Modelling the effects of meteorological parameters on water temperature using artificial neural networks

Merve Temizyurek and Filiz Dadaser-Celik

ABSTRACT

Water temperature affects all biological and chemical processes in water; therefore, it is an extremely important water quality parameter. Meteorological factors are among the most important factors that affect water temperatures. The aim of this study is to develop an artificial neural network (ANN) model to investigate the effects of meteorological parameters on water temperatures at Kızılırmak River in Turkey. Water temperature data were collected from gauging stations on Kızılırmak River, and meteorological data were acquired from the nearest meteorological stations. Air temperature, wind speed, relative humidity, and previous water temperatures were formed the input parameters. The model output included water temperatures. All data were available for the 1995–2007 period, with occasional missing records. The activation functions of the ANN model and the number of neurons in the hidden layer were selected by trial-and-error method to find the best results. The root mean square error and the correlation coefficient between observed and simulated water temperatures were used to assess the model success. The best results were obtained by using sigmoid activation function and scaled conjugate gradient algorithm. This study showed that meteorological data can be used to simulate water temperature with ANN model for Kızılırmak River.

Key words | artificial neural networks, Kızılırmak River, modelling, simulation, water temperature

INTRODUCTION

Water temperature is one of the most crucial parameters for aquatic ecosystem applications. It influences not only the biological conditions and behaviour, but also chemical processes in aquatic systems. The most clear consequence of water temperatures on aquatic organisms is on their development and growth condition (Elliott & Hurley 1997). The aquatic species such as fish, insects, zooplankton, phytoplankton, etc., all have specific temperature ranges (School 2015). Water temperature also controls rates of all chemical reactions. One other important illustration of the effects of temperature on water chemistry is its influence on oxygen solvability. A cool water grips more oxygen than warm water. The increase in water temperature decreases the oxygen holding capacity of water, which can cause more stress on aquatic organisms which rely on oxygen (UCS 2017). Besides, some chemicals are more noxious to aquatic life at higher temperatures (School 2015).

The thermal regime, which is the natural process of heating and cooling of rivers, is greatly dependent on meteorological conditions (e.g. air temperature, solar radiation, and wind speed) and geophysical characteristics of rivers (topography, riparian vegetation, stream discharge, and stream-bed fluxes) (Bogan et al. 2006; Caissie 2006; Caissie et al. 2001). In recent years, there is an increasing concern about river water temperature changes. Webb (1996) examined the trends in river water temperatures based on examination of worldwide data. He showed that water temperatures increased about 1 °C in Europe. Kaushal et al. (2010) showed that river water temperatures increased significantly from 1990 to 2007 in the US. In Turkey, Albek & Albek (2009) examined the long-term trends that occurred in river water temperatures measured at several gauging stations. They showed that river water temperatures had a rising tendency in some stations over Turkey, but more streams had a reducing trend than a raising trend. If global warming continues in the future as projected, more rivers will have an upward temperature trend. Air temperatures are expected to raise by 2–3 °C in Turkey in the near future (Demircan et al. 2014; IPCC 2015). Therefore, it is necessary to examine how vulnerable stream water temperatures are to climatic variations.

doi: 10.2166/wst.2018.058
Models can provide useful information in understanding the relationships between river water temperatures and meteorological and geophysical parameters (Hébert et al. 2015). Usually, models used for prediction of river water temperatures can be classified two ways: deterministic models and data-driven models. Data-driven models focus on the relationships between input and output and generally use statistical techniques to relate water temperatures to the related parameters, such as air temperatures (Caissie 2006). Deterministic models use the physical principles (such as conservation of energy) to identify the relationships. Although deterministic models can provide information about the physical processes that control water temperature changes, extensive data requirements prevent their applications in data scarce regions. Data-driven models such as artificial neural networks (ANNs), support vector machines, boosted regression trees can provide an alternative solution (Caissie 2006).

