Leadership in Sustainable Infrastructure

*Leadership en Infrastructures Durables*

Vancouver, Canada

May 31 – June 3, 2017/ *Mai 31 – Juin 3, 2017*

# STREAMFLOW FORECASTING USING ENSEMBLE WAVELET NEURAL NETWORK MODELS

Khalil, Bahaa[1,3], Elenany, Mohamed[2,4,5] and Adamowski, Jan[3]

[1] Civil Engineering Department, Faculty of Engineering – Matarya, Helwan University, Egypt

[2,]Urban Systems Ltd., Edmonton, Canada

[3] Bioresource Engineering Department, McGill University, Canada

[4] Civil Engineering Department, Faculty of Engineering – Al-Azhar University, Egypt

[5] Corresponding author melenany@urbansystems.ca

**Abstract:** Streamflow forecasting studies have shown that data-driven models are simpler, faster to develop, and provide more accurate and precise results than physical or numerical-based models. In this study, three data-driven models were examined for the short-term forecasting of streamflow at Clearwater River in Alberta, Canada, for one-day, three-days, and seven-days lead-times. The three models are artificial neural network (ANN), two models that are based on de-noising the model predictors using the wavelet-transform (W-ANN), and bootstrap-wavelet-ANN (B-W-ANN) model. The total precipitation, air temperature, snow water, and relative humidity were used as predictors. The ANN model performed significantly better after de-nosing the predictors using wavelet-transforms. Overall, the B-W-ANN model performed best for each of the three lead-times. These results highlight the ability of wavelet-transforms to decompose non-stationary data into discrete wavelet-components, highlighting cyclic patterns and trends in the time-series at varying temporal scales, rendering the data readily usable in forecasting. The good performance of the B-W-ANN model highlights the usefulness of ensemble modeling, and ensuring model robustness along with improved reliability by reducing variance.

## 1 INTRODUCTION

Accurate and reliable streamflow forecasting is an important element of sustainable water resources management (Mohanty et al., 2010; Adamowski and Chan, 2011). Two types of models are commonly used for streamflow forecasting, the physically-based numerical models and data-driven models. The main deficiency of the physically-based models is that this type of models requires large amounts of detailed information that may not be always available and are costly to obtain (Adamowski and Chan, 2011; Yoon et al., 2011). However, the main advantage of the physically-based models is that it helps understanding the underlying mechanisms of the system being modeled. On the other hand, the data-driven models main goal is to establish a direct functional relationship between the target variable (e.g. streamflow) and predictors (e.g. precipitation, air temperature) in spite of the detailed physical processes. The main advantage of data-driven models is that this type of models provides accurate and precise forecasts even under limited available data. Thus, in case of limited available data, and that the main goal of the model is the accurate and precise forecasts rather than understanding the underlying mechanisms of the system being modeled, data-driven models are preferable (Moosavi et al., 2013).

Several data-driven models (e.g. autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), multiple linear regression (MLR), artificial neural networks (ANN), and wavelet transform-ANN hybrid models) have been examined for streamflow forecasting. The ARMA, ARIMA and

MLR models are representative of linear models, which have limited ability to model non-linear relations. The artificial intelligent models in general and ANN in particular have the advantage of modelling non-linear relations. The ANN models are often described as “black box” models that can provide relatively accurate forecasts even with limited available data. The ANN models can adapt to dynamic and non-linear changes, but have serious limitations dealing with non-stationary data (Adamowski and Chan, 2011; Khalil et al., 2015).

Recent research studies have focused on hybrid wavelet transforms-ANN models in order to improve the ability of the ANN models to deal with non-stationary data (e.g. Adamowski and Chan, 2011; Ramana et al., 2013; and Khalil et al., 2015). Wavelet transforms are mathematical functions that help analyzing non-stationary data. In addition, several recent studies have shown that Bootstrap-ANN (B-ANN) models are more robust and enhance the performance of single ANNs (e.g. Zaier et al., 2010; Khalil et al., 2011; 2015). The B-ANN models are developed considering different realizations of the training datasets, which allows the B-ANN models to perform consistently even if the nature of the data are to change in the future. However, to the best of the authors’ knowledge, in spite of the usefulness of the B-W-ANN models, they have never been examined for short-term flood forecasting.

The main goal of this study is to examine B-ANN and B-W-ANN models for short-term flood forecasting (one-day, one-week, and one-month ahead) at the Clearwater River at Draper, Alberta, Canada. These hybrid models that employed wavelet transforms to de-noise the model predictors were compared to conventional ANN models.

## 2 STUDY AREA

In this study, the streamflow of Clearwater River at Draper in the province of Alberta was forecasted. The area served by the Clearwater River drains is about 30,800 km² and Broach Lake in Saskatchewan, at an elevation of 460 m, is the headwaters of this River. Around 20% of the total Athabasca basin to Fort McMurray covers the drainage area of the Clearwater River (Figure 1).

