Regional low-flow frequency analysis using single and ensemble artificial neural networks

T. B. M. J. Ouarda[1] and C. Shu[1]

[1]   In this paper, artificial neural networks (ANNs) are introduced to obtain improved regional low-flow estimates at ungauged sites. A multilayer perceptron (MLP) network is used to identify the functional relationship between low-flow quantiles and the physiographic variables. Each ANN is trained using the Levenberg-Marquardt algorithm. To improve the generalization ability of a single ANN, several ANNs trained for the same task are used as an ensemble. The bootstrap aggregation (or bagging) approach is used to generate individual networks in the ensemble. The stacked generalization (or stacking) technique is adopted to combine the member networks of an ANN ensemble. The proposed approaches are applied to selected catchments in the province of Quebec, Canada, to obtain estimates for several representative low-flow quantiles of summer and winter seasons. The jackknife validation procedure is used to evaluate the performance of the proposed models. The ANN-based approaches are compared with the traditional parametric regression models. The results indicate that both the single and ensemble ANN models provide superior estimates than the traditional regression models. The ANN ensemble approaches provide better generalization ability than the single ANN models.

1.   Introduction and Review

[2]   A large number of engineering activities require the availability of reliable low-flow frequency estimates. Such activities include fish habitat analysis, water quantity and quality management, and environmental impact assessment. For a site with a sufficient amount of historical streamflow records, two traditional approaches can be used to obtain low-flow quantile estimates. The first approach is based on the standard procedure for hydrological frequency analysis, which involves fitting a probabilistic distribution to the observed low-flow data [see, e.g., *Tasker*, 1987]. In the second approach, flow duration curves (FDC) depicting the frequency at which a given flow is equaled or exceeded, are constructed from the available flow record [see, e.g., *Smakhtin et al.*, 1997].

[3]   For ungauged sites, where no historical streamflow record is available, low-flow statistics are frequently estimated using regional regression techniques [*Vogel and Kroll*, 1990, 1992; *Dingman and Lawlor*, 1995; *Eng and Milly*, 2007]. The most used regional regression model has the following generalized form [*Thomas and Benson*, 1970]:

$$Q_{d,T} = a x_1^{\theta_1} x_2^{\theta_2} \cdots x_i^{\theta_i} \cdots x_n^{\theta_n} \qquad (1)$$

where $Q_{d,T}$ is the d-day, T-year low-flow quantile at the site of interest (d, duration; T, return period); x_i is the ith

physiographic or climatic characteristic used for quantile estimation; θ_i is the ith model parameter; n is the total number of site characteristics used in the model; and a is the multiplicative error term. Solving equation (1) using linear regression techniques generally requires linearizing the power form model by a logarithmic transformation to the form

$$\log(Q_T) = \log(a) + \theta_1 \log(x_1) + \theta_2 \log(x_2) + \theta_3 \log(x_3) + \ldots \\ + \theta_n \log(x_n) \qquad (2)$$

Parameters of equation (2) can be estimated using the ordinary, weighed or generalized least squares techniques.

[4]   Other approaches have been attempted for the prediction of low-flow characteristics at ungauged basins. Such approaches include the drainage area ratio method based on the establishment of a linear drainage area discharge relationship, spatial interpolation and regional mapping techniques which assume the existence of a clear relationship between the flow field and explanatory physiographic variables, and the use of regional prediction curves which can be established by standardizing flows at gauged sites by a scale index and then combining the information into a single regional curve. It is also possible to synthetically generate a large number of streamflow time series based on records available at gauging stations and then proceed with the estimation of the regional low-flow characteristics. For a comprehensive review of methods of low-flow estimation at ungauged sites the reader is referred to *Smakhtin* [2001] and *Ouarda et al.* [2008].

[1]INRS-ETE, University of Quebec, Quebec, Quebec, Canada.

[5] Artificial neural networks (ANNs) are biologically inspired computing and modeling tools. One of the major benefits of ANNs is their capability to approximate arbitrary functions given sufficient parameters and training samples. ANNs have been successfully applied to solve a wide range of hydrological and water resources problems. Examples of such applications include rainfall-runoff modeling [*Minns and Hall*, 1996; *Gupta et al.*, 1997], flood routing modeling [*Peters et al.*, 2006], streamflow prediction [*Moradkhani et al.*, 2004], ice dynamics modeling [*Seidou et al.*, 2006] and precipitation estimation [*Hsu et al.*, 1997; *Kuligowski and Barros*, 1998]. *The ASCE Task Committee on Artificial Neural Networks in Hydrology* [2000a, 2000b] and *Govindaraju and Rao* [2000] provided a comprehensive review of the ANN application in hydrology.

