	**0.99**	**0.83**	**1.00**	**0.97**	**0.98**
	a2s-8	0.65	0.83	0.97	0.97	0.91	1.13	0.96	0.97
	RF	0.90	1.16	0.95	0.95	1.36	1.71	0.90	0.93
	BPNN	1.16	1.46	0.92	0.91	1.24	1.51	0.93	0.94
Irondequoit	LSTM	**0.52**	**0.67**	**0.99**	**0.99**	**0.62**	**0.78**	**0.99**	**0.99**
	a2s-8	0.65	0.85	0.99	0.99	0.77	1.01	0.98	0.99
	RF	1.14	1.53	0.95	0.95	1.60	2.16	0.91	0.92
	BPNN	1.47	1.95	0.93	0.92	1.46	1.98	0.92	0.93
Yvonand	LSTM	**0.39**	**0.51**	**0.99**	**0.99**	**0.50**	**0.69**	**0.99**	**0.99**
	a2s-8	0.50	0.66	0.99	0.99	0.56	0.80	0.98	0.98
	RF	0.66	0.89	0.98	0.98	1.01	1.33	0.95	0.95
	BPNN	0.93	1.21	0.96	0.96	1.00	1.32	0.95	0.95
Sion	LSTM	**0.43**	**0.56**	**0.93**	**0.93**	**0.51**	**0.73**	**0.90**	**0.90**
	a2s-8	0.45	0.58	0.93	0.93	0.54	0.75	0.89	0.90
	RF	0.42	0.55	0.93	0.93	0.60	0.82	0.87	0.87
	BPNN	0.56	0.73	0.88	0.88	0.62	0.84	0.87	0.88
Davos	LSTM	**0.33**	**0.42**	**0.98**	**0.98**	**0.39**	**0.50**	**0.97**	**0.97**
	a2s-8	0.51	0.63	0.95	0.95	0.53	0.65	0.95	0.95
	RF	0.41	0.55	0.96	0.96	0.59	0.80	0.92	0.92
	BPNN	0.58	0.77	0.93	0.93	0.58	0.78	0.93	0.93
Yichang	LSTM	**0.61**	**0.78**	**0.98**	**0.98**	**1.64**	**2.04**	**0.85**	**0.86**
	a2s-8	0.72	0.90	0.97	0.97	1.76	2.21	0.82	0.85
	RF	0.89	1.15	0.96	0.96	2.15	2.70	0.74	0.77
	BPNN	1.23	1.57	0.92	0.93	2.15	2.69	0.74	0.77

Note*: The best performances for each gauge are in bold; Yichang station were selected for assessing the performance of LSTM to detect WT variations by climate change and the TGR based on the actual observations.
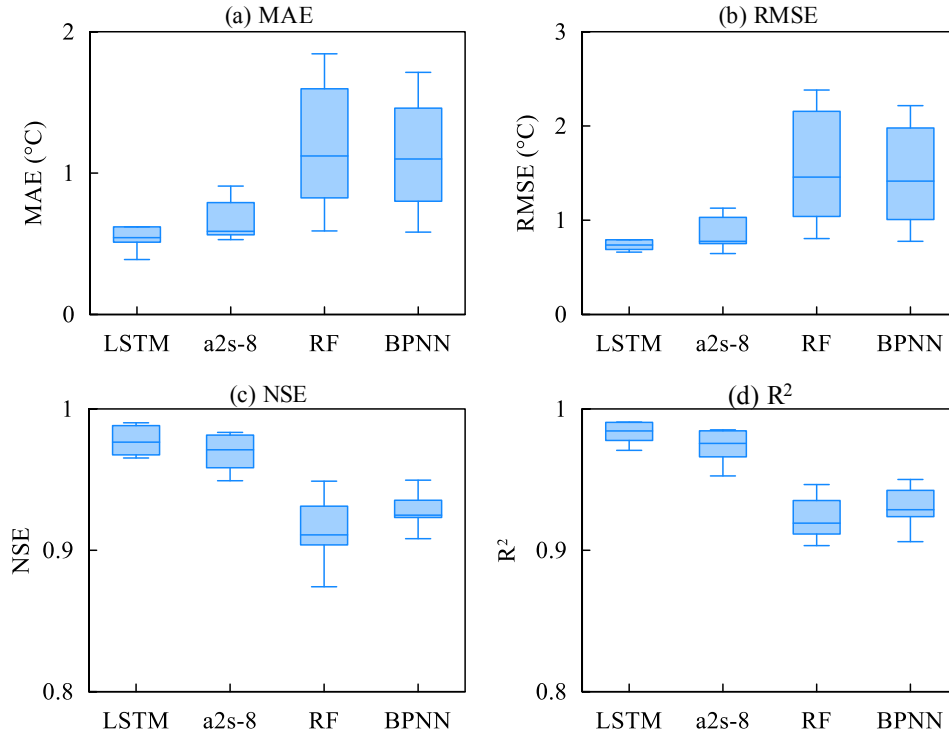
Fig. 2. Boxplots of the MAE, RMSE, NSE, and R2 for the LSTM, a2s-8, RF and BPNN models during the validation period.

construction on the mean daily thermal regime at this location. To conduct this analysis, the post-TGR period was divided into three sub-periods (SP): SP1, 135 m impoundment (Jun. 2003-Sep. 2006), SP2, 156 m impoundment (Oct. 2006-Sep. 2008) and SP3, 175 m impoundment (Oct. 2008-Dec. 2013). For all three subperiods, the annual pattern of mean daily AT fluctuated around the pattern of mean daily AT for the pre-TGR period (Fig. 3(a)-(c)) with mean fluctuations of 0.59 °C in SP1, 0.74 °C in SP2, and 0.39 °C in SP3. Observed mean daily Q values for the three post-dam subperiods (Q$_{post-TGR}$, red lines) decreased in the flood season (June to October) and increased in the dry season (January-March) relative to the pre-TGR period (Q$_{pre-TGR}$, green lines), with the most pronounced differences occurring for SP3 (Fig. 3 (d)-(f)). For WT, the annual pattern of mean daily values following the closure of the TGR reservoir shows that WT relative to the pre-TGR period increased from September-February and decreased from March through June (Fig. 3 (g)-(i)).

