River Flow 2014 – Schleiss et al. (Eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-02674-2

Stream temperature forecasting by means of ensemble of neural networks: Importance of input variables and ensemble size

M.J. Napiorkowski
Environmental Engineering Faculty, Warsaw University of Technology, Warsaw, Poland

A.P. Piotrowski & J.J. Napiorkowski
Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland

ABSTRACT: The models aiming at forecasting or projection of water temperature in natural streams located in cold climate zones, where the seasonality plays important role, are of great importance, as stream temperature is still frequently not measured on site and some tools are needed to evaluate water temperature values for future climatic conditions based on simple hydro-meteorological variables. In several papers Artificial Neural Networks (ANN) were proposed to stream temperature forecasting. However, it is still not clear which hydrological and meteorological variables, among the ones that could be available from Global Circulation Models, are the most significant as ANN model inputs. It is well known that using model ensembles may significantly improve the forecasting accuracy, also in the case of ANN models. However, the impact of ANN ensemble size and of the ensemble aggregation approach on the forecasting accuracy has been rarely studied so far. The present paper aims at both, the problem of the choice of proper ANN input variables, ensemble size and ensemble aggregation approach, at an example of Biala Tarnowska river catchment, located in mountainous part of southern Poland. The meteorological data include declination of the sun, mean, minimum and maximum air daily temperature, which are available from two stations, in addition to the river runoff measured in a single gauging station. The river freezing and melting processes that occur during winter months in the catchment pose a major problem for stream temperature forecasting.

1 INTRODUCTION

Forecasting of water temperature in natural streams located in cold climate zones, where the seasonality plays important role, is of great importance, at least because some tools are needed to evaluate water temperature values for future climatic conditions based on simple hydro-meteorological variables.

Water temperature of rivers is needed in ecological studies, as changes in temperature can significantly impact fish distribution, growth, mortality, production, habitat use and community dynamics (Caissie 2006, St-Hilaire et al. 2012). Models of river water temperature are essential in assessment of the possible impact of climate change on the future biological and chemical processes in rivers (Jeong et al. 2012), as the relation between climate change and the change of water temperature in natural rivers may be relatively complicated (van Vliet et al. 2011). Among various modeling tools, Artificial Neural Networks (ANN) show its ability to properly predict water temperature in rivers (Chenard & Caissie, 2008; Sahoo et al. 2009; Piotrowski et al. 2014).

Water temperature in natural rivers may vary due to natural processes, anthropogenic impacts (Poole & Berman 2001) or due to human-made thermal pollutions (Vega et al. 1998). The present paper puts attention to the first issue. The meteorological parameters affecting stream temperature include air temperature, net solar radiation, cloud cover, relative humidity and wind speed (Bogan et al. 2006). Other parameters affecting stream temperature include: water depth, stream flow rate, groundwater inflow rate and temperature, thermal conductivity of the sediments, wind sheltering and shading, and cooling water inputs (Erickson & Stefan 2000). If sufficient data of weather and stream conditions are available, good estimates of stream temperature can be obtained from deterministic models.

In the last decades ensemble forecasting (learning) is recognized as a valuable strategy within the computational intelligence community. Ensemble forecasting has proven to be effective in solving many real world problems, in particularly in hydrology. The ensemble flood forecasting is a well-known paradigm for years (see the review in Cloke & Pappenburger 2009). Ensemble learning is a distinct concept where decisions of multiple models are combined to improve the prediction performance. It aims at improving the generalization ability and the reliability of the system. Studies have shown that an ensemble system is generally more accurate than any individual model, and its effectiveness has been recognized in different benchmark data sets (Dietterich 2000). The success of the ensemble forecasting results from the degree of diversity within the ensemble. A good ensemble is one in which the models make different errors on the same data point (Soares et al. 2013). In this setting, ANN ensembles have been widely investigated for regression problems.

A number of aggregation algorithms of various complexity are proposed in the literature (Optiz & Shavlik 1996, See & Abrahart, 2001, Zhou et al. 2002, Granitto et al. 2005). Probably the simplest is the idea of computing a median or a mean of forecasts. Within the ensemble each ANN member may have different architecture, may be trained by different optimization algorithms, or may be trained based on different samples of data selected according to some method. Such different samples are usually obtained by means of either bagging or boosting methods (see e.g. Breiman 1996, Granitto et al. 2005, Zheng 2009 and Erdal & Karakurt 2013).

The objective of this paper is verification of the importance of an ensemble modeling for water temperature forecasting in moderately cold climate zones. The impact of number of ensemble members and the ensemble aggregation method on the performance of the model is tested. This study is based on the hydro-meteorological data collected from Biala Tarnowska (Poland) and is a direct continuation of research performed by Piotrowski et al. (2014).