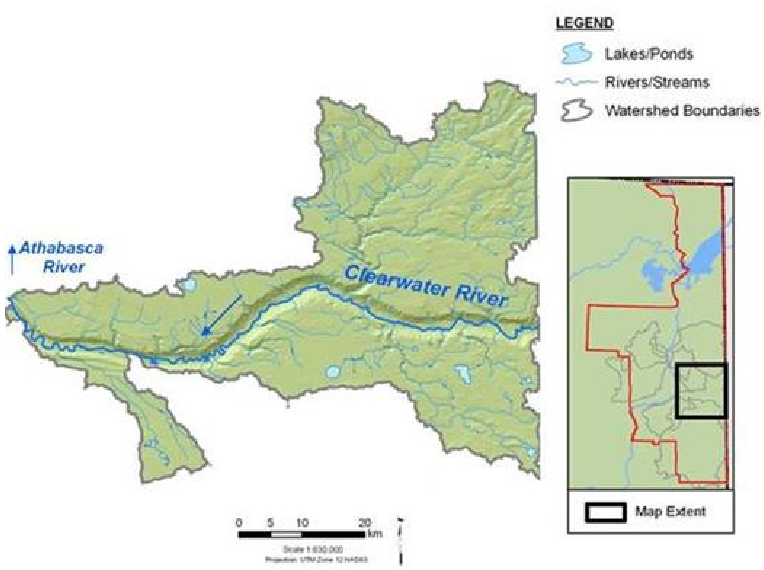


Figure 1: Clearwater River Watershed

The streamflow data has been collected by Water Survey of Canada (WSC) at station 07CD001 for 54 years from January 1st, 1960 to December 31st, 2013. The high flows normally happen in spring as a result of the combination of melted snow with seasonal rainfalls and the Low flows are almost recorded in winter when the precipitation is in the form of snow. In summer, the extreme rainfall has resulted in recorded floods, as well as some low flow periods which may be caused by the dry conditions in watershed. The maximum daily flow recorded in this period is 790 $m^3/s$ on April 30, 1974, and the minimum recorded value is 26.2 $m^3/s$ on February 19, 1982 ("Clearwater River Hydrological Profile", 2016). Different climatological and hydrometrical data were obtained from station T089R09W4 for the period from 1st January, 1960 to 31st December 2013. The available variables are precipitation, snow meltwater, relative humidity, minimum air temperature, and maximum air temperature.

## 3 METHODOLOGY

In hydrological events, the climatological variables may affect the streamflow with time delay. To find the most appropriate predictors with appropriate lag time, the correlation coefficients between each of the available variables and the streamflow at different lags were calculated as well as the streamflow autocorrelation (Table 1). To identify the optimal set of predictors to forecast the streamflow at 1-day, 3-days, and 7-days ahead, cross-correlation, auto-correlation and partial-correlation analyses were applied. Given that in hydrological processes, the climatological variables may affect the streamflow with time delay, cross-correlation was applied considering different lag steps as follows: (i) cross-correlation between the streamflow and all available predictors at different delay steps, and auto-correlation were applied; (ii) the predictor with the highest correlation coefficient was considered the best predictor and should be included in the optimal set of variables; (iii) partial-correlation was then applied between the streamflow and the remaining variables considering the best predictor was included in the optimal set of predictors; (iv) based on partial-correlation results, the second best predictor can be identified as the variable with the highest partial-correlation coefficient; (v) the partial-correlation was again applied with the variables that showed significant partial-correlation coefficient given the two most important predictors included and so forth.

### 3.1 Artificial Neural Network (ANN)

One of the commonly used ANN in hydrological studies is the multilayer perceptron (MLPs) ANN originally proposed by Rumelhart and McClelland (1986), which was employed in this study. The MLP consists of three layers, namely: an input layer, a hidden layer, and an output layer. The size of the neurons in the input layer is the size of the predictors considered to forecast the streamflow. The forecasted streamflow is provided through the single neuron in the output layer. For the size of the neurons in the hidden layer, Shu and Ouarda (2007) recommended that the size of hidden neurons should be less than twice the size of predictors (nI); while, Maier and Dandy (2001) recommended using the following formula to identify the maximum size of the neurons in the hidden layer (nH):

[1] $$nH = \min(2\,nI + 1;\ nTR / nI+1)$$

where nTR is the size of training sample. In this study, the optimal size of hidden neurons was determined based on a trial-and-error process involving varying the size of hidden neurons from one to the minimum of the two criteria proposed by Shu and Ouarda (2007) and Maier and Dandy (2001). In this study, to enhance the nonlinear approximation ability of the ANN, the tan-sigmoid transfer function was used in the hidden neurons, while the linear transfer function was used in the output neuron. The Levenberg–Marquardt (LM) ANN training algorithm (Hagan and Menhaj, 1994) was employed in this study, because it is one of the most efficient and fast algorithms for training ANNs (Khalil et al., 2011).