To reconstruct WT in the absence of the reservoir, which mainly affects discharge of the Yangtze River, measured values of Q at Yichang gauge were replaced in the calibrated LSTM model with the daily inflow discharge to the TGR (Q$_{post-TGR (non-TGR)}$), the Q that would have occurred at Yichang in the absence of the TGR. These inflow values of Q, reported by the China Three Gorges Operation, are derived from information on the change in water storage in the TGR, a storage-capacity curve for the reservoir, and the outflow of the TGR. The reconstructed natural discharge (Q$_{post-TGR (non-TGR)}$, blue lines) for SP1 did not show obvious differences compared with the observed values of Q (Q$_{post-TGR}$, red lines) for this period (Fig. 3 (d)). The Q$_{post-TGR (non-TGR)}$ for SP2 exhibited an increase in October, but the rest of the period was similar to Q$_{post-TGR}$ (Fig. 3 (e)). Notably, the reconstructed discharges (Q$_{post-TGR (non-TGR)}$) for SP3 are greater than observed Q$_{post-TGR}$ for the flood season (June-October) and lower than the observed Q$_{post-TGR}$ in the dry season (January-March) (Fig. 3 (f)). This increase in the range of mean daily discharge over the year for the reconstructed inflow Q$_{post-TGR (non-TGR)}$ compared to the observed Q$_{post-TGR}$ reflects the influence of the TGR on flow extremes as it is operated for flood control and power generation.

The reconstructed discharges (Q$_{post-TGR (non-TGR)}$) along with

measured values of AT were used to predict the WT for the 2003–2013 period that would have occurred in the absence of the TGR. Based on these predictions the impacts of the TGR (Δ_{TGR}) and climate change (Δ_{CLI}) on mean daily WT at Yichang were quantified as:

$$\Delta_{TOT} = WT_{obs,post-TGR} - WT_{obs,pre-TGR} \quad (15)$$

$$\Delta_{CLI} = WT_{sim,post-TGR} - WT_{sim,pre-TGR} \quad (16)$$

$$\Delta_{TGR} = WT_{obs,post-TGR} - WT_{sim,post-TGR} \quad (17)$$

$$\varepsilon = WT_{sim,pre-TGR} - WT_{obs,pre-TGR} \quad (18)$$

where Δ_{TOT} represents the total change of WT before and after dam; Δ_{CLI} is the contribution of changes in AT and natural discharge to change in WT; Δ_{TGR} is the change in WT ascribed to the construction of TGR (assuming no other human perturbations except for the TGR); ε is the bias of model (i.e., mean error); $WT_{obs,post-TGR}$ and $WT_{obs,pre-TGR}$ are observed values of WT in the post-TGR and pre-TGR periods, respectively; $WT_{sim,post-TGR}$ is the simulated WT in the absence of the TGR, using the LSTM model with the reconstructed discharge Q$_{post-TGR (non-TGR)}$ and the observed AT as the inputs; $WT_{sim,pre-TGR}$ is the simulated WT in the pre-TGR period based on the observed Q and AT; and $\Delta_{TOT} = \Delta_{CLI} + \Delta_{TGR} + \varepsilon$. With regard to determining the influence of the TGR on WT, the model can be considered reliable if $|\varepsilon| \ll |\Delta_{TGR}|$. Under these conditions, the error associated with model bias is much smaller than the predicted impact of the TGR on WT.

The results are presented in a compact form by the climatological reference year, in which the value of a variable for each day of the year is the average of all available observed values for that day over the period of record (29 February of leap year was not included) (Cai et al., 2018). Simulations using the LSTM model reveal that the absolute value of ε (0.02 °C) is small compared to absolute values of Δ_{TGR}, indicating that the model is reliable for determining the influence of the TGR on WT. The largest absolute values of ε for particular months occur in August and September, but generally $|\varepsilon| < |0.5\Delta_{TGR}|$ (Table 3).

