ISSN 2201-2796

Comparing ensembles of decision trees and neural networks for one-day-ahead streamflow prediction

Onur KARAKURT, Ms.C.
Doctoral Student, Gazi University, Civil Engineering Department
Celal Bayar Bulvarı, 06570, Maltepe
Ankara/ TURKEY
karakurtonur@yahoo.com

Halil Ibrahim ERDAL, Ph.D.
Expert, Turkish Cooperation and Coordination Agency (TİKA)
Atatürk Bulvarı No:15 Ulus
Ankara/ TURKEY
halilerdal@yahoo.com

Ersin NAMLI, Ph.D.
Research Assistant, İstanbul University, Department of Industrial Engineering
34320, Avcılar
Istanbul,TURKEY
enamli@istanbul.edu.tr

Hacer YUMURTACI-AYDOGMUS, Ph.D.
Assistant Professor, Akdeniz University, Alanya Engineering Faculty, Industrial Engineering Department
Merkez Mh. Sigorta Cd. Kestel Alanya
Antalya, TURKEY
hacer913@gmail.com

Yusuf Sait TURKKAN, Ph.D.
Assistant Professor, İstanbul University, Department of Industrial Engineering
34320, Avcılar
Istanbul,TURKEY
ysturkan@istanbul.edu.tr

Abstract: Ensemble learning methods have received remarkable attention in the recent years and led to considerable advancement in the performance of the regression and classification problems. Bagging and boosting are among the most popular ensemble learning techniques proposed to reduce the prediction error of learning machines. In this study, bagging and gradient boosting algorithms are incorporated into the model creation process for daily streamflow prediction.

This paper compares two tree-based ensembles (bagged regression trees BRT & gradient boosted regression trees GBRT) and two artificial neural networks ensembles (bagged artificial neural networks BANN & gradient boosted artificial neural networks GBANN). Proposed ensembles are benchmarked to a conventional ANN (multilayer perceptron MLP). Coefficient of determination, mean absolute error and the root mean squared error measures are used for prediction performance evaluation. The results obtained in this study indicate that ensemble learning models yield better prediction accuracy than a conventional ANN model. Moreover, ANN ensembles are superior to tree-based ensembles.

Keywords: artificial neural networks, bagging (bootstrap aggregating), decision trees, ensembles, gradient boosting, streamflow prediction

I. INTRODUCTION AND LITERATURE REVIEW

Ensemble machines (Anctil and Lauzon 2004; Snelder et al. 2009) have become very popular in the last decade. The constituent members of an ensemble machine are termed as base predictors and the base learning algorithm commonly used in building ensemble machines are artificial neural networks (ANN) and decision trees (DT) (Zhang et al. 2008). Bagging (Chou et al. 2011; Shu and Quarda 2009) and boosting (Snelder et al. 2009, Zaier et al. 2010) are two popular ensemble techniques which come from the same ideology and are designed to overcome problems with weak predictors (Hancock et al. 2005). Bagging (acronym for bootstrap aggregating) is one of the earliest method which was proposed by Breiman (1996) to reduce the prediction error of learning machines. Boosting (also known as arcing) creates a linear combination out of many models for performing supervised learning. Each model is dependent on the preceding models (Friedman 2002). Although bagging and boosting both combine the outputs from different predictors, they differ in the ways to permutate the training data and to combine the predictions coming from their base predictors (Zhang 2008). To sum up, they are among the simplest to implement ensemble techniques, which can reduce variance when combined with the base learner generation, with a good performance (Wang et al. 2011).

In the last decade, ensemble learning methods have been used in the modeling and predicting of hydrologic variables in different research areas. However, bagging and boosting are popular ensemble machine learning techniques; to the best of our knowledge ensemble methods have not been implemented extensively in hydrological time series analysis especially in streamflow estimation. Cannon and Whitfield (2002) investigated the use of ensemble averaging as a part of the streamflow prediction modeling process using neural network models. Tiwari and Chatterjee (2011) explored the potential of wavelet and bootstrapping techniques to develop an accurate and reliable ANN model for daily discharge forecasting and reported that the model, which used the capabilities of both bootstrap and wavelet techniques, was more accurate and reliable . Araghinejad et al. (2011) investigated both generation and combination techniques of artificial neural networks (ANN) ensembles and proposed a new performance function for generating neural network ensembles. ANN ensembles were applied on the peak discharge forecasting of the floods of Red River in Canada and the seasonal streamflow forecasting of Zayandeh-rud River in Iran. The results of the study indicated that the application of the ensemble ANNs through the proposed method can improve the probabilistic forecast skill for hydrological events. Jeong and Kim (2005) used an ensemble neural network (ENN) for forecasting monthly inflows to the Daecheong dam in Korea. The ENN combined the outputs of member networks using the bagging method. The overall results showed that the ENN performed the best among the various rainfall-runoff models. Li et al. (2010) applied bagging to construct modified support vector machines (SVM) models to reduce the variance in the prediction model on streamflow at the Shihmen Reservoir, Taiwan. The results showed that the modified SVM-based model outperformed the prediction ability of the other models. Tiwari and Chatterjee (2010) developed a hybrid wavelet–bootstrap–ANN model for hourly flood forecasting. They used five years hourly water level data for monsoon season from five gauging stations in Mahanadi River basin, India and results indicated that wavelet–bootstrap–ANN model can improve flood forecasting results.