### 3.2 Wavelet-Artificial Neural Network (W-ANN)

Several studies have shown that data preprocessing using wavelet transforms (WT) improves ANN performance for monthly reservoir inflow (Coulibaly et al., 2000), drought forecasting (Kim and Valdés, 2003), suspended sediment forecasting (Partal and Cigizoglu, 2008), streamflow forecasting (Adamowski and Sun, 2010), and groundwater level forecasting (Adamowski and Chan, 2011; Khalil et al., 2015).

There are two main types of WT: continuous wavelet (CWT) and discrete wavelet transform (DWT). For streamflow forecasting, input data are usually discretely sampled, which makes the DWT more appropriate. The “Haar a Trous” WT introduced by Zheng et al. (1999) was used in this study as the most appropriate DWT for forecasting studies (Mallat, 1999). Consider $C_0$ as the original time series and $C_s$ as the approximation component at scale s, then the WT ($W_s$) (details component) is given by the following set of equations (Murtagh et al., 2004):

[2] $$W_s(k) = C_{s-1}(k) - C_s(k)$$

[3] $$C_s(k) = \frac{1}{2}\left(C_{s-1}(k) + C_{s-1}\left(k - 2^s\right)\right)$$

where $k$ is the position (within the time series) at which the wavelet transform is calculated, $h(l)$ is the low pass filter, and $l = \{0.5, 0.5\}$. The W-ANN model is based on ANN, where the input variables were preprocessed using the "Haar a Trous" wavelet analysis. In the W-ANN, the input variables were the approximation components of the input variables instead of the original input variables used in the conventional ANN model.

In general, "*the DWT requires that the input data have a number of values that is an integer power of two*" (Khalil et al., 2015). Given that the size of available records is around 19724 (daily records from January 1st, 1960 to December 31st, 2013), the resolution levels is limited to 14 ($2^{14}$ = 16384). For the first decomposition level, the time series under analysis is transformed into two components, an approximation component (A1) and a detailed component (D1). The approximation component is considered the de-noised component. For the second decomposition level, the A1 is transformed into a detailed component (D2) and a second level approximation component (A2). Thus, after the second decomposition level, the original input time series is transformed into two detailed components (D1 and D2) and an approximation component (A2), and so on for higher decomposition levels (e.g., Third decomposition level provides D1, D2, D3; and A3; Fourth decomposition level provides D1, D2, D3, D4, and A4, and so on). In the W-ANN models, the A components of the input variables were used instead of the original variables. The best decomposition level was identified for the W-ANN using the Pearson correlation coefficient ($r$) between the stream flow and different approximation components (A1 to A14).

[4] $$r = \sum_{i=1}^{n}\left(A_i - \overline{A}\right)\left(y_i - \overline{y}\right) \Bigg/ \sqrt{\sum_{i=1}^{n}\left(A_i - \overline{A}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}$$

where $y_i$ and $A_i$ are, respectively, the streamflow and approximation component values at time *i*, and $\overline{y}$ and $\overline{A}$ are the mean values of the streamflow and approximation component, respectively; $n$ is the size of records. Among the 14 approximation components that were obtained based on the 14 decomposition levels, the approximation component that showed the highest correlation with the streamflow was selected as the predictor.

### 3.3 Bootstrap- Wavelet-Artificial Neural Network (B-W-ANN)

The B-W-ANN consists of a group of W-ANN members, where each of these W-ANN members is trained for the same problem, and their results are combined to produce the B-W-ANN output. The construction of the B-W-ANN consists of two steps (Merz, 1998): the first step is the generation of the W-ANN members constructing the ensemble; and the second step is to combine the outputs of the W-ANN members to produce the B-W-ANN output. In this study, the bootstrap aggregation technique was selected to generate the ensemble members, and simple averaging was used to combine their outputs.

In the bootstrap, assume that the size of the training dataset TR is nTR, ([X1, y1],…., [XnTR, ynTR]), where X is the predictors matrix and y is the streamflow vector. To generate different bootstrap replicates (W-ANN members) (TB), for each member, random sampling with replacement of nTR cases from the original TR data was applied, while all of the nTR cases have equal probability (1/nTR) to be selected. Thus, each bootstrap replicate (TB) may have many cases that are selected more than once, while other cases are never selected. The W-B-ANN members can then be trained using the various TB generated by this process. The B-W-ANN members have the same structure, transfer function, size of hidden neurons, and training algorithm as defined for the ANN and W-ANN models.