Both approaches have been used for modelling river water temperatures. In a previous study, Marcé & Armengol (2008) compared three different approaches in water quality modelling. They used deterministic, empirical, and hybrid formulations. The results showed that the hybrid approach gave the best results, the empirical approach presented the second best results, and the deterministic approach gave the worst results. DeWeber & Wagner (2014) used ANNs for estimating daily mean water temperatures in the eastern USA. They estimated water temperatures according to meteorological parameters, land shape, and land cover characteristics. Efremova & Pal’shin (2003) analysed the vertical distribution of annual mean water temperatures in lakes as a function of local climate factors and lake morphometry during maximum water temperature periods. As a result, it is found that the maximum temperature in the 5 m layer depends on latitude and the main factors at 10 m or less than 10 m depth are sky morphometry and river water input. Luo et al. (2011) used different stochastic models for modelling water quality. They evaluated spatio-temporal trend analysis on 92 major rivers in Japan to monitor water quality. Gu & Li (2002) carried out a sensitivity analysis to assess river water temperature changes in reaction to hydrological and meteorological circumstances. The sensitivity analysis represented quantitative evidences concerning river temperatures being more sensitive to flux, upstream entry temperature, air temperature, moisture and solar radiation than wind speed, channel geometry, and morphometry. Ozaki et al. (2003) conducted a statistical analysis using meteorological and river water quality data to estimate influences of air temperatures on river water quality. They examined different factors for different time scales. Increase in air temperature results in increase in biochemical oxygen demand and solid matter and reduction in dissolved oxygen in almost all rivers. Erickson & Stefan (2000) investigated the effects of scale, river boundaries and inputs and local climate on air temperatures and river water temperatures. In this study, the relationship between air and river temperatures is found to be affected by dams, reservoirs, artificial heat inputs, underground water inputs and wind. Null et al. (2013) analysed and modelled the effect of climate change on river water temperature in Sierra Nevada Watershed. They determined that for each 20°C increase in air temperature, there is a 1.6°C increase for water temperature. Mohseni et al. (1998) examined the effects of climate change on the river ecosystem on weekly time scales. They observed that the relationship of stream/air temperature is affecting snow melting and underground water inputs. Bogan et al. (2005) investigated the relationships between river water temperatures and equilibrium temperatures. They observed that the relationship between them was almost linear and, in comparison with air temperature, equilibrium temperature is a better indicator for heat exchange. Benyahya et al. (2007) examined the statistical models used in river water temperatures modelling and compared them in terms of advantages and disadvantages. Gao et al. (2012) investigated the interrelationship between climate change and water temperatures using statistics, linear gradient exchange, cumulative anomaly, and variability coefficient methods. They used climatic and hydrological data for 50 years in three areas at Anhui, South China. Maier & Dandy (1996) used ANNs for estimating water salinity as a water quality parameter in the River Murray 14 days in advance. The results have 5.3 to 7.0% error. Ahmad et al. (2017) used feed forward ANNs for modelling water quality index of Perak River basin. Yeon et al. (2008) applied a water quality forecasting model using artificial intelligence. They used an ANN model and the adaptive neuro-fuzzy inference system. They compared different ANN models in rainy and non-rainy periods to monitor water quality. Chenard & Caissie (2008), Hadzima-Nyarko et al. (2014), Moustris et al. (2011), Singh et al. (2009) and Palani et al. (2008) used ANN applications for modelling water temperatures.

This study aims to develop an ANN model to investigate the effects of meteorological parameters on water temperatures at Kızılirmak River in Turkey. ANN modelling approach was selected as it is one of the most widely used and reliable data-driven methods. Kızılirmak River is one of the most important river systems in Turkey. It originates
in Eastern Anatolia around 39.8°N 38.3°E and flows first to the west and then to the north and finally discharge into the Black Sea. It has an average flow of 184 m³ s⁻¹. Total length is 1,355 km. The river basin covers an area of 53,800 km² and flows through regions, which have continental climates. In this study, we developed an ANN model for predicting water temperatures based on meteorological parameters.

MATERIALS AND METHODS

Data used

The data used in this study consist of water temperature data and meteorological data. Water temperature data were obtained from four gauging stations located on the Kızılırmak River. The data were on the monthly timescale and cover the 1995–2007 period. The locations are shown in Figure 1. The water temperature time series had some missing records. Missing records were due to extreme climatic conditions, when water sampling was not possible.

Meteorological data used in the analysis included air temperature, wind speed, and relative humidity data obtained from two meteorological stations in the Kızılırmak River Basin. The meteorological data were also available for the 1995–2007 period, with occasional missing records. Meteorological parameters selected have been used by many previous studies (e.g. Arganis et al. 2009) that focus on water temperatures, and data for them are collected regularly by state meteorology departments all over the world.