A comparison between the observed WT in the post-TGR period, the predicted WT after TGR, and the predicted WT in the assumed case
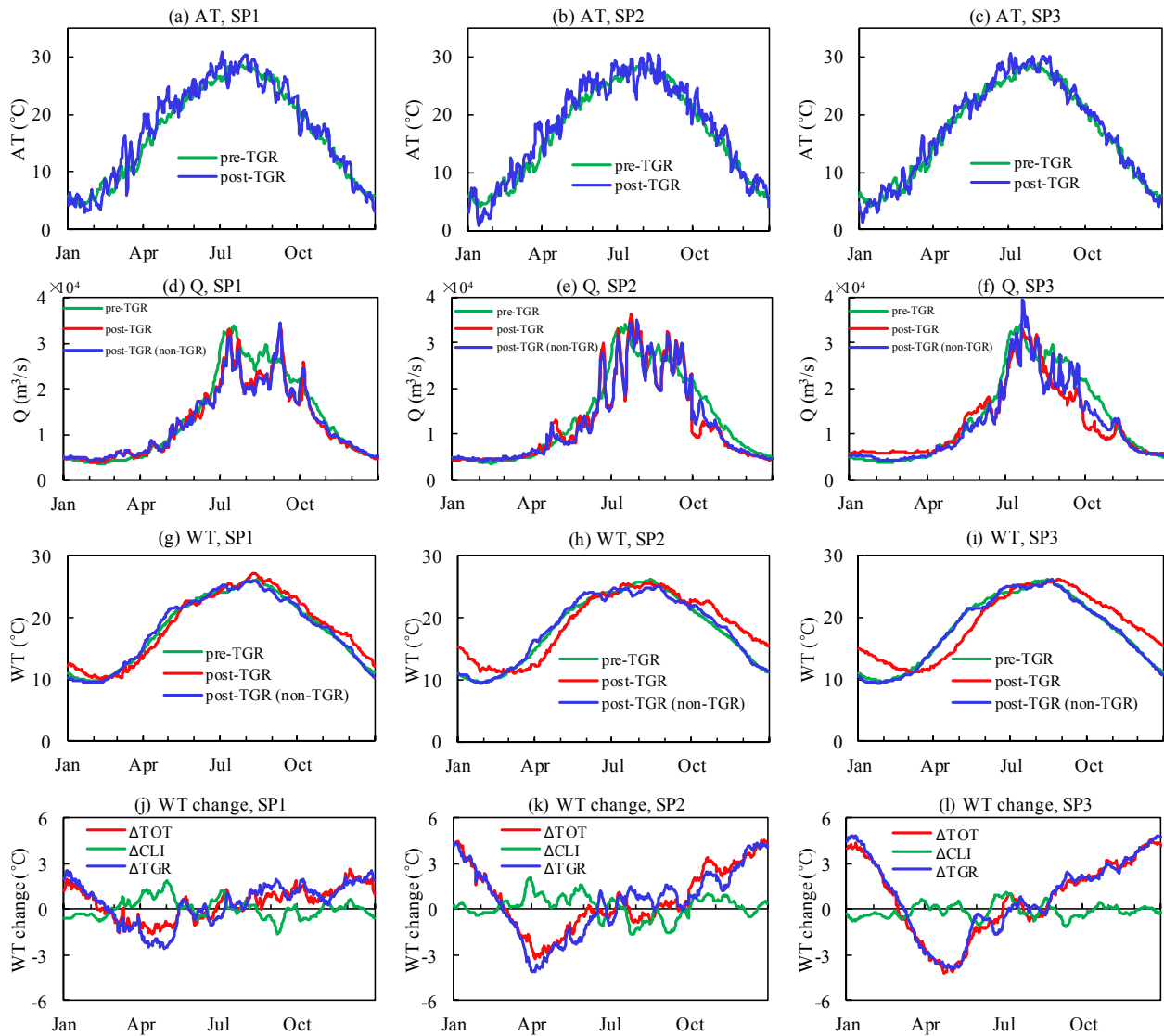
Fig. 3. Seasonal dynamics (climatological year) of (a-c) AT, (d-f) Q, and (g-i) WT at Yichang station during pre-TGR and three subperiods (SP1, SP2, SP3) in the post-TGR periods*. Note*: Plots (j-l) show the WT changes between the pre-TGR and three subperiods, respectively, indicating the forcing by climate change (green lines) and TGR (blue lines). The blue lines in (d-f) (Q$_{post-TGR\ (non-TGR)}$) represent the reconstructed Q in the post-TGR, which are assumed to be the daily inflow to TGR. The blue lines in (g-i) (WT$_{post-TGR\ (non-TGR)}$) represent the reconstructed WT using the LSTM model, with the reconstructed Q and observed AT as input variables. The green lines (WT$_{pre-TGR}$) and the red lines (WT$_{post-TGR}$) in (g-i) denote the observed WT in the pre-TGR and post-TGR periods, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Changes of WT at Yichang station in the post-TGR period (135 m, 156 m and 175 m impoundments) relative to the pre-TGR period (1983–2002) caused by changes in AT (ΔCLI) and by the construction of the TGR (ΔTGR) *.

Period	WT change	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Annual
SP1(135 m impoundment, Jun.2003-Sep.2006)	Δ_{TOT}	1.35	0.10	−0.89	**−1.20**	−0.27	−0.49	0.42	0.73	0.99	0.90	1.32	**1.93**	0.41
	Δ_{CLI}	−0.43	−0.43	0.33	**1.12**	0.59	−0.05	0.34	−0.12	**−0.57**	−0.46	0.17	0.03	0.05
	Δ_{TGR}	1.70	0.34	−1.06	**−2.19**	−0.72	−0.42	0.19	1.13	1.26	1.25	1.05	**1.98**	0.38
	ε	0.07	0.19	−0.16	−0.13	−0.15	−0.03	−0.10	−0.28	0.30	0.11	0.09	−0.09	−0.02
SP2(156 m impoundment, Oct.2006-Sep.2008)	Δ_{TOT}	3.39	1.26	−1.19	**−2.71**	−1.30	−0.21	0.11	−0.33	0.29	2.44	2.71	**3.89**	0.70
	Δ_{CLI}	−0.01	−0.07	0.67	0.74	**0.96**	0.12	−0.27	−0.70	**−0.86**	0.69	0.52	−0.01	0.15
	Δ_{TGR}	3.33	1.14	−1.69	**−3.32**	−2.12	−0.30	0.49	0.64	0.84	1.64	2.09	**3.99**	0.56
	ε	0.07	0.19	−0.16	−0.13	−0.15	−0.03	−0.10	−0.28	0.30	0.11	0.09	−0.09	−0.02
SP3(175 m impoundment, Oct.2008-Dec.2013)	Δ_{TOT}	3.80	1.75	−1.33	**−3.46**	−2.91	−0.96	0.37	−0.01	1.62	2.35	2.98	**4.15**	0.69
	Δ_{CLI}	−0.52	−0.37	−0.08	0.05	−0.12	0.01	**0.34**	0.01	**−0.60**	−0.01	−0.11	−0.03	−0.12
	Δ_{TGR}	4.24	1.93	−1.09	**−3.38**	−2.64	−0.94	0.13	0.26	1.92	2.24	3.00	**4.27**	0.83
	ε	0.07	0.19	−0.16	−0.13	−0.15	−0.03	−0.10	−0.28	0.30	0.11	0.09	−0.09	−0.02

*Note: Δ_{TOT} is the total change of WT before and after the operation of TGR; ε denotes the model bias; the maximum contributions in each category are in bold; and all values are in °C.

without TGR was shown in Fig. 4. It can be clearly seen that the predicted WT in the post-TGR (WT$_{post-TGR (sim)}$, green lines) are not close to the observed values after TGR (WT$_{post-TGR}$, red lines), and the green lines are also significant substantially different from the blue lines (the predicted WT in the assumed case without TGR). The LSTM was trained based on the statistical relationships for the observed WT-observed Q and the observed WT-observed AT in the pre-TGR period. Although the observed Q did not show significant alterations to the reconstructed Q in the SP1 and SP2, but they are significantly different in the SP3, the period of strong forcing by meaning the forcing by TGR. Furthermore, the statistical relationships between WT-Q was clearly disturbed in the post-TGR (Liu et al., 2018). Thus, the WT$_{post-TGR (sim)}$ (green lines) predicted by the LSTM model which was trained by the pre-TGR data showed obvious differences both to the WT$_{post-TGR}$ (red lines) and the WT$_{post-TGR (non-TGR)}$ (blue lines). This indicated the importance of reconstructing the natural discharges that have occurred without TGR before predicting the WT in the absence of the dam.