For the optimal size of the bootstrap members, several studies have suggested that ten members can attain sufficient generalization ability (e.g. Agrafiotis et al., 2002), while, Opitz and Maclin (1999) have shown improvement of the generalization ability when the size increases to 15. Zaier et al. (2010) and Khalil et al. (2015) used 20 members for the estimation of ice thickness on lakes, and groundwater levels, respectively. Khalil et al. (2011) used 15 members for the estimation of water quality characteristics at ungauged sites. In this study, the optimal size of the bootstrap members was identified using a trial-and-error approach, by evaluating the performance of the B-W-ANN model using sizes ranging from two to 20.

### 3.4 Evaluation Procedures and Criteria

The split-samples validation approach was utilized for the validation of the three models. The split-samples approach consists of dividing the available data into two main sets, a set for the model calibration (44 years) and the second set for model validation (10 years), and the calibration data set was divided into 80% for training and 20% for training testing. Using the validation data set, the streamflow forecasts were evaluated against the observed values. The evaluations were conducted using the following three indices: the bias (BIAS), root mean squared error (RMSE), and the Nash-Sutcliffe criterion (NASH). These indices were computed according to the following equations:

[5] $$\mathrm{BIAS} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i - y_i$$

[6] $$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

[7] $$\mathrm{NASH} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where $y_i$ and $\hat{y}_i$ are, respectively, the observed and forecasted streamflow values at time *i*; $\bar{y}$ and $\bar{\hat{y}}$ are, respectively, the mean values of the observed and forecasted streamflow values; $n$ is the size of records.

## 4 RESULTS

Table 1 shows the correlation coefficients between streamflow and different possible predictors at lag-time from zero to 10 days. The correlation coefficients shown in Table 1 are all significant at 95% confidence level. Results of correlation analysis between the precipitation and streamflow showed an increasing pattern from lag-0 (r = 0.0894) to lag-7 (r = 0.1766), which indicates that rainwater from different places within the catchment may take around seven days to reach the main stream. Thus, the best precipitation predictor is the precipitation at lag-7. Similarly, the minimum air temperature at lag-3 showed the best correlation coefficient for the Tmin ($Tmin_{t-3}$). Thus, according to the cross-correlation and auto-correlation analyses, to forecast the streamflow for any lead-time ($Q_{t+1}$; $Q_{t+3}$; and $Q_{t+7}$) the best predictor is the stream flow at time (t) ($Q_t$). Partial-correlation analysis was then applied between the streamflow and the remaining predictors given $Q_t$ is considered the best predictors and included in the set of optimal variables. Results of the partial-correlation analysis are presented in Table 2.

Table 2 showed that the $S_t$ and $Tmax_t$ are not significantly correlated, while $Tmin_t$, $P_{t-7}$, and $RH_t$ are significantly correlated, where $P_{t-7}$ showed the highest partial correlation coefficient (0.266). Results of the partial-correlation analysis between streamflow and each of the $Tmin_t$ and $RH_t$ giving that $Q_t$ and $P_{t-7}$ were included are given in Table 2, which showed that both the $Tmin_t$ and $RH_t$ can be excluded. Thus for the stream flow forecasting at 1-day ahead the optimal set of predictors are $Q_t$ and $P_{t-6}$. Similarly, for the forecasting of streamflow three-days ahead ($Q_{t+3}$), $Q_t$ and $P_{t-4}$ were selected as the optimal set of predictors, and for the forecasting the streamflow seven-days ahead ($Q_{t+7}$), $Q_t$ and $P_t$ were the best predictors.

Table 1: Cross-Correlation and auto-correlation results

| Lag-day | Q | Snow Meltwater | Temp Max | Temp Min | Precipitation | Relative Humidity |
|---|---|---|---|---|---|---|
| Lag-0 | 1.0000* | **-0.4598*** | **0.4866*** | 0.4948* | 0.0904* | **-0.1940*** |
| Lag-1 | **0.9941*** | -0.4573* | 0.4846* | 0.4954* | 0.1185* | -0.1843* |
| Lag-2 | 0.9806* | -0.4539* | 0.4826* | 0.4958* | 0.1425* | -0.1760* |
| Lag-3 | 0.9627* | -0.4495* | 0.4812* | **0.4959*** | 0.1589* | -0.1706* |
| Lag-4 | 0.9418* | -0.4441* | 0.4800* | 0.4955* | 0.1683* | -0.1677* |
| Lag-5 | 0.9189* | -0.4376* | 0.4786* | 0.4943* | 0.1730* | -0.1669* |
| Lag-6 | 0.8948* | -0.4301* | 0.4770* | 0.4927* | 0.1751* | -0.1678* |
| Lag-7 | 0.8702* | -0.4217* | 0.4752* | 0.4910* | **0.1765*** | -0.1694* |
| Lag-8 | 0.8455* | -0.4124* | 0.4734* | 0.4889* | 0.1761* | -0.1717* |
| Lag-9 | 0.8210* | -0.4024* | 0.4715* | 0.4864* | 0.1749* | -0.1742* |
| Lag-10 | 0.7971* | -0.3917* | 0.4694* | 0.4839* | 0.1726* | -0.1770* |