ANN techniques have been widely applied in the modeling and estimating of hydrologic variables: Moghaddamnia et al. (2009) studied on the evaporation estimation methods based on artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS) techniques. Chaves and Chang (2008) proposed an intelligent reservoir operation system based on an evolving artificial neural network (ANN) and applied to the operation of the Shihmen Reservoir in North Taiwan. Besaw et al. (2010) developed and tested two ANNs to forecast streamflow in ungauged basins. The model inputs include time-lagged records of precipitation and temperature were used to train and test the methods. Kisi (2009) applied a conjunction model (neurowavelet) for forecasting daily intermittent streamflow. The comparison results revealed that the suggested model could significantly increase the forecast accuracy of single ANN in forecasting daily intermittent streamflows. Cigizoglu (2004) predicted of daily suspended sediment data by multi-layer perceptrons (MLPs).

The organization of this paper is as follows. The method section is devoted to bagging, gradient boosting and multilayer perceptron. Application and empirical results section describes the data, performance statics, application details and empirical results. Finally, some discussions, conclusions and future study directions are given in the last section.

II. Methods

2.1. Multilayer Perceptron

The output signal for the lth neuron in the nth layer is given by

$$y_l^n(t) = \varphi[\sum_{j=1}^{p} w_{lj}^n(t) y_j^{n-1}(t) + \Psi_l^n] \tag{1}$$

where $\varphi(\cdot)$ is the activation function, w_{lj}^n is the connection weight, t is the time index and $\Psi_l^n = w_{l_0}^n(t)$ is the weighted bias (Barai and Pandey, 1995). The learning law practiced here for weight adaptation is the error back-propagation algorithm. The back-propagation algorithm is a chain learning process which is using for minimizing error (Arditi and Tokdemir, 1999).

In this study, the parameters for MLP were: the number of hidden layers was 1, 3 and 5; the learning rate was 0.2, 0.3 and 0.4; the momentum factor was 0.2, 0.3, and 0.4; and the training epochs were 500, 1,000 and 1,500. The experiments indicated that the best MLP parameters were as follows: the number of hidden layers was 3; the number of the learning rate was 0.3; the momentum factor was 0.2; and the training time was 500.

2.2. Bagging

Bootstrap resampling method (Efron 1979) and aggregating are the basis of Bagging. Variety in Bagging is derived by using bootstrapped replicas of the learning data. The main goal of bagging is minimizing variance in the estimation process (Mert et al., 2012). Different learning sub-datasets are drawn at random with replacement from the entire learning dataset (Wang et al. 2009). Separate models are produced and are used to predict the entire learning data from aforesaid sub-datasets. Many of the original instances may be repeated in the resulting training set whereas others may be omitted (Erdal et al., 2012). Then various estimated models are aggregated by using the mean for regression problems or majority voting for classification problems (Pino et al.

2008). After several regression models are constructed, the average value of the predictions of each regression model gives the final prediction.

A fast decision tree (FDT) was used as the base learning algorithm of BRT and the primary parameters for the FDT were the following: number of folds; the minimum total weight; and number of seeds. The bagging parameters were the size of each bag (as a percentage); the number of iterations; and the number of seeds. In this case, the values for these parameters were 5, 2 & 1 for FDT and 100, 50, and 1 for bagging respectively. Bagged ANN model was created by using bootstrap aggregating of ANN. The parameters of ANN were same as mentioned in multilayer perceptron section.

2.3. Gradient Boosting

Schapire introduced the first boosting algorithm in 1990 and in 1996 Freund and Schapire introduced the AdaBoost algorithm. The main idea of the boosting is improving the performance of prediction using a learning process that generates many models from the same data (Erdal and Karakurt, 2012). Boosting creates an additive structure to improve the final prediction by computing model weights based on the predictions of the previous models (Hancock et al. 2005).

In this study, the gradient boosting technique was used as the boosting algorithm, which was first introduced by Friedman (Friedman 2001, 2002). One of the most powerful meta-learning techniques is gradient boosting, which is a statistical method of fitting an additive model of base functions. Gradient boosting is an important advance in machine learning because it extends and improves the conventional decision trees and ANN models. A more sophisticated version of bagging is described in Friedman (2002).