Artificial neural networks

ANN models were developed for modelling the effects of meteorological parameters on water temperatures at Kızılırmak River Basin. ANN modelling is a widely accepted approach by many disciplines for modelling complicated real-world systems. An ANN is described as a structure consisting of intense interrelationship with each other and adaptable simple process components (called artificial neurons and nodes). They are capable of large-scale parallel calculation for data processing and presentation (Jain et al. 1996). Artificial models are advantageous because (i) they comply with non-linear data much better, (ii) they provide correct estimation in indefinite data variability and measurement errors, (iii) it applies the tolerance of high parallelism, rapid processing, and hardware error, (iv) they allow learning and adaptation to update the reaction of system interior configuration to change of environment, and (v) they are capable of applying the model to non-learned generalising data. The main goal of the calculation based on ANN (neurocalculation) is to develop mathematical algorithms to enable ANN for learning, imitating data processing, and getting information like a human brain (Hinton 1989). An ANN consists of a set of process components which have interrelationships with each other. These process components form a network through three layers organized as input layer, hidden layer and output layer. Input layer includes the inputs to a model. The output layer is the layer where the model produces outputs. The connections between input and output layers are provided through hidden layers. An ANN can be described an

Figure 1 | The location of streamflow gauging stations in the Kızılırmak River Basin.
oriented graphic, which is in the model each neuron is shown as a transfer function, $f_i$.

$$y_i = f_i \left( \sum_{j=1}^{n} w_{ij} x_j - \Theta_i \right)$$  \quad (1)

In Equation (1), $y_i$ is the output of neuron $i$, $x_j$ is the $j$th input to neuron, $i$ and $j$ are the weights of nodes between neurons. $w_{ij}$ is the threshold (or bias) of neuron. $\Theta_i$ is the bias. Usually, $f_i$ is non-linear like S-shaped function or Gauss function.

ANN training can be split into two parts based on classification of its connections: feed forward and feedback. In the feed forward method, there is no connection from bigger number neuron to the smaller neuron in all nodes. All connections are from small number neurons to bigger. If there is no enumerated method like this, then the ANN is feedback. Feed forward networks consist of single-layer perceptron, multilayer perceptron and radial basis function nets. Competitive networks, Kohonen’s SOM, Hopfield network and ART models form recurrent/feedback networks (Jain et al. 1996).

**Model development**

In this study, the dependent variable was selected as monthly river water temperature and the independent variables were previous water temperature, monthly mean air temperature, monthly mean wind speed and monthly relative humidity (Figure 2). Monthly water temperatures measured at four stations and meteorological data obtained from the nearest meteorological stations to these gauging stations were used. The model output consisted of water temperatures.

As an activation function, sigmoid and hyperbolic tangent were used. Scaled conjugate gradient algorithm was used to train the model. To obtain the best results, in the hidden layer different number neurons (from one to ten neurons) were used.

Model training was performed by using 50% of the data, for model testing the 25% of data were used, and the remaining data were used as holdout. The activation functions of the ANN model and the number of neurons in the hidden layer were regulated by trial-and-error method. The accomplishment of the model was assessed by calculating the
root mean square error (RMSE) (Equation (2)) and the coefficient of determination ($R^2$) (Equation (3)) between observed and simulated water temperatures.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (T_{\text{observed},i} - T_{\text{simulated},i})^2}{n}}$$  \hspace{1cm} (2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (T_{\text{observed},i} - T_{\text{simulated},i})^2}{\sum_{i=1}^{n} (T_{\text{observed},i} - T)^2}$$  \hspace{1cm} (3)

In Equation (2) and (3), $n$ denotes the number of data used and $T$ denotes water temperatures. $T$ denotes the mean of the observed water temperatures.

**RESULTS AND DISCUSSION**

The statistical properties of water temperature and meteorological variables

We first analysed the general characteristics of data and the relationships between the parameters. Figure 3 shows the changes in monthly water temperature values from 1995 to 2007. Missing records can be identified in this figure. The missing records in time series of stations 1501, 1503 and 1517 were 14 months, 19 months and 2 months, respectively. Station 1535 had no missing data.

The statistical properties of the same data are shown in Table 1. The monthly mean temperatures during the 1995–2007 period were between 0 °C and 28 °C. For this period, the mean temperature in four stations were about 12–13 °C and the standard deviations were about 7–8 °C. The maximum water temperatures were observed in July and August. The minimum water temperatures were observed in January and February.

Meteorological variables used in the study were obtained from two stations (namely, Zara and Kayseri). The changes in mean monthly air temperature, mean monthly wind speed, and mean monthly relative humidity measured for the 1995–2007 time period were shown in Figure 4. Air temperature values in Zara Station were between −9 °C and 24 °C, where the mean air temperature was 9.1 °C. The mean wind speed at Zara Station was 2 m s$^{-1}$ and the mean relative humidity was 65.7%. In Kayseri Station, air temperatures changed from −6 °C to 26 °C and the mean temperature was 11.6 °C. The mean wind speed in Kayseri Station was 2.8 m s$^{-1}$ and the mean relative humidity was 60%.