[6] Recently, the use of ANNs for hydrological regionalization has been attracting increasing attention. As an alternative to regressive methods, single ANN and ANN ensemble models were introduced by *Shu and Burn* [2004] for regional flood estimation at ungauged sites. The application to selected catchments in the United Kingdom (UK) indicates that the nonlinearity introduced by the ANN models allows them to outperform parametric regression methods.

[7] The generalization ability of a single ANN can be improved by using a properly designed ANN ensemble. *Dawson et al.* [2006] applied ANNs to index flood and flood quantile estimation for 870 catchments across the UK. The results obtained from the ANNs are comparable in accuracy with those obtained by the Flood Estimation Handbook (FEH) [*Reed and Robson*, 1999] models. *Shu and Ouarda* [2008] applied the adaptive neurofuzzy inference system (ANFIS) and ANN approaches for regional flood frequency analysis, and the results indicated that both approaches show a better generalization ability than the parametric regression methods. *Shu and Ouarda* [2007] also developed a canonical correlation analysis (CCA) based single ANN and ANN ensemble models for improved regional flood estimation at ungauged sites.

[8] Despite the increasing popularity of ANN-based methodologies, they have never been applied to the regional frequency prediction of low-flow characteristics. A relatively limited number of publications dealt with drought forecasting and drought risk assessment based on ANNs [*Crespo and Mora*, 1993; *Incerti et al.*, 2007; *Mishra et al.*, 2007; *Morid et al.*, 2007; *Ochoa-Rivera*, 2008]. It is the intent of this paper to apply ANN modeling techniques to regional low-flow estimation at ungauged sites. Specifics regarding the design and implementation of ANNs are discussed. Two ensemble techniques, bagging and stacking, are introduced to improve ANN generalization ability. The ANN models are also compared with the traditional regression models for low-flow estimation based on data from the province of Quebec, Canada.

[9] The remainder of this paper is organized as follows. In section 2, a general introduction to the single and ensemble ANNs for regional low-flow analysis is provided. In section 3, a description of the study area is provided. In section 4, the details concerning the configuration of the single and ensemble ANNs, the estimation models to be compared, and the evaluation methodology are presented. In section 5, the results obtained by applying the proposed approaches are presented and discussed. Finally, in section 6, the conclusions of this work and recommendations for further research are presented.

2. General Background on the Use of Single and Ensemble ANN Models for Regional Low-Flow Analysis

2.1. Single ANN Model

[10] Among the various types of ANNs that have been developed over the years, multilayer perceptrons (MLPs) (also known as multilayer feed-forward networks) originally proposed by *Rumelhart and McClelland* [1986] are the most commonly used and well researched class of ANNs. A typical MLP has an input layer, at least one hidden layer, and an output layer. The layers are interconnected through weighted links from lower layer to higher layer, without lateral or feedback connections. The input layer receives values of the input variables for a given problem. There could be one or more hidden layers lying between the input and output layer. The output layer provides the ANN prediction and represents model output. Transfer functions used in the neurons of the hidden and output layers, which introduce nonlinearity to the network, play an important role in determining the behavior of an ANN.

[11] For the problem of low-flow quantile estimation at ungauged sites, an ANN model is used to approximate the functional relationship between the physiographical variables and the hydrologic variables which act as the input and output, respectively, of an ANN. Parameters in an ANN, which are called weights in the ANN literature, are determined through supervised training. A Bias unit helps convergence of the weights to an acceptable solution, and it can be thought of as a unit which is connected to the hidden and output layer and has an constant output of 1. An extra degree of freedom in the weight space is introduced by adding a bias unit. Once a network is trained and tested it can be given new input information to predict the output. During the training process, network parameters must be optimized until the prediction error made by the network is minimized and the network reaches the specified level of accuracy. The error of a particular configuration of the network can be computed by comparing the ANN generated results with the desired outputs. The differences are combined together by an error function to give the network error. A MLP can be trained in either an incremental or batch style. In incremental training, network parameters are updated each time a training case is presented to the network. In batch training, network parameters are not updated until all the training cases are presented. Although it is arguable which training style results into better generalization ability, the sequential mode generally has a higher speed of learning [*Haykin*, 1994].