In this study, the best configuration parameters for the gradient boosting models (GBRT & GBANN) were; the number of iterations was 600 and the shrinking was 0.09. The parameters of ANN were same as described before in multilayer perceptron section.

III. CASE STUDY

In this research, the daily streamflow data of Üçtepe Observation Station (No: 1818) on the Seyhan River in the Mediterranean Sea Region of Turkey is selected. The location of the station is shown in the Fig.1 and the drainage basin lies between the 35^{o}27 17’ E and 37^{o}25 25’N. The Drainage area of the Seyhan River is 13,846 km^2. The daily observed streamflow data is collected from 1982 to 2006. The observed data is for hydrologic years, i.e. the first month of the year is October and the last month of the year is September.

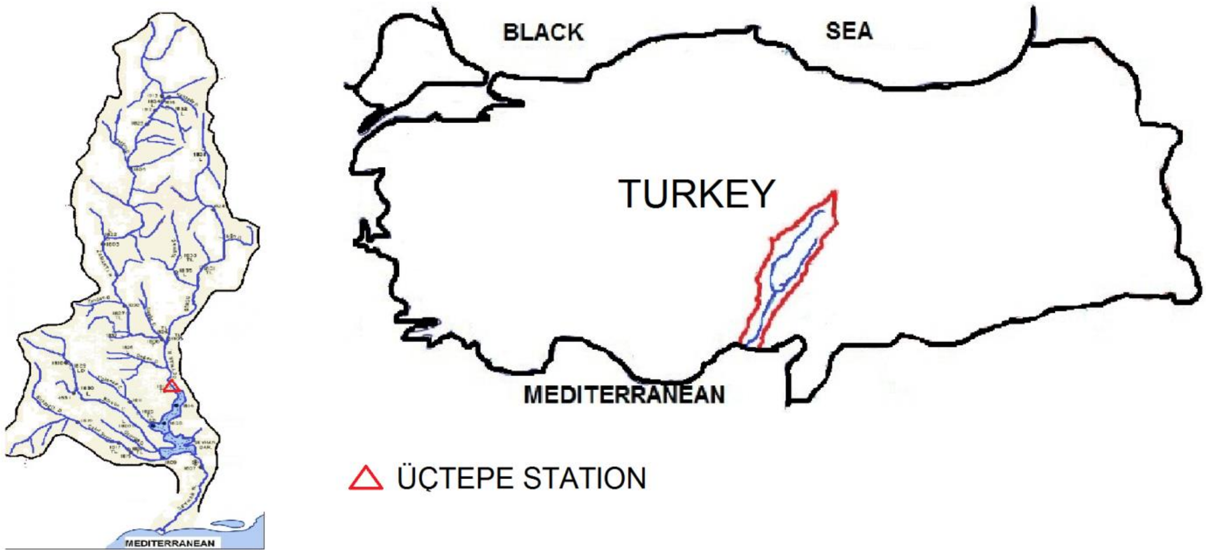


Fig.1 The Üçtepe Station on the Seyhan River in Turkey

IV. Application and Empirical Results

4.1. Application and performance measures

This study investigated bagging and gradient boosting ensembles of artificial neural networks and decision (regression) trees in one day ahead streamflow forecasting. A conventional ANN (multilayer perceptron) employed as the benchmark model. Model structures of bagging and gradient boosting ensembles developed in the present study are shown in Fig. 2 & Fig. 3 respectively.