*Significant correlation coefficient at 95% confidence level

Table 2: Partial-Correlation Results

| Included | $S_t$ | $Tmax_t$ | $Tmin_{t-3}$ | $P_{t-7}$ | $RH_t$ |
|---|---|---|---|---|---|
| $Q_t$ | -0.002 | 0.009 | 0.038* | **0.266*** | 0.080* |
| $Q_t$ & $P_{t-7}$ | | | -0.004 | | -0.008 |

*Significant correlation coefficient at 95% confidence level

### 4.1 Artificial Neural Network (ANN)

Based on cross-correlation, auto-correlation and partial-correlation analyses the optimal set of predictors for 1-day streamflow forecasting ($Q_{t+1}$) consists of $Q_t$ and $P_{t-7}$, while for $Q_{t+3}$ the predictors are $Q_t$ and $P_{t-4}$, and for $Q_{t+7}$ the best predictors are $Q_t$ and $P_t$. The size of the neurons in the hidden layer was identified using a trial-and-error approach from one hidden neuron to nH, which was equal to ten neurons based on equation 1. The selection of the optimal size of neurons in the hidden layer was based on the performance of the ANN model (Table 3). Results of the performance measures (Table 3) indicated that using three hidden neurons provided the best performance for $Q_{t+1}$ forecasting with the lowest RMSE values and highest NASH. Similarly, for $Q_{t+3}$ and $Q_{t+7}$, the optimal size of hidden neurons was five and six hidden neurons, respectively. Table 4 shows the performance measures for the three ANN models.

Table 4 shows that the performance of the 1-day ANN model is more accurate and precise than the 3-days ANN model, and the 3-days ANN model is more accurate and precise than the 7-days ANN model. As a measure of precision, the RMSE values deteriorated from 9.598 $m^3$/sec for the 1-day ANN model, to 25.75 $m^3$/sec for the 3-days ANN model, and 48.23 $m^3$/sec for the 7-days ANN model. Similarly, as a measure of accuracy, the BIAS values were -0.04 $m^3$/sec for 1-day ANN model, -0.08 $m^3$/sec for 3-days ANN model, and -0.24 $m^3$/sec for the 7-days ANN model, which indicate that the accuracy deteriorate slightly as the lead-time increase, but the three models showed minor underestimation (Figure 2). This deterioration in the model performance is directly related to the reduction in the auto-correlation coefficients between the target variables $Q_{t+1}$, $Q_{t+3}$, and $Q_{t+7}$ and the predictor $Q_t$, which were 0.9941, 0.9627, and 0.8702, respectively (Table 1).

Table 3: Performance measures for 1-day ANN forecasting model using different sizes of hidden neurons

| Neurons | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| RMSE | 9.612 | 9.593 | **9.598** | 9.615 | 9.620 | 9.623 | 9.649 | 9.665 | 9.606 | 9.656 |
| BIAS | -0.037 | 0.001 | **-0.040** | -0.062 | -0.070 | -0.030 | -0.085 | 0.264 | -0.092 | -0.072 |
| NASH | 0.9898 | 0.9899 | **0.9899** | 0.9898 | 0.9898 | 0.9898 | 0.9898 | 0.9897 | 0.9898 | 0.9897 |

Table 4: Performance measures for 1-day, 3-days, and 7-days streamflow ANN forecasting models

| Model | ANN | | | W-ANN | | | B-W-ANN | | |
|---|---|---|---|---|---|---|---|---|---|
| L- time | 1-day | 3-days | 7-days | 1-day | 3-days | 7-days | 1-day | 3-days | 7-days |
| RMSE | 9.5981 | 25.7541 | 48.2235 | 8.733 | 21.955 | 40.0355 | 8.635 | 11.823 | 13.727 |
| BIAS | -0.0405 | -0.0803 | -0.3408 | -0.008 | -0.046 | -0.1682 | 0.002 | 0.003 | 0.007 |
| NASH | 0.9899 | 0.9270 | 0.7445 | 0.9957 | 0.9808 | 0.8427 | 0.9980 | 0.9896 | 0.9793 |