[12] There are also other training algorithms collectively known as second-order training algorithms that can significantly increase training speed. These methods include the conjugate gradient algorithm, the quasi-Newton algorithm and the Levenberg-Marquardt (LM) algorithm [*Bishop*, 1995]. Among these algorithms, the LM algorithm [*Hagan and Menhaj*, 1994] is considered to be the most efficient for training median sized artificial neural networks and works extremely well in practice [*Burney et al.*, 2004]. The LM

algorithm is a variation of the Gauss-Newton algorithm, and it is developed to approach second-order training speed without having to compute the Hessian matrix. The new configuration of weight w in step n using LM algorithm is computed as

$$w(n) = w(n-1) - \left(J^T J + \mu I\right)^{-1} J^T e(n) \tag{3}$$

where the Jacobian matrix J contains the first derivatives of the network errors with respect to weights and biases, I is the identity matrix, and e is a vector of network errors which is the difference between the actual and the desired value of the network output. Learning parameter μ is modified based on the development of the error function E. More specifically, μ is multiplied by the decay rate β ($0 < \beta < 1$) if E decreases in a new step, whereas μ is divided by β if E increases in a new step. The LM algorithm is selected in this paper to train the ANNs, and the parameters of the algorithm are discussed in section 4.2.

2.2. ANN Ensemble Model

[13] An ANN ensemble consists of a set of ANNs which are trained for the same task. Theoretical work by *Hansen and Salamon* [1990] suggested that the generalization ability of a single ANN can be significantly improved using an ensemble of ANNs with a plurality consensus scheme for a classification problem in which the final classification results are determined by the majority of the networks. *Krogh and Vedelsby* [1995] proved that the generalization ability of an ensemble is largely controlled by two factors: the average generalization ability and the average ambiguity of the ensemble members. There are also several practical works, such as *Drucker* [1997], *Sharkey* [1999], *Dietterich* [2000] and *Carney and Cunningham* [1999], which showed that the performance of a single ANN can be improved by using appropriately selected ensemble techniques. Generally two major steps are involved in creating an ANN ensemble. The first step is to generate individual ensemble members, and the second step is to combine the prediction from the ensemble members to produce a unique output. *Opitz and Maclin* [1999] provided a review of the popular ensemble methods. *Shu and Burn* [2004] conducted a comprehensive evaluation of six ANN ensemble techniques for hydrological application. *Shu and Ouarda* [2007] applied bagging and averaging ensemble techniques to improve ANN performance for regional flood estimation at ungauged sites.

[14] Two widely used methods for creating ensemble members are bagging [*Breiman*, 1996a] and boosting [*Freund and Schapire*, 1996; *Schapire*, 1990]. Both of these methods rely on resampling algorithms to obtain different training sets for the component predictors. Bagging [*Breiman*, 1996a] is an acronym for "bootstrap aggregation," and it is based on bootstrap statistical resampling technique [*Efron and Tibshirani*, 1993]. In bagging, each ANN is trained independently on a training set sampled with replacement from the original data set. Suppose there are m instances in the training set T. The training set of a member network, T_B, is generated by sampling with replacement m times from the original training set T. Each instance of the original data set has a probability of $1/m$ to be selected. There could be a number of instances that are repeated several times in T_B,

while others may be left out. The probability that an instance from T will not be part of a bootstrap resampled training set is $(1 - 1/m)^m$, which is close to 0.37 for a large sample.

[15] The boosting algorithm was originally proposed by *Schapire* [1990]. The algorithm generates a series of ensemble members which are trained with different distributions of the original training data. The algorithm starts by training the first predictor with the original training set. Then the training set of a new predictor is sampled from the original data set based on a performance adjusted distribution. The distribution ensures that high probabilities of being sampled are assigned to the training cases for which the predicted values obtained from the previous predictor differ significantly from their desired values. Thus training cases which failed to be correctly predicted by the previous predictor will have a greater chance of appearing in the new training set than those correctly predicted. Thus, the sequentially generated predictors by the boosting algorithm are specialized in different parts of the observation space.

[16] A number of authors, such as *Carney and Cunningham* [1999] and *Zhang* [1999], have shown that bagging is an effective approach to improve model generalization ability for both regression and classification problems. *Drucker* [1997, 1999] and *Sharkey* [1999] showed that boosting could generate better results than bagging in many situations. However, boosting is very sensitive to the outliers [*Bauer and Kohavi*, 1999]. Due to this major drawback of the boosting algorithm, the bagging algorithm is used in the present paper to generate the individual ensemble networks.

[17] Simple averaging is the most used method for combining members of an ANN ensemble [*Wolpert*, 1992]. Stacked generalization is a more complicated approach that has generated significant interest among researchers during the recent years. Suppose that K member networks are generated, and the ith training case has a desired value y_i and obtains a predicted value $\hat{y}_i^k$ from the kth network. Combining the networks using simple averaging is defined as

$$\hat{y}_i = \frac{1}{K} \sum_{k=1}^{K} \hat{y}_i^k, \qquad i = 1 \ldots m \tag{4}$$

This method of combination has the benefits of easy implementation and improved performance [*Perrone and Cooper*, 1993; *Bishop*, 1995].