In Piotrowski et al. (2014) the detailed comparison of the performance of nature inspired optimization methods and Levenberg–Marquardt (LM) algorithm in Multi-Layer Perceptron ANNs (MLP) training was presented. Large number of metaheuristics, including Differential Evolution, Particle Swarm Optimization, Evolution Strategies, multialgorithms and Direct Search methods were compared with LM algorithm on MLP training for the described case study. The impact of population size and some control parameters of particular metaheuristics on the ANN training performance was verified. It was found that despite widely claimed large improvement in nature inspired methods during last years, the vast majority of them are still outperformed by LM algorithm on the selected problem. Due to this finding and the speed of gradient-based methods, in the present study all ANNs are trained by means of LM algorithm.

2 BIALA TARNOWSKA RIVER CATCHMENTS

The valley of Biala Tarnowska River is located in the central part of the Polish Carpathians. The source is located in the Low Beskid at altitudes of 730 meters (Carpathian belt, southern Poland). The total length of the river is 101.8 km and the catchment area to the Koszyce Wielkie gauging station (10 km to the south-west from the city of Tarnow) equals 956.9 km^2. Biala Tarnowska catchment is very narrow and extends from the border with Slovakia. The majority of the river has unregulated banks and is in a natural state. Fields, pastures, meadows, and natural vegetation predominate in the catchment of the upper and middle portion of the river. Dominant geology can be defined as sandstone and shale flysch. Biala Tanowska catchment is divided into two different parts. The south section, representing about 25% of the catchment, is a wooded mountain part with the average slope of 10‰. The north part representing almost 75% of the basin, characterized by deep river valleys (mostly agricultural hills and foothills), is generally deforested. The river slope in the northern part is in the range of 0.9–5 ‰.

The highest precipitation (up to 100 mm/month), and hence frequent spates are observed during summer months. Average (high/low) temperature is (1°C/–5°C) in January and (25°C/13°C) in July.

According to the Köppen Climate Classification Biala Tarnowska river is located within the Humid Continental Zone. It may freeze during winter months and river ice may occur between November and April. If the Biala Tarnowska River is frozen, it is assumed in this paper that its water temperature equals 0°C.

One lead-day forecasting of Biala Tarnowska River temperature in Koszyce Wielkie village is performed according to hydro-meteorological measurements collected between November 1983 and October 2000. The data collected before November 1990 compose training set, the data collected between November 1990 and October 1995 are included into validation set, the rest of data form the testing set.

In the present study the MLP input variables are selected based on expert knowledge about the major factors that impact the river temperatures in moderately cold climate zones (Chenard & Caissie 2008). Air temperature is used as a predictor variable in the water temperature model, because it is a major component in calculating net changes of heat flux at the water surface. From the same reason net solar radiation should be taken into account. Water temperature is inversely related to river discharge, reflecting a reduced thermal capacity under decreasing flow volumes, hence river discharge is an additional input variable. Hence, taking into account data availability in Biala Tarnowska catchment the following measurements are considered as input variables:

- declination of the sun (Sun),
- daily average (Tavr) air temperature in Tarnow,
- daily maximum (Tmax) air temperature in Tarnow,
- daily average (Navr) air temperature in Nowy Sacz,
- daily maximum (Nmax) air temperature in the Nowy Sacz,
- daily runoff (Q) in Koszyce Wielkie gauging station,
- daily average water temperature (WT) in Koszyce Wielkie, which is to be predicted by the ANN.

3 MULTI-LAYER PERCEPTRON NEURAL NETWORKS

Multi-Layer Perceptron ANN consists of nodes grouped into input, hidden and output layers. Single hidden layer is considered enough to approximate continuous differentiable functions, but there is no widely accepted rule regarding the number of hidden nodes (Zhang et al. 1998). MLP neural network is defined as:

$$y^P = v_0 + \sum_{j=1}^{J} v_j f\left(w_{j0} + \sum_{k=1}^{K} w_{jk} z_k \right) \tag{1}$$

where y^p is a predicted value of the output variable, z^k, $k = 1, \dots$ K represents input variables, w and v are MLP weights (parameters to be optimized), J is the number of hidden units and f is the so-called activation function. In the present paper the popular logistic activation function is used:

$$f(a) = \frac{1}{1 + \mathrm{e}^{-a}} \tag{2}$$

The Mean Square Error (MSE) objective function is required by LM algorithm and is used in the present study

$$MSE(\mathbf{w},\mathbf{v}) = \min_{\mathbf{w},\mathbf{v}} \frac{1}{N} \sum_{n=1}^{N} \left(y_n^P(\mathbf{w},\mathbf{v}) - y_n \right)^2 \tag{3}$$

where y is the measured value of the output variable and N is the number of observations.