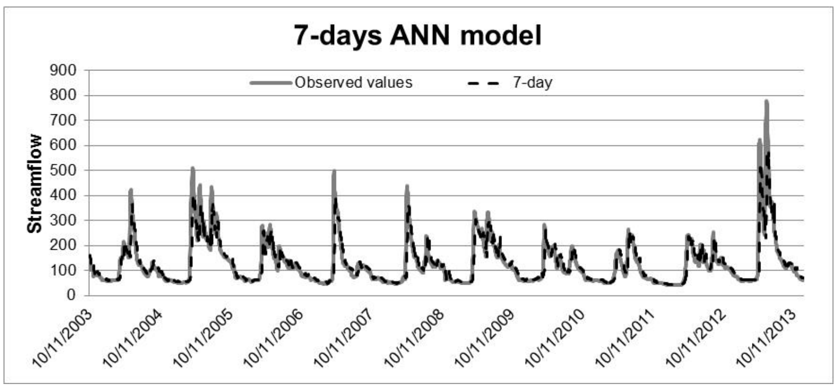


Figure 2: Observed and forecasted 7-days ANN model

## 4.2 Wavelet-Artificial Neural Network (W-ANN)

For the W-ANN models, the best decomposition level and best size of hidden neurons were identified based on the indices of model performance using the trial-and-error approach. The performance measures were computed for all possible combinations of the size of hidden neurons (1 to 10 neurons) and decomposition levels (1 to 14 levels). Based on the performance measures, for the 1-day W-ANN forecasting model, Table 5 shows 140 RMSE values that represent all possible combinations of 1 to 10 hidden neurons and 1 to 14 decomposition levels. From Table 5, the minimum RMSE value was 8.73 $m^3$/sec corresponds to the first decomposition level and four neurons in the hidden layer.

Similarly, for the 3-days and 7-days ($Q_{t+3}$ and $Q_{t+7}$) W-ANN forecasting models, the first decomposition level was enough to de-noise the predictors, while the size of neurons in the hidden layers were 3 and 5, respectively. Table 4 shows the performance measures for the three W-ANN models, which indicate that similar to the ANN models, the performance of the 1-day W-ANN model was more accurate and precise than the 3-days W-ANN model, and the 3-days W-ANN model was more accurate and precise than the 7-days W-ANN model. Based on the BIAS, RMSE and NASH values, comparing the performance of the ANN models with that of the W-ANN models indicated the better performance of the W-ANN models at any of the three lead-times considered. Figure 3 showed the observed and forecasted 7-days W-ANN model.

The significant improvement in the performance of the three W-ANN models over the ANN models was due to the use of the approximation components (de-noised components) of the predictors rather than the original values, which confirmed the usefulness of using the wavelet transform to de-noise the predictors.

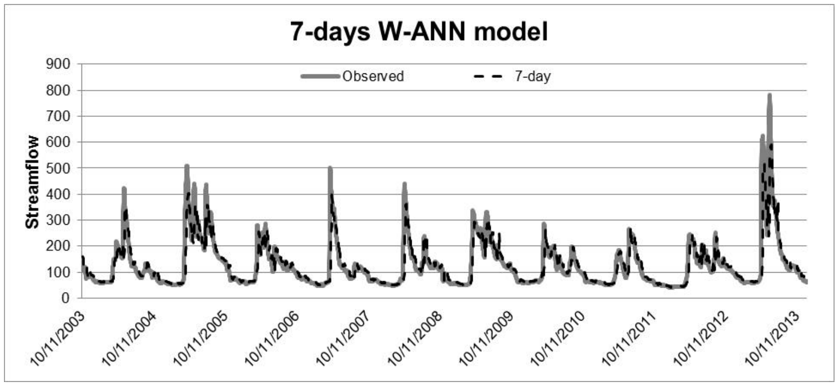


Figure 3: Observed and forecasted 7-days W-ANN model

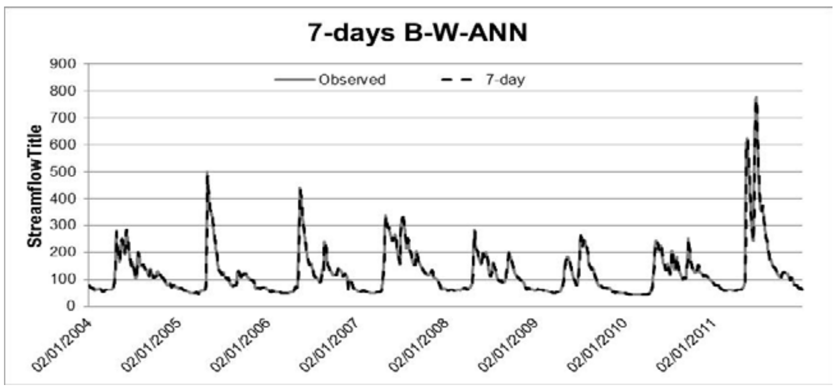


Figure 4: Observed and forecasted 7-days B-W-ANN model

### 4.3 Bootstrap-Wavelet-Artificial Neural Network (B-W-ANN)

For each of the three B-W-ANN models, the size of the bootstrap was identified using a trial-and-error approach, by evaluating the performance of the B-W-ANN models using bootstrap sizes ranging from two to 20. For instance, Figure 5 showed the RMSE values corresponding to the 1-day B-W-ANN forecasting models, using two to 20 bagging replicates, where a size of 16 showed the lowest RMSE value.