[18] Stacked generalization or stacking [*Wolpert*, 1992] improves the generalization ability of a classification or regression model by using a two-layer architecture. In stacking, the way that the outputs from the predictors or classifiers at level 0 are combined needs not to be linear but through a level 1 generalizer which is also trained to minimize an error function. Suppose there are K predictors (in our case the ANNs) at level 0, and the data set used for training the level 1 generalizer has a size of n. The purpose of level 1 training is to find the coefficients $\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_K$ to minimize the following function suggested by *Breiman* [1996b] for a regression problem:

$$W = \sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{K} c_k \hat{y}_i^k \right]^2 \qquad c_k > 0 \tag{5}$$
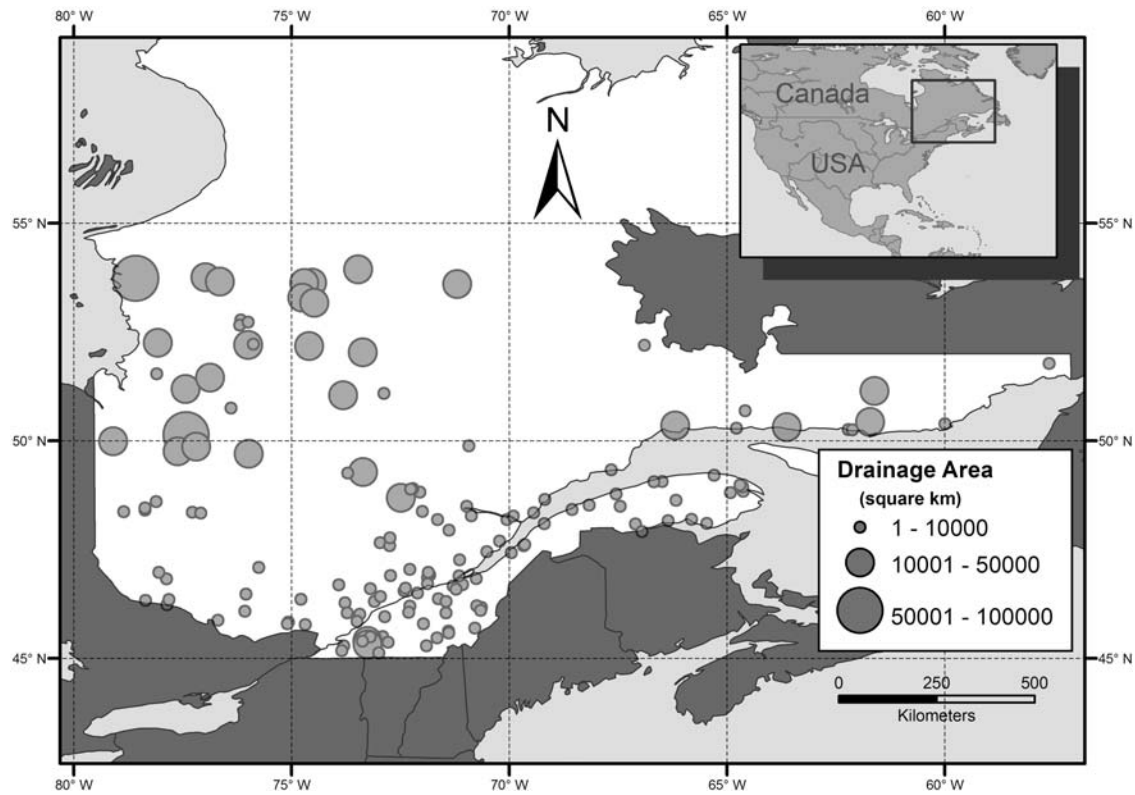
Figure 1. Hydrometric stations across the province of Quebec, Canada.

where pattern i has a desired output y_i and the prediction from the kth level 0 predictor is $\hat{y}_i^k$. After the coefficients $\hat{c}_1$, $\hat{c}_2, \ldots, \hat{c}_K$ are identified, the ensemble prediction for the jth pattern can be constructed by

$$\hat{y}_j = \sum_{k=1}^{K} c_k \hat{y}_j^k \qquad (6)$$

Equation (6) minimizes the squared absolute differences between the observed and predicted values. However, this process, when used to determine the coefficients, may be dominated by those patterns with a large error. A better option was suggested by *Shu and Burn* [2004] and consists in minimizing the squared relative difference:

$$V = \sum_{i=1}^{m} \left[\frac{y_i - \sum_{k=1}^{K} c_k \hat{y}_i^k}{y_i} \right]^2 \qquad c_k > 0 \qquad (7)$$

This option was adopted in the present paper. *Drucker* [1997], *Hu and Tsoukalas* [2003] and *Shu and Burn* [2004] compared the two approaches for combining ensemble members. The results of these two studies indicated that stacked generalization provides better estimation than averaging.