3.1 *Method to prevent overfitting*

The main problem with the practical application of ANNs is the possible overfitting to the training data (Holmstrom & Koistinnen 1992, Prechlet, 1998). ANN overfitting is understood as fitting the ANN weights not only to the signal but also to a noise that is always present in the training sample. The possibility of overfitting depends on the ANN size, number of training patterns, the covering of feature space by the data and their quality.

To prevent overfitting in this paper only relatively simple MLP architectures are considered. However, the choice of simple MLP architecture, although important, is not sufficient. Hence in the present paper the simple early stopping technique is used according to Prechlet's (1998) Generalization Loss (GLα) class, which was developed for gradient-based training methods and turned out successful in water-related applications (Piotrowski and Napiorkowski, 2013). The data set is divided into three parts: training (TR), validation (V) and independent testing (TE), hence three different MSE values are computed in each iteration (MSE_{TR}, MSE_V and MSE_{TE}). In the case of Biala Tarnowska river this three data sets are of almost equal size, each one is composed of 10-years long measurements.

In case of gradient-based learning algorithms the derivatives and the step size are determined according to MSE_{TR} only. In this paper the LM training method is stopped at generation t at which (see Prechlet, 1998)

$$GL(t) = \left(\frac{MSE_V(t)}{MSE_V(c)} - 1 \right) > \alpha \tag{4}$$

where α is set to 1.2 and $c < t$ is the number of the generation at which the lowest value of MSE_V was obtained. The training may also be stopped after pre-defined number of function calls which is set to 1000, longer training does not lead to further improvement. After termination, the best solution returned by the algorithm is chosen according to the performance for the validation data, i.e. the solution with the lowest $MSE_V(c)$.

3.2 *Initialization and bounds*

ANN weights are frequently initialized within a small limited range around 0 (Thimm and Fiesler, 1997; Zhang et al. 1998). After previous experience of the authors (Piotrowski et al. 2014), in this study the ANN weights are generated randomly from uniform distribution within [−1,1].

In the case of MLP input and output variables are frequently linearly normalized to [0,1] interval (Zhang et al. 1998), what is also done in the present paper.

Table 1. MLP structures with description of input variables.

Structure	No. param.	Input variables
13-9-1	136	S(t-1), S(t-2), Tavr(t-1), Tarv(t-2),
13-7-1	106	STavr(t-3_7), Tmax(t-1), Tmax(t-2),
13-5-1	76	Navr(t-1), Navr(t-2), Nmax(t-1),
13-4-1	61	Nmax(t-2), Q(t-1), Q(t-2)
9-7-1	78	
9-6-1	67	S(t-1), S(t-2), Tavr(t-1), Tavr(t-2)
9-5-1	56	STavr(t-3_7), Tmax(t-1), Tmax(t-2),
9-4-1	45	Q(t-1),Q(t-2)
9-3-1	34	
8-7-1	71	S(t-1), S(t-2),
8-6-1	61	Tavr(t-1), Tavr(t-2), STavr(t-3_7),
8-5-1	51	Tmax(t-1), Tmax(t-2),
8-4-1	41	Q(t-1)
8-3-1	31	
7-7-1-T_{MAX2}	64	
7-6-1-T_{MAX2}	55	S(t-1), S(t-2),
7-5-1-T_{MAX2}	46	Tavr(t-1), Tavr(t-2),
7-4-1-T_{MAX2}	37	STavr(t-3_7), Tmax(t-1), Q(t-1)
7-3-1-T_{MAX2}	28	
7-5-1-T_{AVER2}	46	S(t-1), S(t-2), Tavr(t-1), Tavr(t-2), Tmax(t-1), Tmax(t-2), Q(t-1)
7-5-1-SUN2	46	S(t-1), Tavr(t-1), Tavr(t-2), STavr(t-3_7), Tmax(t-1), Tmax(t-2), Q(t-1)

3.3 *Data sets for water temperature forecasting at Biala Tarnowska river*

The considered input variables and MLP architectures are shown in Table 1. Notations are given in the section 2 of the text, except STavr(t-3_7), which is a sum of the daily average air temperatures measured in Tarnow 3 to 7 days before t.

Note that there are no autoregressive inputs and the past water temperatures are never used to model the future ones in this study. There are two reasons of that. Firstly, the water temperature forecasting models are sometimes expected to provide forecasts for the future climate conditions, and secondly, the water temperature is still not always measured in natural rivers.

3.4 *Median results achieved in single model case*

Table 2 depicts the median performance from 50 runs that are convenient for the comparison with forecasts obtained by means of aggregation algorithms of ANN ensembles. The lowest values for a particular data set are bolded. This results are the same as reported in Piotrowski et al. (2014).