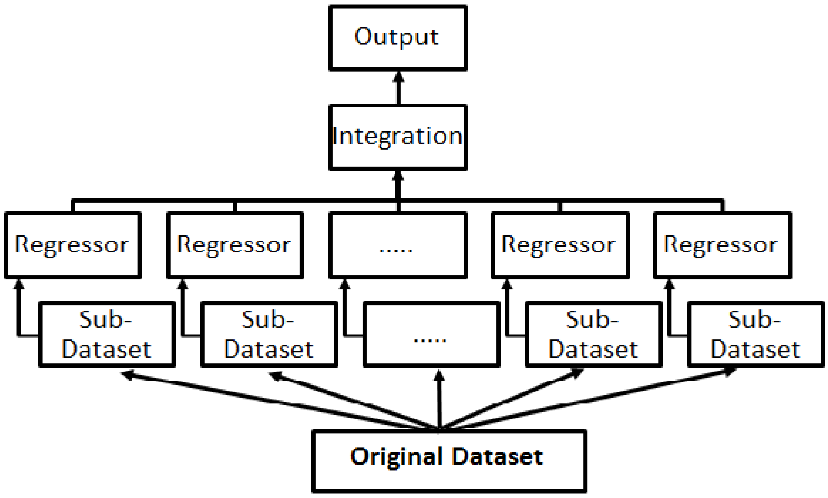


Fig.2 Bagging ensemble model structure

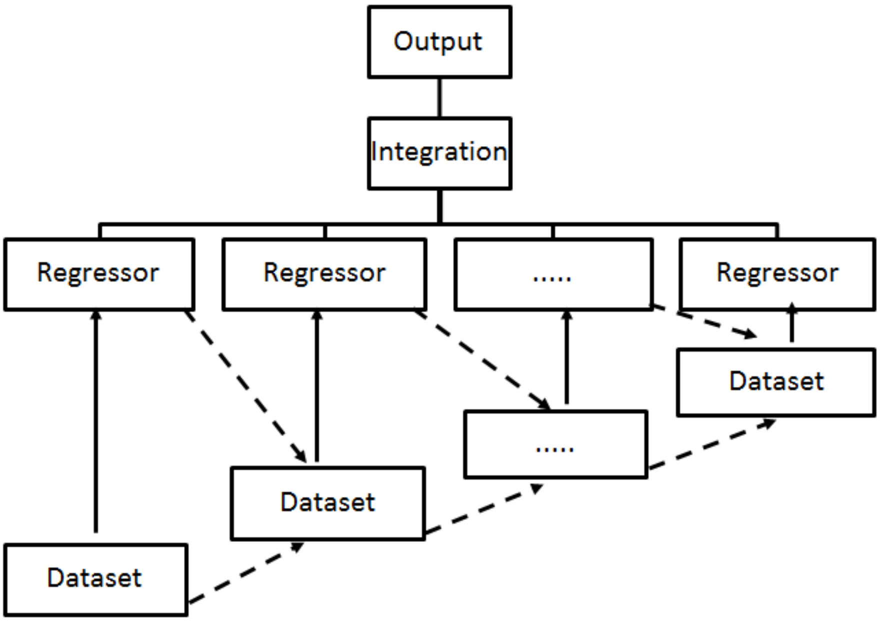


Fig.3. Gradient boosting ensemble model structure

The evaluation was conducted with the k-fold cross validation. Ten-folds cross validation technique was used to choose parameters that yielded the best results. First, the data was randomized and then data was partitioned into three parts as training set (8 distinct folds), cross-validation set (1 fold) and testing set (1 fold). The training set was employed for the model training and the testing set was used to evaluate the accuracy of models. The predictive ensemble machine learning models proposed in this study were evaluated by using tree performance measures:

Coefficient of determination (R^2):

$$R^2 = \left(\frac{n\sum y.y' - (\sum y)(\sum y')}{\sqrt{(\sum y^2) - (\sum y)^2}\sqrt{(\sum y'^2) - (\sum y')^2}} \right)^2 \quad (2)$$

where y = actual value; y' = predicted value; and n = number of data samples.

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{\sum (y' - y)^2}{n}} \quad (3)$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1,n} |y - y'| \quad (4)$$

Also six numerical descriptors were computed to investigate the statistical relation between observed streamflow data and predicted streamflow data;

• Maximum discharge (Max Q)

• Minimum discharge (Min Q)

• Mean of discharge (Mean Q)

• Variance of discharge (Var Q)

• Maximum under-prediction (MUP)

• Maximum over-prediction (MOP)

We purposely do not give the training performance statistics, because fair testing accuracy gives no guarantee for a low test error.

4.2. Empirical Results

The results of all the performance measures (i.e., R^2, MAE & RMSE) and numerical descriptors (i.e., Max Q, Min Q, Mean Q, Var Q, MUP & MOP) for the proposed ensemble machine learning models are summarized in Table 1 and Table 2. The best model for determining R^2 was the BANN model (R^2=0.9347), which had an MOP value of 313.37 m^3/s and MUP value of -686.94 m^3/s. The second best model was GBANN (R^2 = 0.9251) with MOP value of 385.78 m^3/s and UP value of -737.15 m^3/s, and the third best model was GBRT (R^2=0.9228) with MOP value of 465.95 m^3/s and MUP value of -713.86 m^3/s, which obtained results was very similar to GBANN for accuracy R^2. The 4th model BRT (R^2=0.8998) with MOP value of 362.71 m^3/s and MUP value of -678.58 m^3/s was slightly better than the ANN model. The 5th model was ANN, which exhibited the worst predictive capabilities with R^2=0.8942 (MOP=460.79 m^3/s & MUP=-699.94 m^3/s).