The operation of TGR (Δ_{TGR}) is the major factor contributing to change in mean annual water temperature before and after the construction of the reservoir, accounting for 93% of the increase in WT (Δ_{TOT}) for SP1, 81% of the increase for SP2, and 119% of the increase for SP3 (Table 3). By contrast changes in WT related to AT and natural discharge (climate) range from 11% (SP1) to 22% (SP2). Thus, most change in WT at Yichang before and after the construction of the TGR can be attributed to the influence of the reservoir, rather than to changes in climate.

Reservoir construction also is the major factor contributing to change in seasonal variation of WT before and after reservoir construction. Cooling in warm seasons (March-June) and warming in cold seasons (September-February) are mainly attributable to the operation of TGR, consistent with the results of Cai et al (2018). The largest effects of forcing by the TGR occur in April and December during the 175 m impoundment (Table 3). In April values of Δ_{TGR} of -2.19 °C (SP1), -3.32 °C (SP2), and -3.38 °C (SP3) accounted for 183% (SP1), 123% (SP2) and 98% (SP3) of total cooling, whereas in December values of Δ_{TGR} of 1.98 °C (SP1), 3.99 °C (SP2), and 4.27 °C (SP3) accounted for 103% of total warming in all three post-TGR subperiods.

Variations in mean daily WT induced by climate change (Fig. 3 (j)-(l)) (Δ_{CLI}, green lines) fluctuate above and below the value of zero throughout the year, resulting in small average annual Δ_{CLI} values of 0.05 °C (SP1), 0.15 °C (SP2), and -0.12 °C (SP3) (Table 3). Climate-induced change is most distinct in April (average $\Delta_{CLI} = 0.64$ °C for the three subperiods) and September (average $\Delta_{CLI} = -0.68$ °C for the three subperiods). In general, climate change had its largest impact on WT at Yichang during SP2 with a range of -1.67 °C $< \Delta_{CLI} < 2.04$ °C. The smallest effect was for SP3, where -1.18 °C $< \Delta_{CLI} < 1.15$ °C.

The annual pattern of changes in WT induced by the TGR (Δ_{TGR}, blue lines) matches the total WT changes (Δ_{TOT}, red lines) (Fig. 3(j)-(l)) for all three subperiods, confirming the major role of the TGR in producing changes in water temperature. Patterns of Δ_{TGR} by day for each month

and each subperiod (Fig. 5) illuminate in greater detail the monthly patterns (Table 3). These patterns confirm that the TGR acted as a cold source from the beginning of March to early July and a warm source from mid-July to late February. This pattern reflects the large thermal inertia of the TGR, which delays the seasonal circle of WT (Toffolon et al., 2010; Cai et al., 2018). More precisely, the TGR effects can be roughly divided into four stages: increasing cooling from March through mid-May, decreasing cooling from mid-May through early July, increasing warming from mid-July to December, and decreasing warming during January and February. The largest cooling effects occurred in early April to early May and the most substantial warming happened in late December and early January. The effect of increasing the water level, or volume of water, in the TGR is evident for cooling in that the amount of maximum cooling increases between SP1 and the two later subperiods by about 1.5 °C. Moreover, between SP2 and SP3 the timing of maximum cooling shifts from the beginning of April to the beginning of May. The duration of maximum cooling also increases from SP2 to SP3. In general, warming from mid-July to late February increases as reservoir level increases (SP1 to SP3) with this effect most pronounced in January and February when temperature increases for SP3 clearly exceed those for SP1 and SP2. These patterns generally reflect an increase in thermal inertia as the volume of the reservoir increases (Toffolon et al., 2010).

5. Discussion

A comparison between the deep learning method (LSTM) and three benchmark models (air2stream, RF and BPNN) at nine gauges on seven different rivers in different physiographic settings demonstrates the improved forecasting accuracy of LSTM for predicting mean daily water temperatures based on air temperature and discharge. Previous studies using data for six of the nine gauges (three Swiss and three USA gauges) (Zhu et al., 2019b; Piotrowski and Napiorkowski, 2018) examined in the present study revealed that the air2stream model outperformed conventional data-driven methods (feedforward neural network, Gaussian process regression, decision tree, the Mohseni et al. (1998) model, the van Vliet et al. (2006) model, linear regression). The present study compared the air2stream model (8-parameter version) to Random Forest (RF) and Back Propagation Neural Network (BPNN). The a2s-8 model outperformed than the RF and BPNN at the same studied gauges, but did not perform as well as the LSTM. In contrast, although the improvements by LSTM for the MAE and RMSE relative to the a2s-8 were relatively small at two of the gauges (1.7%-5.6% and 1.2–2.7%, respectively), the LSTM decreased the MAE and RMSE at four of the gauges by 8.8%-26.4% and 11.5%–23.1%, respectively. The reason why the LSTM outperforms than the benchmark models is the input and output of LSTM are considered as a two-time-series sequence. The LSTM is deep in time, which not only updates the past state, but also can learn when to forget and how long to keep the state information (Shen, 2018).