The 7-5-1-T_{MAX2} architecture performs best according to training-independent testing data. 7-5-1-T_{MAX2} is composed of 7 inputs, 5 hidden nodes and a single output. T_{MAX2} means that

Table 2. The 50-run averaged performance of different MLP structures trained with LM algorithm.

	Training	Validation	Test	
Structure	Median	Median	Median	Best
13-9-1	0.974	1.179	1.064	0.878
13-7-1	1.010	1.209	1.111	0.919
13-5-1	1.015	**1.163**	0.941	0.871
13-4-1	1.061	1.182	0.949	0.831
9-7-1	0.965	1.169	0.980	0.867
9-6-1	0.989	1.168	0.977	0.864
9-5-1	1.011	1.175	0.953	0.839
9-4-1	1.037	1.181	0.920	0.836
9-3-1	1.131	1.238	0.979	0.840
8-7-1	**0.964**	1.166	0.923	0.824
8-6-1	0.978	1.176	0.918	0.840
8-5-1	1.014	1.173	0.911	0.824
8-4-1	1.058	1.176	0.896	0.838
8-3-1	1.097	1.210	0.950	0.840
7-7-1-T_{MAX2}	0.969	1.169	0.920	0.831
7-6-1-T_{MAX2}	1.004	1.187	0.912	0.822
7-5-1-T_{MAX2}	1.027	1.175	**0.884**	0.820
7-4-1-T_{MAX2}	1.071	1.178	0.885	**0.802**
7-3-1-T_{MAX2}	1.137	1.222	0.964	0.830
7-5-1-T_{AVER2}	1.325	1.194	1.067	0.968
7-5-1-SUN2	1.300	1.557	1.252	1.200

this architecture was obtained from 8-5-1 one by discarding the daily maximum air temperature measured 2 days before the day for which the forecast is performed, the architectures 7-5-1-T_{AVER2} and 7-5-1-SUN2 were obtained in similar fashion by discarding the average air temperature or declination of the sun measured 2 days ago, respectively.

4 IMPROVEMENT BY MEANS OF AGGREGATION OF ANN ENSEMBLE

The temperature forecasting performance may be improved by means of aggregation algorithms of ANN ensembles. A large number of very different methods to construct ensemble members and perform aggregation of the results have already been proposed in the literature. Some approaches that aim at using different training subsets for each ensemble member, like bagging, have strong theoretical background (see Breiman, 1996). Unfortunately in case of many applications, the main assumptions like uncorrelated errors of ensemble members are not fulfilled (Jeong and Kim, 2005).

The performance of various approaches is frequently very similar (see the comparison presented by Granitto et al. 2005). In the present paper, following the results obtained by Zheng (2009), only the simplest methods are applied, in which data from the training set are used by each ensemble member and aggregated forecasts $y^{p}_{agg,n}$ are estimated as a mean or median of the forecasts $y^{p}_{i,n}$ performed by NEM particular members (NEM is a Number of Ensemble Members) randomly selected form the set of 50 MLP. Four different ensembles are tested, namely the ensembles with NEM = 5, 10, 20 and 50 (all trained) members for each considered MLP structure. In such a simply way the impact of the ensemble size on the performance of temperature forecast can be examined. The structures of MLPs are defined in Table 1 and the 50-run median performances obtained for each structure by single models are reported in Table 2.

In this paper the aggregated mean square error (MedianA and MeanA), based on aggregation by means of median and mean, respectively, are defined as follows:

$$MedianA = \frac{1}{N}\sum_{n=1}^{N}\left(y_n^{P,agg} - y_n\right)^2$$
$$y_n^{P,agg} = median(y_n^{P,i}, \quad i = 1, ..., NEM) \tag{5}$$

and

$$MeanA = \frac{1}{N}\sum_{n=1}^{N}\left(y_n^{P,agg} - y_n\right)^2 \tag{6}$$

$$y_n^{P,agg} = \sum_{i=1}^{NEM} y_n^{P,i}/NEM \tag{7}$$

In above equations NEM ensemble members are chosen (NEM set to 50, 20, 10 and 5 are tested in the study) and N is the number of data in particular set (training, validation or testing). The results obtained for Biala Tarnowska river by means of aggregation algorithms of ANN ensembles are reported in Table 3 for training data, in Table 4 for validation data and in Table 5 for testing data.

The lowest mean and median values for each ensemble (i.e. each column) are bolded.

The best results for training data set are obtained by means of 13-9-1 MLP, i.e. when input set is formed by all considered variables and the number of neurons in the hidden layer is the highest. This most complex architecture is defined by as many as 136 weights, so it is not surprising that the best performance for training data does not result in good performance for the validation or testing data sets.