Table 1. R^2, MAE and RMSE statistics for five predictive models over 10 folds

	Performance measures		
	R^2	RMSE (m^3/s)	MAE (m^3/s)
ANN	0.8942	74.34	47.04
BANN	0.9347	58.45	31.46
GBANN	0.9251	62.65	36.59
BRT	0.8998	72.88	40.01
GBRT	0.9228	63.20	29.67

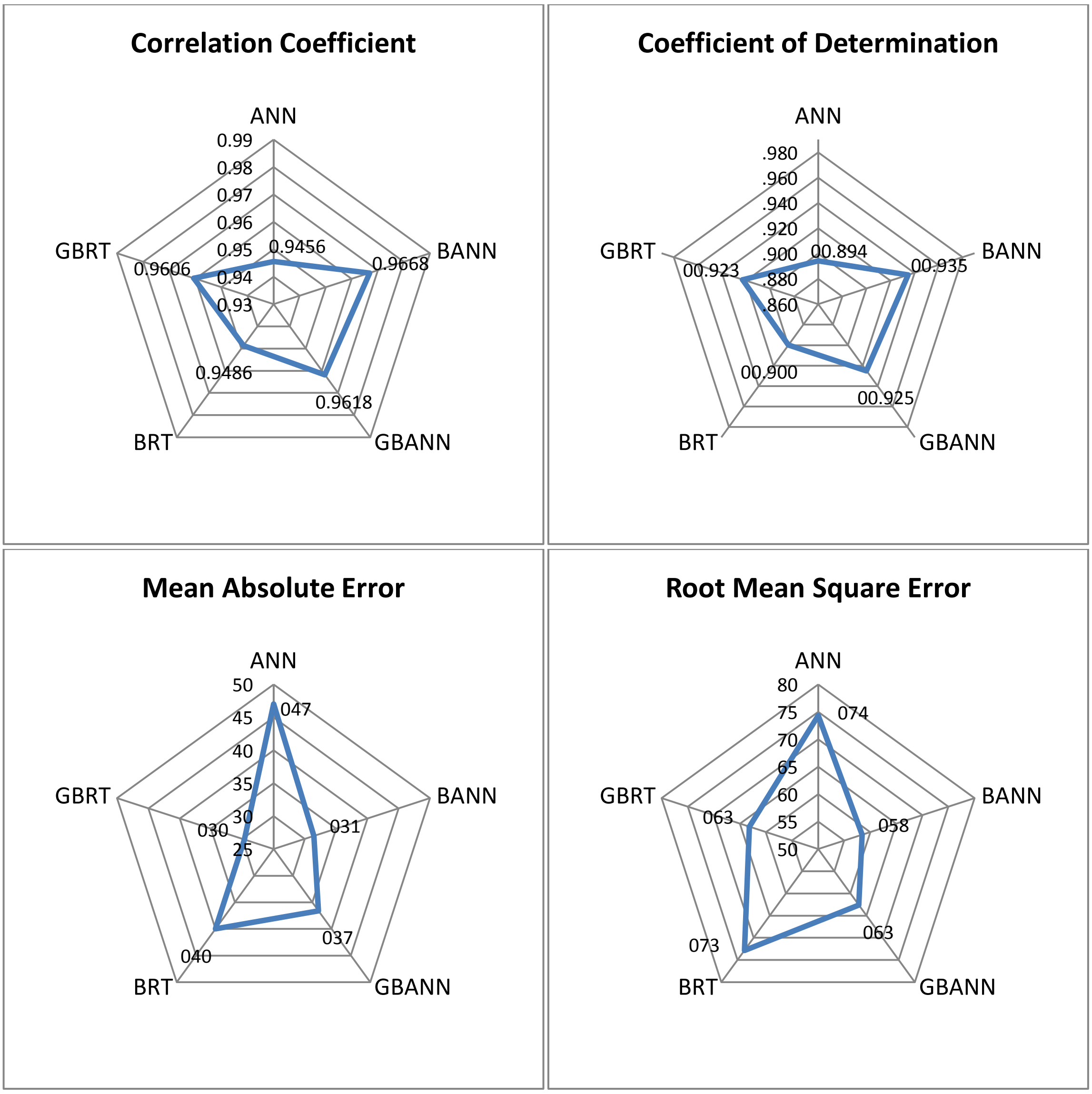
Table 2. Numerical descriptors for five predictive models and observed data

	Numerical descriptors					
	Min Q (m^3/s)	Max Q (m^3/s)	Mean Q (m^3/s)	Var Q	MOP (m^3/s)	MUP (m^3/s)
Observed	49.10	1473.00	216.92	47052.60		
ANN	26.75	1292.24	223.04	49748.26	460.79	-699.94

BANN	37.76	1206.85	223.78	50078.62	313.37	-686.94
GBANN	29.67	1457.12	224.02	50185.52	385.78	-737.15
BRT	59.65	1223.44	217.27	47206.33	362.71	-678.58
GBRT	72.30	1212.78	217.26	47200.76	456.95	-713.86

Clearly, Table 1 indicates the direct relationship between R^2 and RMSE. The best model for minimizing RMSE was BANN (58.45 m^3/s), the 2nd model was GBANN (62.65m^3/s), the 3rd model was GBRT (63.20 m^3/s), the 4th model was BRT (72.88 m^3/s) and finally the worst model was ANN (74.34 m^3/s).