Figure 3 | Water temperatures at Kızılırmak River for the 1995–2007 period.
The effects of meteorological parameters on water temperatures

The effects of meteorological parameters on Kızılrınkırmak River water temperature values were determined by correlation analysis. In the correlation analysis, the relationships between water temperature values and the water temperature value belongs to previous month, air temperature, relative humidity and mean wind speed were examined (Table 2). Because of continuous temperature exchange between rivers and air, the most powerful interaction is observed with air temperature. Correlation coefficients calculated between water temperatures and air temperatures were between 0.87 and 0.92. The second most powerful relationship in examined parameters was observed between water temperatures and the temperatures belongs to previous month (the correlation coefficient changed between 0.75 and 0.78). The reason of this situation is the autocorrelation between water temperature values. Water temperature value has a powerful interaction with the temperature value of previous month. The relationships between water temperature and relative humidity and wind speeds were comparatively weak.

Modelling with ANNs

As activation functions, S-shape function and hyperbolic tangent function were used. The model was trained with scaled conjugate gradient algorithm. To obtain best results in modelling, one to ten neurons were used in the hidden layer. To determine the best configuration, trial-and-error method was used based on calculated RMSE and R² values.

For each station, different numbers of neurons gave the best results in the hidden layer. S-shaped activation function with scaled conjugated gradient algorithm provided the best results for all stations. Figure 5 provides a comparison of the water temperature values measured and simulated for the 1995–2007 period for four stations. In Figure 6, measured water temperatures were drawn against simulated water temperatures. The best configuration

Table 1 | The statistical properties of water temperatures measured at Kızılrınkırmak River for the 1995–2007 period

<table>
<thead>
<tr>
<th>Station</th>
<th>Minimum (°C)</th>
<th>Maximum (°C)</th>
<th>Mean (°C)</th>
<th>Standard deviation (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1501</td>
<td>-1</td>
<td>28</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>1503</td>
<td>0</td>
<td>27</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>1517</td>
<td>0</td>
<td>24</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>1535</td>
<td>0</td>
<td>26</td>
<td>13</td>
<td>8</td>
</tr>
</tbody>
</table>

Figure 4 | Air temperature, wind speed and relative humidity values measured at Zara and Kayseri meteorological stations for the 1995–2007 period.
provided RMSE values of 2.49, 2.10, 2.37 and 2.64 °C for stations 1501, 1503, 1517, and 1535, respectively, for the 1995–2007 period. The R² values for the same stations and same time period were 0.90, 0.93, 0.93, and 0.94, respectively. Train, test and holdout error values were very close to each other. These results showed that ANN approach was successful in modelling the relationships between water temperatures and meteorological parameters.

The input parameters in this study include previous water temperature, mean air temperature, mean wind speed, and mean relative humidity. A sensitivity analysis was carried on by calculating the relative consequence of each variable in the ANN model. In this application, the ANN models developed by excluding a single input variable each time and the remains were compared with each other. S-shaped activation function and scaled conjugated gradient algorithm gave the best results in previous analysis were used in this application. The best results were obtained when four input parameters (air temperature, A, previous water temperature, P, wind speed, W, relative humidity, H), were used (Table 3). The largest error value was observed when air temperature is excluded (case 4). This situation showed that air temperature is the most effective parameter of used parameters for modelling on water temperature, as expected. The error values calculated for the models that include air temperatures but exclude other parameters, were not very different than the initial conditions. Among all parameters, wind speed seems the least important parameter.

Table 2  | The correlations between water temperatures and previous water temperatures, air temperatures, relative humidity and wind speed

<table>
<thead>
<tr>
<th>Station/parameter</th>
<th>Station name</th>
<th>Previous water temperature (°C)</th>
<th>Air temperature (°C)</th>
<th>Wind speed (m s⁻¹)</th>
<th>Relative humidity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1501</td>
<td>Yamula</td>
<td>0.75</td>
<td>0.87</td>
<td>0.43</td>
<td>−0.58</td>
</tr>
<tr>
<td>1505</td>
<td>Yahşiyan</td>
<td>0.76</td>
<td>0.91</td>
<td>−0.17</td>
<td>−0.72</td>
</tr>
<tr>
<td>1517</td>
<td>Şefaatli</td>
<td>0.75</td>
<td>0.92</td>
<td>−0.11</td>
<td>−0.75</td>
</tr>
<tr>
<td>1535</td>
<td>Söğütülihan</td>
<td>0.78</td>
<td>0.92</td>
<td>0.50</td>
<td>−0.65</td>
</tr>
</tbody>
</table>

Figure 5  | The comparison of observed and simulated water temperatures for the 1995–2007 period.
The correlations between the observed and simulated water temperatures for the 1995–2007 period.