Two architecture, namely 8-5-1 and 7-5-1-T_{MAX2} perform best according to training-independent testing data for ANN ensembles aggregated by means of Mean and Median. This confirms that most important inputs for temperature forecast at Koszyce Wielkie are declination of the sun, the average and maximum temperature measured in Tarnow (close to Koszyce Wielkie) in last 2 days and flow measured 1 day ago.

One may note that for the testing data set the MeanA and MedianA are lower than the median MSE obtained by individual MLP models. Moreover, for 7-5-1 ANN, the increase in the number of ensemble members gives better results, both for MeanA and MedianA. However, the best among 50 MLP models usually perform better than the aggregated MLP prediction.

The problem is just that basing on training and validation data sets no one is able to determine, which single MLP would perform best for the testing data. On the other hand, aggregation of the forecasts achieved by MLP models of various performance may improve the results. A good ensemble is one in which the models make different errors on the same data point.

In Figures 1–4 the results of one lead-day water temperature forecasting by means of ANN ensembles are illustrated for four seasons. To get better visual comparison of observed and predicted temperatures in a single Figure, 2 characteristic moths for each season are depicted. The best architecture for testing data set is selected, namely 7-5-1-T_{MAX2}. Similarly to the results included in Tables 3–5 the forecastings of four different ensembles with 5, 10, 20 and 50 elements are plotted. Additionally, the

Table 3. The results for ANN ensembles aggregated by means of Mean and Median—training data.

Ensemble	50		20		10		5	
ANN	Mean A	Median A	Mean A	Median A	Mean A	Median A	Mean A	Median A
13-9-1	0.969	**0.928**	0.938	**0.920**	**0.898**	**0.907**	0.979	**0.888**
13-7-1	1.023	0.967	1.066	1.058	1.039	1.016	1.043	0.972
13-5-1	0.977	0.977	0.975	0.984	0.982	0.988	0.984	1.026
13-4-1	1.007	1.020	0.998	1.011	0.999	1.008	1.034	1.052
9-7-1	0.911	0.919	0.926	0.933	0.948	0.958	0.951	0.973
9-6-1	0.938	0.943	0.951	0.961	0.934	0.945	0.938	0.943
9-5-1	0.966	0.969	0.971	0.976	0.967	0.969	0.984	0.991
9-4-1	0.998	1.001	1.002	1.004	1.019	1.022	1.010	1.012
9-3-1	1.078	1.108	1.078	1.103	1.094	1.127	1.116	1.148
8-7-1	0.933	0.939	0.949	0.961	0.951	0.958	0.936	0.940
8-6-1	0.938	0.943	**0.925**	0.925	0.940	0.938	0.941	0.942
8-5-1	0.965	0.973	0.965	0.970	0.975	0.976	0.962	0.962
8-4-1	1.023	1.030	1.015	1.024	1.008	1.019	1.011	1.024
8-3-1	1.065	1.065	1.062	1.068	1.044	1.066	1.030	1.058
7-7-1-T_{MAX2}	**0.931**	0.934	0.932	0.937	0.922	0.925	**0.912**	0.914
7-6-1-T_{MAX2}	0.954	0.958	0.956	0.959	0.956	0.958	0.958	0.965
7-5-1-T_{MAX2}	0.981	0.989	0.981	0.987	0.993	1.000	0.975	0.984
7-4-1-T_{MAX2}	1.021	1.031	1.027	1.042	1.030	1.044	1.044	1.063
7-3-1-T_{MAX2}	1.083	1.100	1.080	1.087	1.088	1.097	1.082	1.076
7-5-1-TAVER	1.262	1.276	1.259	1.278	1.291	1.311	1.296	1.332
7-5-1-SUN	1.265	1.276	1.274	1.295	1.262	1.285	1.251	1.259

Table 4. The results for ANN ensembles aggregated by means of Mean and Median—validation data.