In this study, MAE was also used to evaluate the average prediction ability, which was slightly inconsistent with the other performance measures; the best results were obtained by the GBRT model (MAE=29.67). Fig. 4 shows the performance measures given in Table 1.



ISSN 2201-2796

Fig. 4. Results of the proposed five predictive models

4.3. Distributions of streamflow

The numerical descriptors are depicted with boxplots presented in Fig.5. The boxes indicate the interquartile ranges, the whiskers show the 5th and 95th percentile of observed and predicted data, dots indicate values outside the range and the horizontal line within each boxes indicate the median values. Skewness, a description of streamflow distribution asymmetry, is shown in the figure. Typically, streamflow data are positively skewed, placing the mean in the upper half of the data. The degree of positive skewness illustrates that streamflow typically occurs as many small events with a few large events that elevate the mean. The predictions fit the historical observed data well. The flood seasons typically have peak discharges which create a wide range of skew values. The BANN and GBAN models did a fairly good job at capturing the observed data, however the overall performance of BRT, GBRT and ANN were good when compared to the patterns of the observed streamflow data.

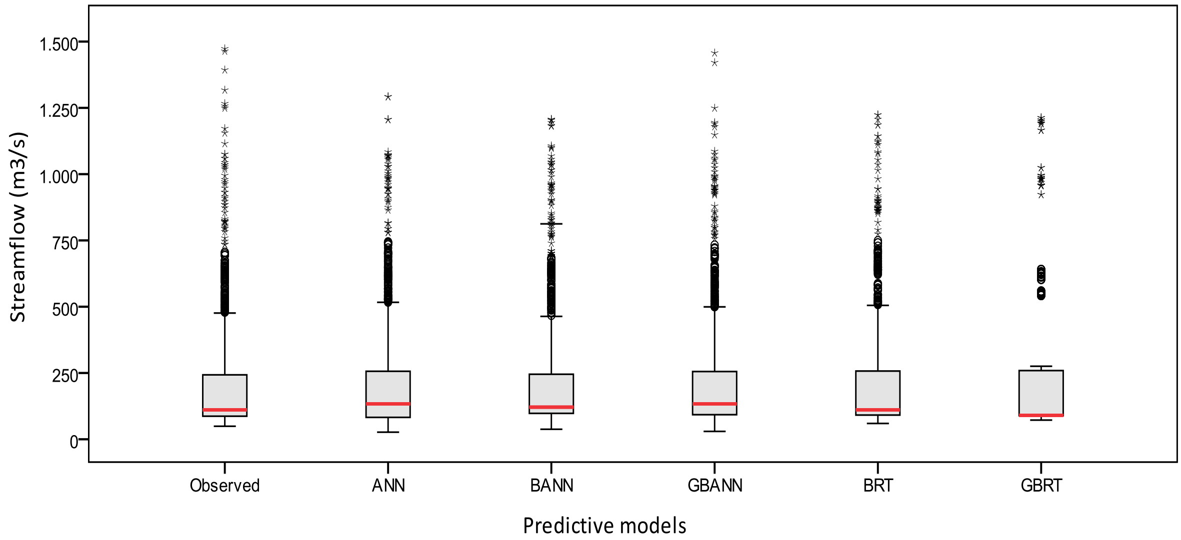


Fig. 5. Box plots of daily observed and predicted streamflows' distributions

The numerical descriptors calculated above (Table 2) for the Seyhan River in Turkey suggests that all models produce statistically similar streamflow predictions and distributions when compared with the measured flow data. Fig. 6 depicts the distributions of the models and the measured flow data statistically. The figure shows that the best results obtained by BANN and GBANN gave a better fit to a straight line than ANN, BRT and GBRT did, which indicated that these techniques were more accurate for predicting streamflow.