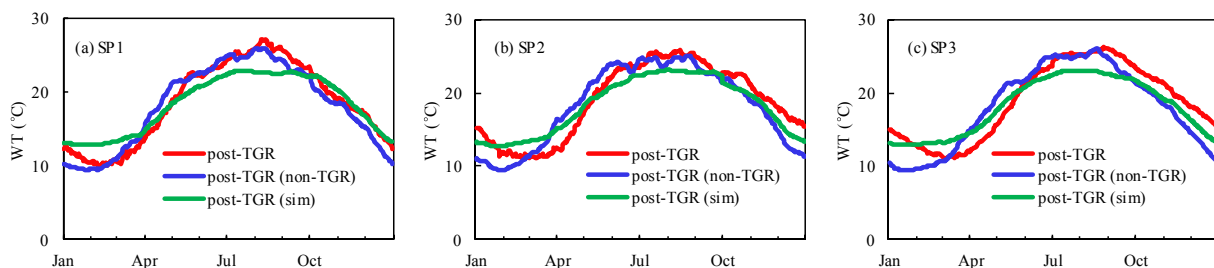The calibrated LSTM model successfully reconstructed the daily

Fig. 4. Comparisons between the observed WT after TGR (post-TGR, red lines), the predicted WT after TGR (post-TGR (sim), green lines) and the predicted WT in the assumed case without TGR (post-TGR (non-TGR), blue lines) *. Note*: The WT$_{post-TGR (sim)}$ was predicted by the LSTM using the observed Q and AT as input variables. The WT$_{post-TGR (non-TGR)}$ was predicted by the LSTM with the reconstructed Q and the observed AT as inputs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
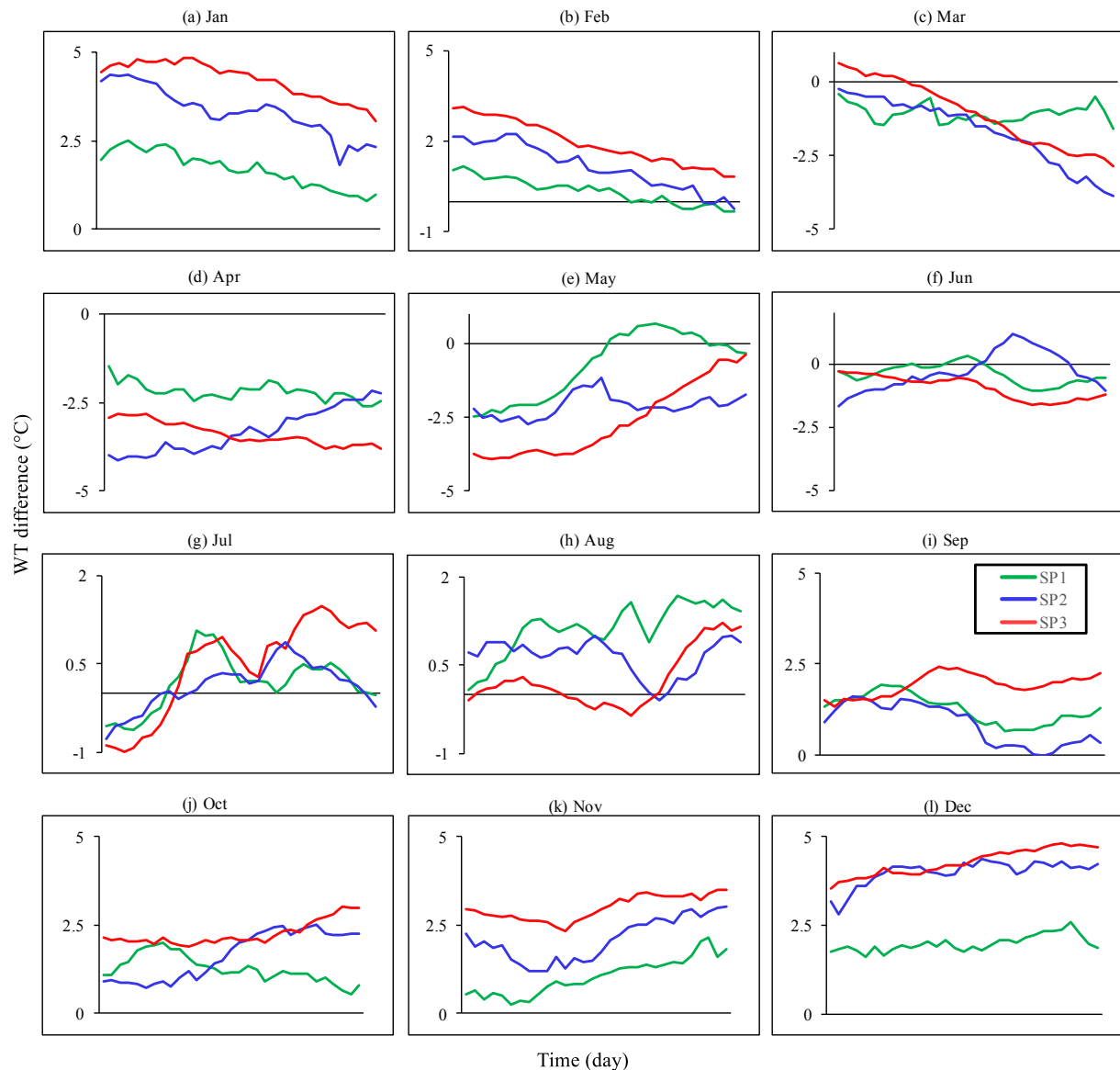
Fig. 5. WT differences induced by the TGR in three impoundments relative to the pre-TGR period on the daily scale.

water temperatures at Yichang on the Yangtze River that would occur in the absence of TGR, which helped to quantify the influence of changes in air temperature and discharge, surrogates for the influence of climate, versus the influence of the TGR, on changes in mean daily water temperatures at Yichang. Past studies have revealed that TGR has changed the thermal regime of the Yangtze River on the annual and seasonal scale using the air2stream model and non-reconstructed Q (Cai et al., 2018), or using the linear regression model and reconstructed Q from a conventional artificial neural network model (Tao et al., 2020). The differences between the observed WT in the post-TGR, the predicted WT after TGR, and the predicted WT in the assumed case without TGR demonstrated it's necessary to reconstruct the natural discharge Q that would occur in the absence of TGR before reconstructing the WT without the dam (Fig. 4). In addition, the results of the present study confirm the efficiency of LSTM to reconstruct natural thermal conditions in the absence of a large reservoir. These results also indicate that that climate change has had a small impact on mean daily WT change at Yichang (Fig. 3) – a finding consistent with previous efforts to unravel the effects of climate-related and dam-related effects on WT at Yichang (Tao et al., 2020). The daily timescale resolution of the analysis shows that the in-ertial effects of the reservoir affect the thermal regime of the Yangtze

river under its present operational stage (175 m) by producing periods of increasing cooling in the spring, decreasing cooling in early summer, increasing warming from mid-summer to early winter, and decreasing warming in mid-late winter. The greatest cooling occurs in mid-April to mid-May (cooling) and the greatest warming in late December and early January. In short, LSTM acts as a useful new tool to predict the daily water temperatures and to detect at high temporal resolution the im-pacts of climate change and reservoirs on the thermal regime of rivers.