Ensemble	50		20		10		5	
ANN	Mean A	Median A	Mean A	Median A	Mean A	Median A	Mean A	Median A
13-9-1	1.202	1.133	1.151	1.120	1.108	1.116	1.097	1.101
13-7-1	1.255	1.167	1.286	1.242	1.249	1.189	1.265	1.170
13-5-1	1.121	1.112	1.116	1.114	1.129	1.119	1.109	1.136
13-4-1	1.128	1.134	1.119	1.127	1.123	1.133	1.134	1.143
9-7-1	1.107	1.112	1.110	1.114	1.128	1.139	1.108	1.125
9-6-1	**1.107**	**1.108**	1.109	1.115	**1.099**	**1.103**	1.107	1.108
9-5-1	1.114	1.119	1.127	1.136	1.122	1.131	1.134	1.137
9-4-1	1.127	1.126	1.130	1.131	1.143	1.144	1.128	1.122
9-3-1	1.172	1.203	1.173	1.203	1.189	1.221	1.217	1.236
8-7-1	1.113	1.112	1.119	1.120	1.110	1.112	1.095	1.084
8-6-1	1.110	1.110	**1.108**	**1.107**	1.105	1.107	1.096	1.109
8-5-1	1.111	1.116	1.106	1.114	1.122	1.125	1.114	1.114
8-4-1	1.130	1.133	1.128	1.129	1.131	1.130	1.132	1.129
8-3-1	1.154	1.152	1.156	1.154	1.144	1.164	1.135	1.161
7-7-1-T_{MAX2}	1.115	1.112	1.109	1.111	1.106	1.109	**1.103**	**1.101**
7-6-1-T_{MAX2}	1.119	1.120	1.109	1.115	1.107	1.108	1.109	1.118
7-5-1-T_{MAX2}	1.115	1.118	1.112	1.118	1.112	1.111	1.118	1.120
7-4-1-T_{MAX2}	1.128	1.133	1.132	1.138	1.133	1.140	1.152	1.159
7-3-1-T_{MAX2}	1.164	1.181	1.164	1.165	1.174	1.185	1.175	1.166
7-5-1-TAVER	1.143	1.150	1.127	1.130	1.133	1.142	1.136	1.149
7-5-1-SUN	1.516	1.526	1.515	1.532	1.508	1.527	1.514	1.526

Table 5. The results for ANN ensembles aggregated by means of Mean and Median—testing data.

Ensemble	50		20		10		5	
ANN	Mean A	Median A	Mean A	Median A	Mean A	Median A	Mean A	Median A
13-9-1	0.969	0.928	0.938	0.920	0.898	0.907	0.979	0.888
13-9-1	0.857	0.845	0.844	0.846	0.834	0.852	0.848	0.877
13-7-1	0.892	0.857	0.925	0.916	0.883	0.863	0.891	0.857
13-5-1	0.816	0.819	0.814	0.820	0.821	0.829	0.849	0.858
13-4-1	0.830	0.836	0.828	0.835	0.832	0.838	0.859	0.864
9-7-1	0.833	0.839	0.835	0.832	0.840	0.857	0.835	0.859
9-6-1	0.817	0.824	0.819	0.831	0.827	0.838	**0.817**	0.824
9-5-1	0.825	0.830	0.830	0.831	0.848	0.844	0.867	0.868
9-4-1	0.834	0.839	0.833	0.838	0.856	0.860	0.851	0.857
9-3-1	0.876	0.911	0.872	0.895	0.891	0.922	0.908	0.930
8-7-1	0.821	0.817	0.822	0.822	0.821	0.820	0.830	0.828
8-6-1	0.811	0.811	0.821	0.821	0.825	0.827	0.828	0.845
8-5-1	0.808	0.812	**0.806**	**0.804**	0.821	**0.815**	0.824	0.825
8-4-1	0.822	0.821	0.827	0.826	0.834	0.841	0.886	0.853
8-3-1	0.854	0.857	0.859	0.852	0.854	0.876	0.849	0.874
7-7-1-T_{MAX2}	0.825	0.820	0.817	0.809	0.839	0.835	0.854	0.845
7-6-1-T_{MAX2}	0.815	0.816	0.810	0.813	0.819	0.822	0.838	0.850
7-5-1-T_{MAX2}	**0.805**	**0.808**	0.809	0.812	**0.809**	0.817	0.820	**0.820**
7-4-1-T_{MAX2}	0.939	0.952	0.937	0.938	0.948	0.954	0.962	0.973
7-3-1-T_{MAX2}	1.192	1.200	1.199	1.210	1.203	1.211	1.215	1.229
7-5-1-TAVER	0.821	0.822	0.824	0.826	0.836	0.831	0.847	0.845
7-5-1-SUN	0.862	0.882	0.851	0.862	0.861	0.866	0.871	0.876

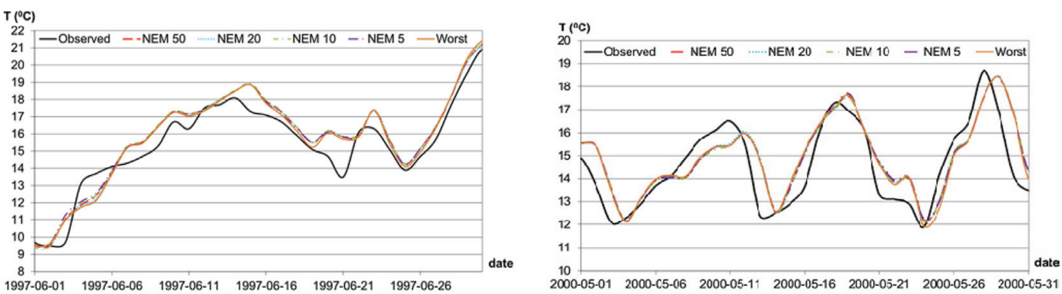

Figure 1. One lead-day water temperature forecasting by means of ANN ensembles for one month in spring—testing data set.