3. Study Area

[19] The ANN approaches proposed in this paper are applied to the hydrometric station network of the southern part of the province of Quebec, Canada. The same network was used in a regional low-flow frequency study by *Ouarda et al.* [2005]. The winter and summer mean temperatures of the study area are in the ranges of $[-21°C, -10°C]$ and $[12°C, 20°C]$, respectively. Due to the strong seasonal variation in the flow regime of the region of study, it is more appropriate to study winter and summer low flows separately. In the present paper, low-flow quantiles $Q_{T,d}$ corresponding to return periods of $T = 2, 5$ and 10 years and durations of $d = 7$ and 30 days are examined for the winter and for the summer seasons. These quantiles are commonly used in Canada for the purpose of water quality control and fish habitat protection. To ensure the quality of the low-flow study, catchments selected from the 190 hydrometric stations managed by the Ministry of the Environment of the province of Quebec (MENV) should meet the following criteria [*Ouarda et al.*, 2005].

[20] 1. A historical flow record of at least 10 years is required.

[21] 2. The gauged catchment should present a natural flow regime.

[22] 3. The historical data at the gauging stations should pass the Kendall test of stationarity [*Kendall*, 1975] and the nonparametric independence test by *Wald and Wolfowitz* [1943].

[23] As a result, 134 and 129 sites were selected for the analysis of $Q_{T,30}$ and $Q_{T,7}$, respectively, during the summer season. During the winter season, 135 and 133 sites were selected for the analysis of $Q_{T,30}$ and $Q_{T,7}$, respectively. The selected catchments are located between 45°N and 55°N, and between 55°W and 80°W. The area of these catchments ranges from 572 km^2 to 96,600 km^2 with a median value of 3077 km^2. The locations of the selected hydrometric stations are shown in Figure 1.

Table 1. Descriptive Statistics of the Selected Physiographical and Meteorological Variables

Variable	Units	Notation	Mean	Maximum	Minimum	Standard Deviation
Basin area	km^2	A	5655.52	96600	0.70	11685.7
Percent of the basin occupied by lakes	%	PLAKE	6.33	32.00	0.00	6.57
Percent of the basin occupied by forest	%	PFOR	85.78	100.00	6.50	15.97
Annual mean degree days <0°C	degree day	DJBZ	1635.15	2963.10	920.60	529.29
Summer mean liquid precipitation	mm	PLME	464.51	664.00	306.00	77.40
Average number of days with temperature >27°C	-	NJH27	12.28	36.60	0.80	7.57
Curve number	-	CN	45.08	78.20	21.00	-

[24] A set of physiographical and meteorological variables that are most suitable to explain the low-flow processes are used in this study. These variables are the basin area (A), the percentage of the basin area occupied by lakes (PLAKE), the percentage of the basin area covered by forest (PFOR), mean annual degree days below 0°C (DJBZ), average summer/autumn liquid precipitation (PLME), average number of days for which the mean temperature exceeds 27°C (NJH27) and a soil characteristic: the curve number (CN). The value of 27°C represents a specific regional hydrology and climatology threshold for Quebec indicating the average of July maximum temperatures. The summary statistics and a brief description for these variables are presented in Table 1. *Ouarda et al.* [2005] provide more details concerning the description of these variables and the methodology used to build the physiographical and meteorological database.

[25] At-site frequency analysis was carried out for low-flow data at each station. Local low-flow quantiles $Q_{T,d}$ corresponding to the various return periods T and durations d (days) are estimated. All following commonly used statistical distributions were considered in the fitting process: Gumbel (EV1), Weibull (W2), two-parameter lognormal (LN2), three-parameter lognormal (LN3), generalized extreme value distribution (GEV), Gamma (G), Pearson type III (P3), log-Pearson type III (LP3) and generalized Pareto (GP). At each station, the Bayesian information criterion is used to identify the distribution that best fits the data. The appropriate distribution is then used for the local estimation of low-flow quantiles. For summer low flows, the distributions LN2, G, W2 and EV1 were selected 79, 40, 35 and 22 times, respectively, as appropriate distributions and provided excellent fit to the data. For winter low flows, the same distributions were selected 39, 45, 39 and 50 times, respectively, as appropriate distributions. The other distributions accounted for a very small number of stations. Correlations between the various independent variables and local low-flow quantiles are computed and presented in Table 2.