The analysis in the present study assumes that the construction of TGR is the main significant anthropogenic perturbation during the post-TGR period. Previous work that has attempted to decompose the changes in the thermal regime of the Yangtze River at Yichang suggests that other factors such as land use change or sewage discharges may have impacts on WT (Cai et al., 2018; Tao et al., 2020). The effect of sewage discharge has been estimated as<0.1 °C (Cai et al., 2018), which is small compared to the total WT change before and after dam. The effect of other factors, such as land use, are difficult to quantify. Ac-cording to the Tao et al. (2020) model, warming in WT at Yichiang related to the TGR occurred only from December through February in SP3; otherwise, the TGR produced cooling during all seasons in SP1, SP2, and SP3. These results differ from those of the LSTM model, which

indicate that the TGR has produced warming in WT over a substantial portion of the year (July through February), especially in SP3. The Tao et al. (2020) model attributed impacts related to factors other than AT, Q, and the TGR, including other human impacts, to a factor $\Delta_{\beta0}$, which contributed substantially to WT variations. However, the extent to which the factor $\Delta_{\beta0}$ isolates non-reservoir related effects, specifically other human-related effects, such as land-use changes, remains uncertain and is in need of further investigation. Differences between the predictions of the Tao et al. (2020) model and LSTM model may also partly reflect modelling uncertainty (e.g., the selection of prediction model, the scenario reconstruction in the absence of TGR, the model bias) and the time scale used for analysis (monthly versus daily). Full consideration of the impacts of other factors, such as land use change, on water temperature regimes requires comprehensive data on this change, which is beyond the scope of the present study, but which can be considered in future research. Although land use effects could be important, the close association of the pattern of total change in WT with the pattern of change associated with the TGR (Fig. 3 (j)-(l)) implies that the total change is associated mainly with the effect of the reservoir.

6. Conclusions

This study has verified that a deep learning model, LSTM, is a powerful tool for accurately forecasting river water temperatures and for reconstructing natural thermal regimes at a daily timescale that would occur in the absence of a large dam. Major conclusions include:

- LSTM outperformed three benchmark models (air2stream, RF, BPNN) based on metrics of predictive accuracy for nine gauges on seven rivers in China, the United States, and Switzerland.
- LSTM predictions of river water temperatures in the absence of the Three Gorges Reservoir indicate that the reservoir has had a much larger impact on changes in temperature than have changes in climate.
- The influence of the dam on water temperatures has been greatest in the spring (mid-April to mid-May) when it has produced cooling of about 4 °C, and in early winter (late December to early January), when it has produced warming of about the same amount.

Additional testing and evaluation of the deep learning approach to forecasting river water temperatures is needed to determine the extent to which this approach can be generalized. As with all data-driven approaches, the capability of the model to predict accurately beyond the domain of the data used in model calibration remains unclear. Nevertheless, the present study indicates that in cases where forecasting is a primary aim, the method provides accurate predictions when calibrated using historical river data for on water temperature, discharge and air temperature. It also shows that the method is useful for deciphering how human-related disturbances, such as reservoir construction, have influenced thermal regimes of rivers. In this sense, the deep learning model not only is a powerful forecasting tool, but also is capable of providing explanatory insight into causes in changes in river water temperature.

CRediT authorship contribution statement

Rujian Qiu: Methodology, Formal analysis, Writing - original draft. **Yuankun Wang:** Conceptualization, Supervision. **Bruce Rhoads:** Validation, Formal analysis. **Dong Wang:** Project administration. **Wenjie Qiu:** Data curation, Investigation. **Yuwei Tao:** Investigation. **Jichun Wu:** Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2021.126016.

References

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Buentello, J.A., Gatlin III, D.M., Neill, W.H., 2000. Effects of water temperature and dissolved oxygen on daily feed consumption, feed utilization and growth of channel catfish (Ictalurus punctatus). Aquaculture 182 (3–4), 339–352. https://doi.org/10.1016/S0044-8486(99)00274-4.

Cai, H., Piccolroaz, S., Huang, J., Liu, Z., Liu, F., Toffolon, M., 2018. Quantifying the impact of the Three Gorges Dam on the thermal dynamics of the Yangtze River. Environ. Res. Lett. 13 (5), 054016. https://doi.org/10.1088/1748-9326/aab9e0.

CAISSIE, D., 2006. The thermal regime of rivers: a review. Freshw. Biol. 51 (8), 1389–1406. https://doi.org/10.1111/fwb.2006.51.issue-810.1111/j.1365-2427.2006.01597.x.

Cha, Y., Cho, K.H., Lee, H., Kang, T., Kim, J.H., 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.

Chai, Y.F., Li, Y.T., Yang, Y.P., Zhu, B.Y., Li, S.X., Xu, C., Liu, C.C., 2019. Influence of climate variability and reservoir operation on streamflow in the Yangtze River. Sci. Rep. 9 (1), 1–10. https://doi.org/10.1038/s41598-019-41583-6.

Cole, J.C., Maloney, K.O., Schmid, M., McKenna Jr, J.E., 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.

DeWeber, J.T., 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.Z., Shrestha, N.K., Wang, J.Y., 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.

Durance, I., Ormerod, S.J., 2009. Trends in water quality and discharge confound longterm warming effects on river macroinvertebrates. Freshw. Biol. 54 (2), 388–405. https://doi.org/10.1111/j.1365-2427.2008.02112.x.

Ficklin, D.L., Stewart, I.T., Maurer, E.P., 2013. Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California. Water Resour. Res. 49 (5), 2765–2782. https://doi.org/10.1002/wrcr.20248.

Grabowski, Z.J., Watson, E., Chang, H., 2016. Using spatially explicit indicators to investigate watershed characteristics and stream temperature relationships. Sci. Total Environ. 551–552, 376–386. https://doi.org/10.1016/j.scitotenv.2016.02.042.