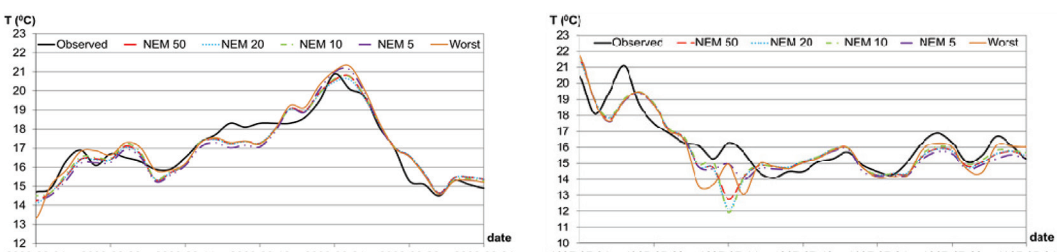

Figure 2. One lead-day water temperature forecasting by means of ANN ensembles for one month in summer—testing data set.

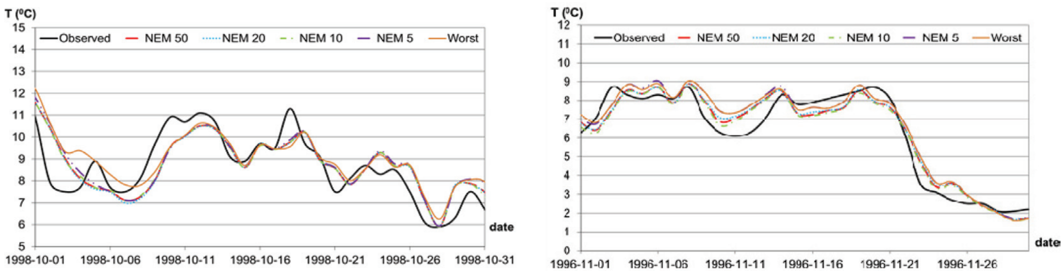

Figure 3. One lead-day water temperature forecasting by means of ANN ensembles for one month in autumn—testing data set.

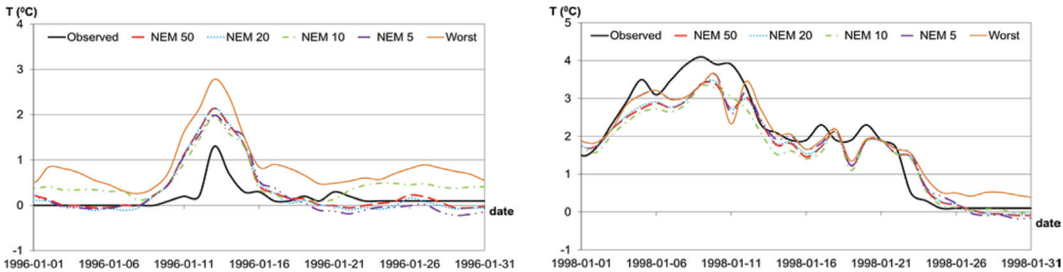

Figure 4. One lead-day water temperature forecasting by means of ANN ensembles for one month in winter—testing data set.

results of forecast by means of the worst among 50 MLP models are included.

Both, the median and the average aggregation methods give very similar results. However, the median aggregation method is chosen since it eliminates the cases of very poor predictions for particular ensemble elements. Such poor predictions may happened when forecasting model is tested on data much different from those in training data set.

The best results are observed for spring and summer months (Figs. 1 and 2). Forecasts for June 1996 and May 2000 by means of all considered ensembles and the worst single model do not differ.

Figure 2 depicts temperature trajectories for 2 summer months in 1997 and 2000. For both months, predictions obtained by means of four different integrated ensembles and single model are very close to each other. Only at July 11th 1997 on may observe sudden drop in the forecasted temperature by all models when the actual temperature is rising. Such unexpected behavior results from rapid flow variations. Interestingly, for this single day the best forecast is made by single "worst" model.

Figure 3 depicts temperature forecast for selected months in autumn. According to our expectation, more elements in ensemble results in better forecast. Here the "worst" case of single prediction really suggests the worst prediction.

One lead-day water temperature forecasting by means of ANN ensembles in winter is probably the most interesting (Fig. 4). All aggregated ensembles lead to different predictions. In winter freezing and melting processes do occur in a river, which are very difficult to be predicted. Frequently the models expect a warming of water, whereas in nature the river is still frozen or, even if the water is ice free, its temperature do not increase.