4. Study Methodology

[26] In the present work, the parametric multiple regression method used by *Ouarda et al.* [2005] is compared to the ANN based approaches.

4.1. Multiple Regression Equations

[27] Multiple regression equations are developed for the entire region of study for the estimation of summer and winter low-flow quantiles. Five variables, A, PLAKE, PLME, NJH27 and CN are considered as explanatory physiometeorological variables for the estimation of sum-

mer low-flow quantiles, while four variables including A, PLAKE, PFOR and DJBZ are considered for the estimation of winter low-flow quantiles [*Ouarda et al.*, 2005]. The optimal variables were selected based on stepwise regression analysis. Appropriate transformation is required for a number of hydrological, physiographical and meteorological variables in the regression analysis in order to achieve normality or linearity. For A, DJBZ and for low-flow quantiles, a logarithmic transformation is used. For PLAKE, a square root transformation is used. As a result, the following regression equation is used for the estimation of summer low-flow quantiles [*Ouarda et al.*, 2005]:

$$\log(Q_{T,d}) = \log(\beta_0) + \beta_A \log(A) + \beta_{NJH27} NJH27$$
$$+ \beta_{PLAKE}\sqrt{PLAKE} + \beta_{CN}CN + \beta_{PLME}PLME + \varepsilon \quad (8)$$

For the estimation of the winter low-flow quantiles, the regression equation is given by

$$\log(Q_{T,d}) = \log(\beta_0) + \beta_A \log(A) + \beta_{PLAKE}\sqrt{PLAKE}$$
$$+ \beta_{PFOR}PFOR + \beta_{DJBZ}\log(DJBZ) + \varepsilon \quad (9)$$

In equations (8) and (9), parameters β are the regression coefficients and ε are the residuals. More detailed information regarding the development of the two regression equations is presented by *Ouarda et al.* [2005].

4.2. Single and Ensemble ANN Models Adopted in This Study

[28] MLP is selected as the single ANN model (SANN) and the base model of an ANN ensemble to establish the relationship between low-flow quantiles and catchment descriptors. The actual MLP adopted in this paper consists of an input layer, one hidden layer, and an output layer. Proper preprocessing is generally required for both inputs

Table 2. Correlation Between Quantiles and Physiographic Variables

	Summer			Winter		
	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$
A	0.941	0.944	0.927	0.981	0.983	0.975
DJBZ	0.575	0.572	0.566	0.558	0.559	0.556
PLAKE	0.531	0.541	0.530	0.588	0.585	0.583
PFOR	−0.029	−0.031	−0.031	−0.074	−0.066	−0.067
NJH27	−0.344	−0.341	−0.343	−0.308	−0.301	−0.298
PLME	−0.432	−0.429	−0.426	−0.429	−0.428	−0.425
CN	−0.203	−0.214	−0.212	−0.173	−0.183	−0.181

and outputs of an ANN. As a result, inputs are normalized so that they have means of zero and standard deviations of one, while a logarithmic transformation is used for the output layer. *Shu and Burn* [2004] and *Shu and Ouarda* [2007] pointed out that if a linear transformation is used for the output, an ANN may concentrate the effort on learning patterns having large values. By taking logarithms, these problems can be mitigated, since a difference between two logarithmic transformed values measures the ratio of the original values. The added benefit of using logarithmic transformation for low-flow estimation is that quantiles estimated by an ANN cannot fall into unrealistic negative flow. The transfer functions for the hidden nodes and the output nodes use the tan-sigmoid function and the linear function, respectively. The same set of independent variables used for regression analysis is used as the inputs for the ANN models. Thus five input nodes are used in the ANNs for summer low-flow estimation, and four input nodes are used in the ANNs for winter low-flow estimation. Each ANN has one output which is the low-flow quantile to be estimated.

[29] The ANN models developed in this paper are trained using the LM algorithm. The parameters are updated in the batch mode. The scalar parameter μ in equation (3) is adjusted during the network training according to the system performance. The LM algorithm behaves as a gradient descent method with a small step size when the value of μ is large. However, when the value of μ is close to zero, the algorithm approximates the Gauss-Newton method [*Demuth and Beale*, 2003]. The initial value for μ is given as 0.005 in this paper. If a training epoch decreases the performance of the network, the value of μ is multiplied by $\mu_d = 0.1$. If a training epoch increases the performance of the network, the value of μ is multiplied by $\mu_i = 10\$$. The ANNs stop training when μ reaches a preset parameter $\mu_{\max} = 1 \times 10^6$.