Graf, R., Zhu, S., Sivakumar, B., 2019. Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. J. Hydrol. 578, 124115. https://doi.org/10.1016/j.jhydrol.2019.124115.

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 (5), 1379–1394. https://doi.org/10.1007/s11269-014-0557-7.

Haykin, S., 1998. Neural networks: A comprehensive foundation, 2nd Ed. Prentice Hall PTR, Upper Saddle River.

Heddam, S., Ptak, M., Zhu, S., 2020. Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. J. Hydrol. 588, 125130. https://doi.org/10.1016/j.jhydrol.2020.125130.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Isaak, D.J., Wollrab, S., Horan, D., Chandler, G., 2012. Climate change effects on stream and river temperatures across the northwest U.S. from 1980–2009 and implications for salmonid fishes. Clim. Change 113 (2), 499–524. https://doi.org/10.1007/s10584-011-0326-z.

Jackson, F.L., Fryer, R.J., Hannah, D.M., Millar, C.P., Malcolm, I.A., 2018. A spatio-temporal statistical model of maximum daily river temperatures to inform the management of Scotland's Atlantic salmon rivers under climate change. Sci. Total Environ. 612, 1543–1558. https://doi.org/10.1016/j.scitotenv.2017.09.010.

Kaushal, S.S., Likens, G.E., Jaworski, N.A., Pace, M.L., Sides, A.M., Seekell, D., Belt, K.T., Secor, D.H., Wingate, R.L., 2010. Rising stream and river temperatures in the United States. Front. Ecol. Environ. 8 (9), 461–466. https://doi.org/10.1890/090037.

Khosravi, K., Mao, L., Kisi, O., Yaseen, Z.M., Shahid, S., 2018. Quantifying hourly suspended sediment load using data mining models: Case study of a glacierized Andean catchment in Chile. J. Hydrol. 567, 165–179. https://doi.org/10.1016/j.jhydrol.2018.10.015.

Lessard, J.L., Hayes, D.B., 2003. Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. River Res. Appl. 19 (7), 721–732. https://doi.org/10.1002/(ISSN)1535-146710.1002/rra.v19:710.1002/rra.713.

Li, Q.F., Yu, M.X., Lu, G.B., Cai, T., Bai, X., Xia, Z.Q., 2011. Impacts of the Gezhouba and Three Gorges reservoirs on the sediment regime in the Yangtze River China. J. Hydrol. 403 (3–4), 224–233. https://doi.org/10.1016/j.jhydrol.2011.03.043.

Liu, P., Li, L.P., Guo, S.L., Xiong, L.H., Zhang, W., Zhang, J.W., Xu, C.Y., 2015. Optimal design of seasonal flood limited water levels and its application for the Three Gorges Reservoir. J. Hydrol. 527, 1045–1053. https://doi.org/10.1016/j.jhydrol.2015.05.055.

Liu, Z., Chen, X., Liu, F., Lin, K., He, Y., Cai, H., 2018. Joint dependence between river water temperature, air temperature, and discharge in the Yangtze River: the role of the Three Gorges Dam. J. Geophys. Res. Atmos. 123 (21), 11,938–11,951. https://doi.org/10.1029/2018JD029078.

Maheu, A., Caissie, D., St-Hilaire, A., El-Jabi, N., 2014. River evaporation and corresponding heat fluxes in forested catchments. Hydrol. Process. 28 (23), 5725–5738. https://doi.org/10.1002/hyp.10071.

Marzadria, A., Toninab, D., Bellina, A., 2013. Quantifying the importance of daily stream water temperature fluctuations on the hyporheic thermal regime: implication for dissolved oxygen dynamics. J. Hydrol. 507, 241–248. https://doi.org/10.1016/j.jhydrol.2013.10.030.

Matthews, K.R., Berg, N.H., 1997. Rainbow trout responses to water temperature and dissolved oxygen stress in two southern California stream pools. J. Fish Biol. 50 (1), 50–67. https://doi.org/10.1111/j.1095-8649.1997.tb01339.x.

Mohseni, O., Stefan, H.G., 1999. Stream temperature/air temperature relationship: a physical interpretation. J. Hydrol. 218 (3–4), 128–141. https://doi.org/10.1016/S0022-1694(99)00034-7.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: performance measures and evaluation criteria. Trans. ASABE 58 (6), 1763–1785. https://doi.org/10.13031/trans.58.10715.

Mozer, M.C., 1989. A focused backpropagation algorithm for temporal pattern recognition. Complex Systems. 3 (4), 349–381.

Olah, C., 2015. Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/, Accessed date: 19 July 2018.

Olden, J.D., Naiman, R.J., 2010. Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. Freshw. Biol. 55 (1), 86–107. https://doi.org/10.1111/j.1365-2427.2009.02179.x.

Piotrowski, A.P., Napiorkowski, M.J., Napiorkowski, J.J., 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.

Piotrowski, A.P., Napiorkowski, J.J., 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, A.P., Napiorkowski, J.J., 2019. Simple modifications of the nonlinear regression stream temperature model for daily data. J. Hydrol. 572, 308–328. https://doi.org/10.1016/j.jhydrol.2019.02.035.

Poole, G.C., Berman, C.H., 2001. An ecological perspective on in-stream temperature: natural heat dynamics and mechanisms of human-caused thermal degradation. Environ. Manage. 27 (6), 787–802. https://doi.org/10.1007/s002670010188.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. Learning representations by back-propagating errors. Nature 323 (6088), 696–699. https://doi.org/10.1038/323533a0.

Sahoo, G.B., Schladow, S.G., Reuter, J.E., 2009. Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. J. Hydrol. 378 (3–4), 325–342. https://doi.org/10.1016/j.jhydrol.2009.09.037.

Shen, C.P., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour. Res. 54 (11), 8558–8593. https://doi.org/10.1029/2018WR022643.

St-Hilaire, A., Ouarda, T.B.M.J., Bargaoui, Z., Daigle, A., Bilodeau, L., 2012. Daily river water temperature forecast model with a k-nearest neighbour approach. Hydrol. Process. 26 (9), 1302–1310. https://doi.org/10.1002/hyp.v26.910.1002/hyp.8216.