Systematic prediction error that is occasionally observed is caused by the absence of autoregressive input. In some cases the forecasts may differ from the measurements for a number of consecutive days, frequently due to specific meteorological conditions.

5 CONCLUSIONS

In the present paper the Multi-Layer Perceptron Artificial Neural Networks are applied for regression problem of forecasting water temperature in the Biala Tarnowska River located in Poland.

The problems involved in the proper choice of ANN input variables, ensemble size and ensemble aggregation approach are discussed. According to training-independent testing data the architectures

8-5-1 and 7-5-1-TMAX2 outperform the others tested.

Forecasting performance of the neural networks is satisfactory and generally more elements in aggregated ensemble results in better forecast.

Prediction errors observed in winter are related to the river freezing and melting processes, which are very difficult to be predicted accurately.

ACKNOWLEDGMENTS

This paper has been financed from the polish public budget for science (2013–2015) by MNiSW, Grant No. IP2012040672.

REFERENCES

Bogan, T., Othamer, J., Mohseni, O. & Dtefan, H. 2006. Estimating extreme stream temperatures by the standard deviate method. Journal of Hydrology 317: 173–189.

Breiman, L. 1996. Bagging predictors. Machine Learning 24(2): 123–140.

Caissie, D., 2006. The thermal regime of rivers: a review. Freshwater Biology 51(8): 1389–1406.

Chenard, J.F. & Caissie, D. 2008. Stream temperature modeling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. Hydrological Processes 22: 3361–3372.

Cloke, H. & Pappenberger, F. 2009. Ensemble flood forecasting: A review. Journal of Hydrology, 375: 613–626.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine Learning 40: 139–157.

Erdal, H.I. & Karakurt, O. 2013. Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. Journal of Hydrology 477: 119–128.

Erickson, T.R. & Stefan, H.G. 2000. Linear air/water temperature correlations for streams during open water periods. Journal of Hydrologic Engineering 5(3): 317–321.

Granitto, P.M., Verdes, P.F. & Ceccatto, H.A. 2005. Neural network ensembles: Evaluation of aggregation algorithms. Artificial Intelligence 163: 139–162.

Holmstrom, L. & Koistinen, P. 1992. Using additive noise in back-propagation training. IEEE Transactions on Neural Networks 3: 24–38.

Jeong, D.I., Daigle, D. & St-Hilaire, A. 2012. Development of a stochastic water temperature model and projection of future water temperature and extreme events in the Ouelle river basin in Québec, Canada. River Research And Applications, DOI: 10.1002/rra.

Jeong, D.I. & Kim, Y.O. 2005. Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. Hydrological Processes 19: 3819–3835.

Optiz, D.W, Shavlik J.W. 1996. Actively searching for en effective neural network ensemble. Connection Science 8(3–4): 337–353.

Piotrowski, A.P. & Napiorkowski, J.J, 2013. A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modeling. Journal of Hydrology 476: 97–111.

Piotrowski, A.P., Osuch, M. Napiorkowski, M.J., Rowinski, P.M. & Napiorkowski, J.J. 2014. Comparing large number of metaheuristics for artificial neural networks training to predict water temperature in a natural river. Computers & Geosciences 64: 136–151.

Poole, G.C. & Berman, C.H. 2001. An ecological perspective on in-stream temperature: Natural heat dynamics and mechanisms of human-caused thermal degradation. Environmental Management 27(6): 787–802.

Prechlet, L. 1998. Automatic early stopping using cross-validation: quantifying the criteria. Neural Network 11(4): 761–777.

Sahoo, G.B., Schladow, S.G., Reuter, J.E., 2009. Forecasting streamwater temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. Journal of Hydrology 378: 325–342.

See, L. & Abrahart, R.J. 2001. Multi-model data fusion for hydrological forecasting. Computers & Geosciences 27: 987–994.

Soares, S., Antunes C.H. & Araujo R., 2013. Comparison of a genetic algorithm and simulated annealing for automatic neural network ensemble development. Neurocomputing 121(9): 498–511.

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. Hydrological Processes 26: 1302–1310.

Thimm, G. & Fiesler, E. 1997. High-order and multilayer perceptron initialization. IEEE Transactions on Neural Networks 8(2): 349–359.

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 Resources Research 47: art. nr. W02544.

Vega, M., Pardo, R., Barrado, E. & Deban, L. 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research 32(12): 3581–3592.

Zhang, G., Patuwo, B.E. & Hu, M.Y. 1998. Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting 14: 35–62.

Zheng, J. 2009. Predicting software reliability with neural network ensembles. Expert Systems with Applications 36: 2116–2122.

Zhou, Z.H., Wu, J. & Tang, W. 2002. Ensembling neural networks: Many could do better than all. Artificial Intelligence 137: 239–263.