[30] Using proper combination of nonlinear transfer functions in the hidden layers enables an ANN to approximate successfully any complex nonlinear relationships. However, when too many hidden neurons are used in the ANN, the ANN can be easily overtrained due to excessive model degrees of freedom. To overcome the problem of overtraining in ANN modeling, the early stopping and weight decay approaches can be used [*Matignon*, 2007]. The use of early stopping requires dividing the training data into two parts. One part is used for network training, and the other part is used for network validation. The basic idea of early stopping is to stop ANN training when the validation error reaches a desirable minimum. In the weight decay regularization algorithm, the error function, which is minimized during the training phase, is augmented with additional terms that penalize the complexity of the model. The new error function *msereg* with penalty has the following form:

$$msereg = \gamma\, mse + (1 - \gamma)\, msw \qquad (10)$$

where γ is the performance ratio, *mse* is the typical performance function (mean square error) used for ANN training, and *msw* is the mean of the sum of squares of the network weights. Using a simple early stopping strategy such as randomly selecting the validation set is the widely adopted approach for preventing overfitting. However, *Shu and Burn* [2004] indicated that there are two major

problems with early stopping. First, a validation set needs to be extracted from the training set, which may lead to insufficient data being available to successfully train an ANN. Second, how to optimally separate the validation set still remains a major challenge. Thus, the weight decay regularization method is adopted in this paper.

[31] The ANN ensemble (EANN) model is proposed in this paper to improve the generalization ability of the single ANN models. The member networks of the EANN are generated using the bagging algorithm, while the stacked generalization method is used to combine these individual networks to generate a unique output. A total number of ten ANNs are used in each ANN ensemble. Member networks of the ensemble ANN model use the same configuration as the single ANN model.

4.3. Evaluation Criteria

[32] A jackknife evaluation procedure [*Miller*, 1964; *Shao and Tu*, 1995] is used in this paper to compute the goodness of fit statistics and the evaluation criteria necessary for the assessment of the performance of the various regional low-flow frequency models. The advantage of the jackknife procedure is that model accuracy obtained using the procedure is independent of the calibration data [*McCuen*, 2005]. In jackknifing, the flow record of one catchment in the study area is held out from the database, thus the catchment is considered as "ungauged." Then the coefficients in the regional regression models or weights in the ANN models are calibrated or trained using the data from the remaining sites to obtain the estimates of the catchment that is held out. This process is continued until regional estimates are obtained using the proposed models for all the sites in the study area.

[33] A set of five indices is used to evaluate the regional low-flow frequency analysis models proposed in this paper. These indices are the Nash criterion (*NASH*), the root mean squared error (*RMSE*), the relative root mean squared error (*RMSEr*), the mean bias (*BIAS*), and the relative mean bias (*BIASr*). They are computed using the following equations:

$$NASH = 1 - \frac{\sum_{i=1}^{n} (q_i - \hat{q}_i)^2}{\sum_{i=1}^{n} (q_i - \overline{q})^2} \qquad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (q_i - \hat{q}_i)^2} \qquad (12)$$

$$RMSEr = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{q_i - \hat{q}_i}{q_i} \right)^2} \qquad (13)$$

$$BIAS = \frac{1}{n} \sum_{i=1}^{n} (q_i - \hat{q}_i) \qquad (14)$$

$$BIASr = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{q_i - \hat{q}_i}{q_i} \right) \qquad (15)$$
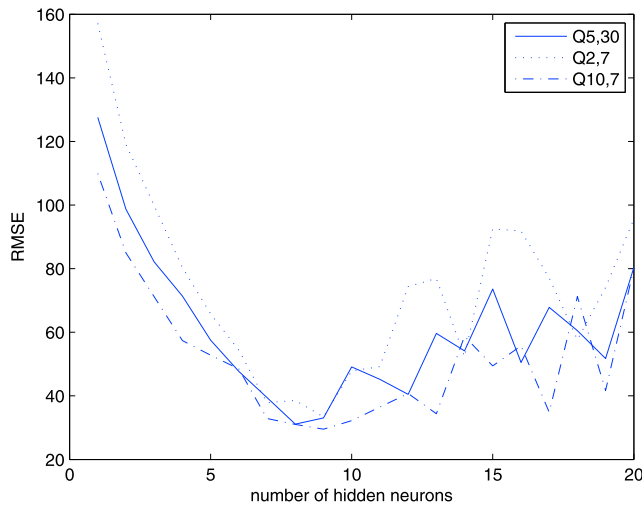
Figure 2. RMSE of summer low-flow estimation using single ANN models with the number of hidden neurons increased from 1 to 20.

where n is the total number of stations being modeled, q_i is the at-site estimate for site i, $\hat{q}_i$ is the estimate obtained from the regional low-flow model for site i, and $\overline{q}$ is the mean of at-site estimation of the n stations.