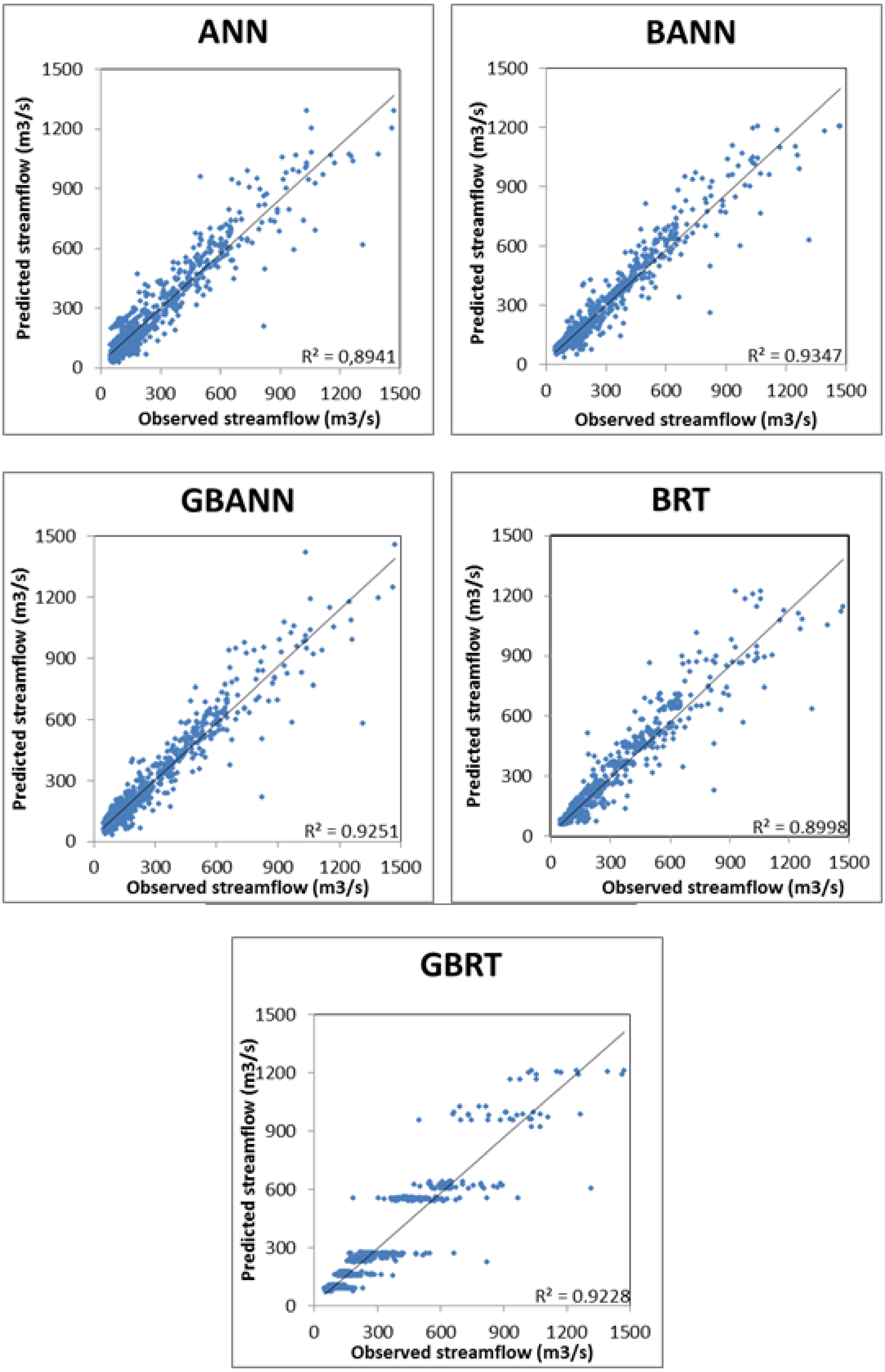


Fig. 6: Observed versus predicted daily streamflow data

V. Conclusion

Ensembles are combinations of several classifiers which are called base predictors and they generally provide better results than their constituent members. Because combining various instances of the same base model can decrease the variance and optimize prediction accuracy.In this study, forecasting of the streamflow measurements, for 1096 days, between October 2004 and September 2007, was conducted by using tree-based ensembles (gradient boosted regression trees GBRT & bagged regression trees (BRT), ANN ensembles (gradient boosted artificial neural networks GBANN & bagged artificial neural networks BANN) and a conventional artificial neural networks (ANN). The proposed predictive ensemble machines were implemented to the dataset by using 10-folds cross validation. The best performing model here was the bagged ANN model for determining coefficient of determination (R^2=0.9347). Second best performing model was gradient boosted ANN which had R^2= 0.9251, closely followed by the gradient boosted RT (R^2=0.9228). Bagged RT (R^2=0.8998) model slightly outperformed than a conventional ANN (R^2=0.8942), which was the worst model. The results indicated that ensemble machine learning models can process the daily streamflow data series better than a conventional ANN and ensembles of ANN yield better results than ensembles of decision trees. In this study, we use only one input-output model for one day ahead streamflow forecasting. Many environmental factors may influence the daily streamflow, however, they are beyond the scope of this study. This could be very important future study issue.

References

Anctil, F. & Lauzon, N. (2004). Generalisation for neural networks through data sampling and training procedures, with applications to streamflow predictions. *Hydrology and Earth System Sciences*. 8(5), 940-958.

Araghinejad, S., Azmi, M. & Kholghi, M. (2011). Application of Artificial Neural Network Ensembles in Probabilistic Hydrological Forecasting. *Journal of Hydrology*, 407(1-4), 94-104.

Arditi,D. & Tokdemir, O.B. (1999). Comparison of Case-Based Reasoning and Artificial Neural Networks. *J. Comput. Civ. Eng.* 13, 162-170.