Table 3 | The results of the sensitivity analysis. It was noted that developed with different input structures RMSE and coefficient of determination ($R^2$) of the ANN models

<table>
<thead>
<tr>
<th>Case number</th>
<th>Station</th>
<th>Model inputs</th>
<th>Number of neurons</th>
<th>RMSE (°C)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>1</td>
<td>1501</td>
<td>A, W, H, P</td>
<td>8</td>
<td>2.61</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>1503</td>
<td>A, W, H, P</td>
<td>1</td>
<td>2.09</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td>1517</td>
<td>A, W, H, P</td>
<td>1</td>
<td>2.44</td>
<td>2.13</td>
</tr>
<tr>
<td></td>
<td>1535</td>
<td>A, W, H, P</td>
<td>3</td>
<td>2.63</td>
<td>2.51</td>
</tr>
<tr>
<td>2</td>
<td>1501</td>
<td>A, W, H</td>
<td>9</td>
<td>2.90</td>
<td>2.40</td>
</tr>
<tr>
<td></td>
<td>1503</td>
<td>A, W, H</td>
<td>10</td>
<td>2.58</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>1517</td>
<td>A, W, H</td>
<td>3</td>
<td>2.34</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>1535</td>
<td>A, W, H</td>
<td>9</td>
<td>2.59</td>
<td>2.42</td>
</tr>
<tr>
<td>3</td>
<td>1501</td>
<td>A, W, P</td>
<td>8</td>
<td>2.59</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>1503</td>
<td>A, W, P</td>
<td>3</td>
<td>1.92</td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>1517</td>
<td>A, W, P</td>
<td>9</td>
<td>2.52</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>1535</td>
<td>A, W, P</td>
<td>5</td>
<td>2.66</td>
<td>2.37</td>
</tr>
<tr>
<td>4</td>
<td>1501</td>
<td>W, H, P</td>
<td>8</td>
<td>3.06</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>1503</td>
<td>W, H, P</td>
<td>2</td>
<td>2.16</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>1517</td>
<td>W, H, P</td>
<td>7</td>
<td>2.84</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td>1535</td>
<td>W, H, P</td>
<td>3</td>
<td>3.13</td>
<td>3.85</td>
</tr>
<tr>
<td>5</td>
<td>1501</td>
<td>A, H, P</td>
<td>8</td>
<td>2.62</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>1503</td>
<td>A, H, P</td>
<td>7</td>
<td>1.84</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>1517</td>
<td>A, H, P</td>
<td>10</td>
<td>2.17</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td>1535</td>
<td>A, H, P</td>
<td>6</td>
<td>2.61</td>
<td>2.48</td>
</tr>
</tbody>
</table>
SUMMARY AND CONCLUSIONS

The aim of this study was to examine the effects of meteorological parameters (air temperature, relative humidity and wind speed) on water temperatures at the Kizilirmak River Basin using ANN modelling approach.

The best results for water temperatures were obtained when four parameters were used as inputs and when S-shape activation function and scaled conjugate gradient algorithm were used. In this configuration, the RMSE values between observed and modelled values for four stations were between 2.10 and 2.64 °C. The correlation coefficients between the observed and modelled values for the same stations were over 0.90. Both correlation analysis and the sensitivity analysis showed that air temperature is the most important parameters affecting water temperatures. It can be inferred from these results that the ANN approach can be used for modelling the effects of meteorological factors on water temperatures.

In this study, water temperatures were estimated by using the interaction between river water temperature and various meteorological parameters. While evaluating the results we should consider that water temperatures are not only influenced by meteorological factors, but also geophysical factors. Therefore, inclusion of geophysical factors could improve the results. However, geophysical factors are not always readily available, which could prevent their widespread use in models. The method described in this paper can be of interest to hydrologists, environmental scientists and ecologists, as it provides a quick estimate of the relationships between water temperatures and meteorological parameters. Meteorological data has been collected regularly, which increases the potential for application of the method for estimation of river water temperatures. Once developed, the models can be used to simulate the effects of climatic changes on river ecosystems.

This study showed that ANN models can be successfully applied by using only meteorological parameters as inputs. In the future, we plan to investigate the applicability of other data-driven models (e.g. support vector machines) for this purpose and to investigate if model prediction accuracy could be improved.

ACKNOWLEDGEMENTS

This study was funded by the Erciyes University Research Fund (Grant No. FYL-2015-5645).

REFERENCES


UCS, Union of Concerned Scientists 2017 Global Warming Impacts. Cambridge, MA, USA.


First received 24 September 2017; accepted in revised form 26 January 2018. Available online 12 February 2018