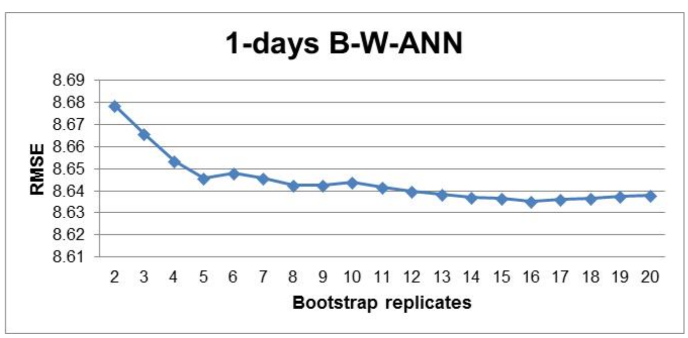


Figure 5: RMSE values for different bootstrap replicates for 1-day B-W-ANN forecasting model

Table 4 showed that all the performance measures corresponding to the B-W-ANN models were better than those corresponding to any of the ANN and W-ANN models at any of the three lead-times considered in this study. It should be emphasized that the ensemble technique was originally designed for cases of limited data. Although the size of data used (19000 records) cannot be considered limited, the use of the ensemble technique improves the W-ANN performance and provides more robust models as indicated by Figure 4. This can also be seen with the RMSE, BIAS and NASH values (Table 4).

Table 4: RMSE values for 1-day W-ANN forecasting model for all possible combinations of the size of hidden neurons and decomposition levels

| | Size of neurons in the hidden layer | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Decomposition level | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 8.80 | 8.80 | 8.76 | **8.73** | 8.81 | 8.77 | 8.80 | 8.81 | 8.77 | 8.77 |
| 2 | 13.78 | 13.78 | 13.74 | 13.73 | 13.80 | 13.74 | 13.77 | 13.82 | 13.75 | 13.74 |
| 3 | 21.09 | 21.05 | 20.99 | 21.10 | 20.97 | 20.99 | 21.01 | 20.96 | 21.09 | 21.01 |
| 4 | 32.67 | 32.77 | 32.69 | 32.84 | 32.84 | 32.62 | 32.93 | 32.72 | 32.72 | 32.81 |
| 5 | 46.44 | 45.23 | 45.17 | 45.26 | 45.79 | 46.11 | 45.79 | 46.61 | 46.20 | 46.06 |
| 6 | 59.71 | 59.15 | 58.71 | 59.71 | 61.83 | 60.43 | 61.21 | 60.50 | 61.31 | 61.63 |
| 7 | 74.66 | 75.02 | 74.68 | 74.74 | 75.08 | 75.52 | 75.85 | 219.56 | 80.22 | 77.46 |
| 8 | 89.04 | 88.40 | 88.42 | 88.61 | 88.57 | 88.33 | 88.77 | 90.14 | 89.98 | 91.09 |
| 9 | 93.44 | 98.99 | 98.45 | 99.87 | 96.94 | 99.53 | 97.07 | 98.35 | 95.70 | 99.10 |
| 10 | 95.42 | 94.10 | 94.72 | 94.09 | 94.28 | 94.26 | 94.30 | 94.64 | 95.06 | 94.74 |
| 11 | 97.02 | 95.53 | 94.21 | 93.82 | 95.69 | 96.89 | 97.89 | 97.59 | 98.14 | 102.29 |
| 12 | 97.55 | 96.24 | 96.58 | 98.55 | 98.17 | 98.82 | 97.97 | 98.24 | 99.76 | 98.37 |
| 13 | 98.10 | 98.70 | 97.14 | 99.48 | 98.18 | 98.32 | 100.53 | 101.43 | 103.48 | 100.55 |
| 14 | 97.22 | 97.49 | 98.91 | 99.37 | 101.24 | 98.85 | 99.77 | 98.53 | 110.03 | 99.90 |

## 5 Conclusions

Three data-driven models were examined in this study for short-term streamflow forecasting. The three models are the artificial neural network (ANN), wavelet-ANN (W-ANN), and bootstrap-W-ANN (B-W-ANN) model. The ANN model main goal was to establish a functional relationship between streamflow and selected predictors, historical streamflow records and precipitation. Selection of the ideal predictors was based on cross-correlation, auto-correlation, and partial-correlation. For the wavelet-based models (W-ANN and B-W-ANN), wavelet analysis was employed to de-noise the selected predictors and ANN was then used to approximate the functional relationship between the streamflow and the de-noised predictors. The three models were developed and applied using data from the Clearwater River at Draper, Alberta. The split-samples validation approach was applied to evaluate the performance of the three models using three performance indices, the root mean square error (RMSE), bias (BIAS), relative RMSE, relative BIAS, and NASH.