Sohrabi, M.M., Benjankar, R., Tonina, D., Wenger, S.J., Isaak, D.J., 2017. Estimation of daily stream water temperatures with a Bayesian regression approach. Hydrol. Process. 31 (9), 1719–1733. https://doi.org/10.1002/hyp.v31.910.1002/hyp.11139.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinf. 9 (1), 307. https://doi.org/10.1186/1471-2105-9-307.

Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43 (6), 1947–1958. https://doi.org/10.1021/ci034160g.

Tetko, I.V., Livingstone, D.J., Luik, A.I., 1995. Neural network studies. 1. Comparison of overfitting and overtraining. J. Chem. Inf. Comput. Sci. 35 (5), 826–833. https://doi.org/10.1021/ci00027a006.

Tao, Y., Wang, Y., Rhoads, B., Wang, D., Ni, L., Wu, J., 2020. Quantifying the impacts of the Three Gorges Reservoir on water temperature in the middle reach of the Yangtze River. J. Hydrol. 582, 124476. https://doi.org/10.1016/j.jhydrol.2019.124476.

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

Toffolon, M., Siviglia, A., Zolezzi, G., 2010. Thermal wave dynamics in rivers affected by hydropeaking. Water Resour. Res. 46 (8), W08536. https://doi.org/10.1029/2009WR008234.

Van Vliet, M.T.H., Franssen, W.H.P., Yearsley, J.R., Ludwig, F., Haddeland, I., Lettenmaier, D.P., Kabat, P., 2013. Global river discharge and water temperature under climate change. Global Environ. Change 23 (2), 450–464. https://doi.org/10.1016/j.gloenvcha.2012.11.002.

Van Vliet, M.T.H., Ludwig, F., Zwolsman, J.J.G., Weedon, G.P., Kabat, P., 2011. Global river temperatures and sensitivity to atmospheric warming and changes in river flow. Water Resour. Res. 47 (2), W02544. https://doi.org/10.1029/2010WR009198.

Vega, M., Pardo, R., Barrado, E., Debán, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res. 32 (12), 3581–3592. https://doi.org/10.1016/S0043-1354(98)00138-9.

Wanders, N., Wada, Y., 2015. Human and climate impacts on the 21st century hydrological drought. J. Hydrol. 526, 208–220. https://doi.org/10.1016/j.jhydrol.2014.10.047.

Wanders, N., van Vliet, M.T.H., Wada, Y., Bierkens, M.F.P., van Beek, L.P.H., 2019. High-resolution global water temperature modelling. Water Resour. Res. 55 (4), 2760–2778. https://doi.org/10.1029/2018WR023250.

Wang, Y.K., Rhoads, B.L., Wang, D., 2016. Assessment of the flow regime alterations in the middle reach of the Yangtze River associated with dam construction: potential ecological implication. Hydrol. Process. 30, 3949–3966. https://doi.org/10.1002/hyp.10921.

Wang, Y.K., Rhoads, B.L., Wang, D., Wu, J.C., Zhang, X., 2018. Impacts of large dams on the complexity of suspended sediment dynamics in the Yangtze River. J. Hydrol. 558, 184–195. https://doi.org/10.1016/j.jhydrol.2018.01.027.

Ward, J.V., Stanford, J.A., 1982. Thermal responses in the evolutionary ecology of aquatic insects. Annu. Rev. Entomol. 27 (1), 97–117. https://doi.org/10.1146/annurev.en.27.010182.000525.

Webb, B.W., Clack, P.D., Walling, D.E., 2003. Water-air temperature relationships in a Devon river system and the role of flow. Hydrol. Process. 17 (15), 3069–3084. https://doi.org/10.1002/(ISSN)1099-108510.1002/hyp.v17:1510.1002/hyp.1280.

Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. Proc. IEEE 78 (10), 1550–1560. https://doi.org/10.1109/5.58337.

Xiao, C., Chen, N., Hu, C., Wang, K.e., Gong, J., Chen, Z., 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. Remote Sens. Environ. 233, 111358. https://doi.org/10.1016/j.rse.2019.111358.

Yan, H., Zhang, J., Zhou, N., Li, M., 2020. Application of hybrid artificial intelligence model to predict coal strength alteration during CO2 geological sequestration in coal seams. Sci. Total Environ. 711, 135029. https://doi.org/10.1016/j.scitotenv.2019.135029.

Yi, Y.J., Wang, Z.Y., Yang, Z.F., 2010. Impact of the Gezhouba and Three Gorges Dams on habitat suitability of carps in the Yangtze River. J. Hydrol. 387 (3–4), 283–291. https://doi.org/10.1016/j.jhydrol.2010.04.018.

Zhang, J.F., Zhu, Y., Zhang, X.P., Ye, M., Yang, J.Z., 2018. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929. https://doi.org/10.1016/j.jhydrol.2018.04.065.

Zhang, L.J., Xue, M., Wang, M., Cai, W.J., Wang, L., Yu, Z.G., 2012. The spatiotemporal distribution of dissolved inorganic and organic carbon in the main stem of the Changjiang (Yangtze) River and the effect of the Three Gorges Reservoir. J. Geophys. Res. Biogeosci. 119, 741–757. https://doi.org/10.1002/2012JG002230.

Zheng, S.R., 2016. Reflections on the Three Gorges Project since its operation. Engineering. 2 (4), 389–397. https://doi.org/10.1016/J.ENG.2016.04.002.

Zhu, S.L., Hadzima-Nyarko, M., Gao, A., Wang, F.F., Wu, J.X., Wu, S.Q., 2019a. Two hybrid data-driven models for modeling water-air temperature relationship in rivers.

Environ. Sci. Pollut. Res. 26 (12), 12622–12630. https://doi.org/10.1007/s11356-019-04716-y.

Zhu, S.L., Nyarko, E.K., Hadzima-Nyarko, M., Heddam, S., Wu, S.Q., 2019b. Assessing the performance of a suite of machine learning models for daily river water temperature prediction. PeerJ 7, e7065. https://doi.org/10.7717/peerj.7065.