Barai, S.V. & Pandey P.C. (1995). Vibration Signature Analysis Using Artificial Neural Networks. *J. Comput. Civ. Eng.* 9, 259-266.

Besaw, L.E., Rizzo, D.M., Bierman, P.R. & Hackett, WR. (2010). Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology*, 386(1-4), 27–37.

Breiman, L. (1999). Using adaptive bagging to Debias regressions. Technical Report No.547 of University of California, Berkeley.

Cannon, A.J. & Whitfield, P.H. (2002). Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. *Journal of Hydrology*, 259:136-151.

Chaves, P. & Chang, F.-J. (2008). Intelligent reservoir operation system based on evolving artificial neural networks. *Advances in Water Resour*ces, 31(6), 926–936.

Chou, J.S. & Chiu, C.K., Farfoura, M., Al-Taharwa, I. (2011). Optimizing the Prediction Accuracy of Concrete Compressive Strength Based on a Comparison of Data-Mining Techniques, *J. Comput. Civ. Eng*., 25(3), 242-263.

Cigizoglu, H.K. (2004) Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons. *Advances in Water Resources*, 27(2), 185–195.

Efron B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7 (1): 1–26.

Erdal, HI, Karakurt, O, (2012). Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms Journal of Hydrology doi:10.1016/j.jhydrol.2012.11.015

Erdal, HI, Karakurt, O, Namli, E., (2012). High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform. Engineering Applications of Artificial Intelligence doi:10.1016/j.engappai.2012.10.014

Freund Y. & Schapire R. E. (1996). Experiments with a new boosting algorithm. In Machine Learning: Proceedings of the Thirteenth International Conference, 325-332.

Friedman J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statist*ics, 29(5):1189-1232.

Friedman J.H. (2002). Stochastic Gradient Boosting. *Computational Statistics and Data Analysis*, 38(4):367-378.

Hancock T, Put R, Coomans D, Vanderheyden Y & Everingham Y, (2005). A performance comparison of modern statistical techniques for molecular descriptor selection and retention prediction in chromatographic QSRR studies, *Chemometrics and Intelligent Laboratory Systems*. 76(2), 185-196.

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

Kisi, Ö. (2009). Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *Journal of Hydrologic Engineering*, 14(8), 773-782.

Li, P-H, Kwon,H-H, Sun,L, Lall,U & Kao,J-J. (2010). A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan. International Journal of Climatology, 30:8, 1256-1268.

Mert, A., Kilic, N¸ Akan, A., (2012). Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats. Neural Comput & Applic, DOI 10.1007/s00521-012-1232-7

Moghaddamniaa, A., Goushehb, M.G., Piric, J., Amind, S. & Hane, D. (2009). Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Advances in Water Resources*, 32(1), 88–97.

Pino-Mejias, R., Jimenez-Gamero, M.D., Cubiles-de-la-Vega, M.D. & Pascual-Acosta, A., (2008). Reduced bootstrap aggregating of learning algorithms. *Pattern Recognition Letters*, 29(3), 265–271.

Schapire E. R. (1990). The Strength of Weak Learnability. *Machine Learning*, 5, 197-227.

Shu C. & Ouarda T.B.M.J. (2009). Regional low-flow frequency analysis using single and ensemble artificial neural networks. *Water Resources Research*, 45, W11428, 16 PP., 2009 doi:10.1029/2008WR007196

Snelder, T.H., Lamouroux, N., Leathwick, J.R., Pella, H., Sauquet, E. & Shankar, U. (2009). Predictive mapping of the natural flow regimes of France. *Journal of Hydrology*, 373: 57-67.

Tiwari, M. K. & Chatterjee, C. (2011). A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. *Journal of Hydroinformatics*, 13 (3), 500-519.

Tiwari, M. K. & Chatterjee, C. (2010). Development of an accurate and reliable hourly flood forecasting model using Wavelet-Bootstrap-ANN (WBANN) hybrid approach. *Journal of Hydrology*, 394 (3-4), 458-470.

Wang, G., Hao J., Mab J. & Jiang H., (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, 38 (1), 223-230.

Wang, Y.R., Chen, Y.J. & Huang, C.F.J. (2009). Applying Neural Network Ensemble Concepts for Modelling Project Success. International Symposium on Automation and Robotics In Construction (ISARC 2009), Austin, TX: University of Texas at Austin, 199 - 204.

Zaier I., Shu C., Ouarda T.B.M.J., Seidou O. & Chebana F. (2010). Estimation of ice thickness on lakes using artificial neural network ensembles. *Journal of Hydrology*, 383 (3-4): 330-340.

Zhang CX, Zhang JS & Wang GW. (2008). An empirical study of using Rotation Forest to improve regressors. *Applied Mathematics and Computation*, 618-629.