The results showed that the B-W-ANN models outperformed the ANN and W-ANN models for the forecasting of the streamflow at the three lead times considered in this study. Results also showed that the performance of the ANN models was significantly improved by de-nosing the predictors using wavelet analysis. More generally, it was demonstrated that data-driven models such as the W-ANN and B-W-ANN models are an interesting alternative to numerical-based models where streamflow forecasting and consequent management activities need to be evaluated on a short-term basis.

The three models developed in this study were designed for short-term streamflow forecasting; exploring these types of models for long-term forecasting is recommended. It is also recommended to examine other artificial intelligence modeling methods such as support vector regression (SVR) and W-SVR.

**Acknowledgements**

Financial support provided by NSERC Discovery Grant held by Jan Adamowski is acknowledged.

**References**


Adamowski, J. and Chan, H.F. 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol., 407(1-4): 28-40.

Adamowski, J. and Sun, K. 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J. Hydrol., 390(1-2): 85-91.

Agrafiotis, D.K., Cedeno, W. and Lobanov. V.S. 2002. On the use of neural network ensembles in QSAR and QSPR. J. Chem. Inf. Comput. Sci., 42: 903–911.

Clearwater River Hydrological Profile. 2016. Regional Aquatics Monitoring Program. Retrieved 10 December 2016, from ( http://www.ramp-alberta.org/river/hydrology/river+hydrology/clearwater+river.aspx

Coulibaly, P., F. Anctil, and B. Bobe´e. 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. J. Hydrol., 230: 244–257.

Hagan, M.T. and Menhaj, M. 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Networks, 5(6): 989–993.

Khalil, B., S. Borda, J. Adamowski, B. Ozga-Zielinski, and A. Donohoe. 2015. Short-term forecasting of groundwater levels under tailings recharge using Wavelet Ensemble Neural Network models. Hydrogeology Journal, 23(1): 121-141.

Khalil, B., T.B.M.J. Ouarda, and A. St-Hilaire. 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J. Hydrol., 405: 277-287.

Kim, T.W. and Valdés, J.B. 2003. Nonparametric approach for estimating return periods of droughts in arid regions. J. Hydrol. Eng., 8 (5): 237-246.

Maier, H.R. and Dandy, G.C. 2001. Neural network based modelling of environmental variables: a systematic approach. Mathematical and Computer Modelling, 33: 669-682.

Mallat, S.G. 1999. A wavelet Tour of signal processing, 2nd ed., Academic Press, London.

Merz, C.J. 1998. Classification and Regression by Combining Models. Ph.D. Thesis, Dep. of Inf. and Comput. Sci., Univ. of Calif, Irvine.

Mohanty, S., Jha, M.K., Kumar, A. and Sudheer, K.P. 2010. Artificial neural network modeling for groundwater level forecasting in a River Island of Eastern India. Water Resour. Manag., 24(9): 1845-1865.

Moosavi, V., Vafakhah, M., Shirmohammadi, B. and Behnia, N. 2013. A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. Water Resour. Manag., 27(5): 1301-1321.

Murtagh, F., J.L. Starck, and O. Renaud. 2004. On neuro-wavelet modeling. Decis. Support Syst., 37: 475-484.

Opitz, D. and Maclin, R. 1999. Popular ensemble methods: an empirical study. J. Artif. Intell. Res., 11: 169–198.

Partal, T. and Cigizoglu, K. 2008. Estimation and forecasting of daily suspended sediment data using wavelet-neural networks. J. Hydrol., 358(3-4), 317-331.

Ramana, R.V., Krishna, B., Kumar, S.R. and Pandey, N.G. 2013. Monthly rainfall prediction using wavelet neural network analysis. Water Resour. Manag., 27: 3697-3711.

Rumelhart, D.E. and McClelland, J.L. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Foundations, vol. 1. MIT Press, Cambridge, Mass

Shu, C. and Ouarda, T.B.M.J. 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. Water Resour. Res., 43, W07438. doi:10.1029/2006WR005142.

Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O. and Lee, K.K. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J. Hydrol., 396(1): 128-138.

Zaier, I., Shu, C., Ouarda, T.B.M.J., Seidou, O. and Chebana, F. 2010. Estimation of ice thickness on lakes using artificial neural networks ensembles. J. Hydrol., 380: 330–340.

Zheng, G., J.L. Starck, J.G. Campbell, and F. Murtagh. 1999. Multiscale transforms for filtering financial data streams. Journal of Computational Intelligence in Finance, 7: 18-35.