5. Results and Discussion

[34] ANN structures need to be optimally designed so that the resulting ANN models can obtain desirable generalization ability. One of the most challenging but important issues when designing a MLP is the determination of the number of neurons in the hidden layer. To illustrate this issue, the performance of a single ANN model is evaluated by sequentially increasing the number of hidden neurons from one to a relatively large number, 20 (Figures 2 and 3). The performance measure, RMSE, is obtained from a tenfold cross-validation procedure for each ANN.

[35] From Figure 2, the ANNs achieve lowest RMSE when the number of hidden neurons of the ANNs for summer low-flow quantiles $Q_{5,30}$, $Q_{2,7}$ and $Q_{10,7}$ estimation is equal to 8, 9 and 9, respectively. If less than 6 neurons are used in the hidden layer, the ANNs do not have sufficient complexity to fully represent the functional relationship between the system inputs and outputs. This results in the problem of underfitting. However, if the number of hidden neurons increases above 10 neurons, the ANN model may suffer from the problem of overfitting. This is caused by the large number of hidden neurons which leads to not having enough training cases to adequately train all the neurons in the ANN.

[36] From Figure 3, we can observe that the ANN models for winter low-flow quantile estimation achieve lowest RMSE when the numbers of hidden neurons for Q5,30, Q2,7 and Q10,7 estimation increase to 11, 13, and 14, respectively. The ANNs tend to underfit if less than 8 neurons are used in the hidden layer. However, if the number of hidden neurons increases over 15, the ANN models may lead to the problem of overfitting. Overall, the performance of the ANN models for both summer and winter low-flow quantile estimation improves dramatically with the initial increase of the number of hidden neurons. For winter low-flow quantile estimation, ANNs show more resistance to overfitting compared to summer low-flow quantile estimation.

[37] The SANN, EANN and the parametric regression models are applied to the study area for summer and winter low-flow quantile estimation. The results obtained using the jackknife validation procedure are presented in Tables 3 and 4, respectively.

[38] The *NASH* criterion is a widely used goodness of fit measure for hydrological models. The criterion indicates a perfect fit if it is equal to 1. Normally the model fit can be considered as good if the *NASH* criterion is above 0.8. All the models evaluated in the present paper, for both summer and winter low-flow quantile estimation, have a *NASH* criterion higher than 0.89. This indicates that all these approaches provide a very satisfactory fit. However, both SANN and EANN methods show better performances in the *NASH* criterion than the parametric regression method for the estimation of all low-flow quantiles evaluated in this paper.

[39] For the estimation of summer low-flow quantiles, both SANN and EANN approaches show a better performance in the *RMSE* and *RMSEr* indices than the regression approach. This indicates that the ANN based approaches provide better prediction accuracy in both absolute and relative scale. For the winter low-flow quantile estimation, both the SANN and EANN models show better performance than the regression model in the *RMSE* criterion. However, the regression model shows better performance than the SANN model in the *RMSEr* criterion. This indicates that the SANN model shows a better performance in the absolute error and a lower performance in the relative error than the regression model for winter low-flow quantile estimation. The EANN approaches have the best performance in both *RMSE* and *RMSEr* indices for winter low-flow quantile estimation. This can be explained by the two level structure of the EANN model. The level 0 training of the EANN model uses mean square error with weight penalty as the performance function which can essentially improve the model performance in the *RMSE* criterion,
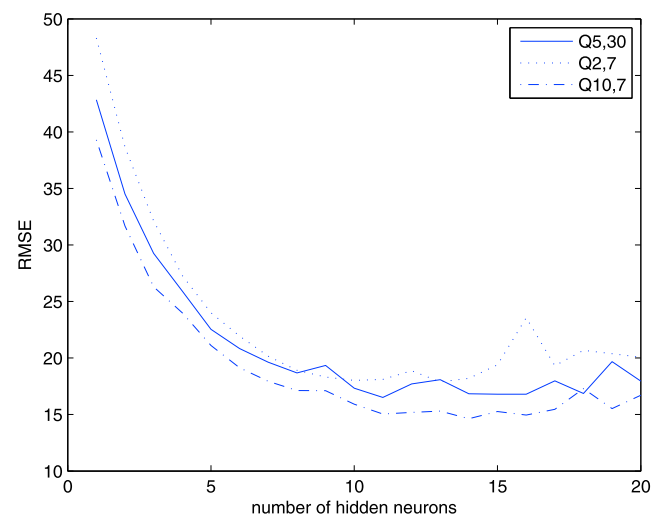

Figure 3. RMSE of winter low-flow estimation using single ANN models with the number of hidden neurons increased from 1 to 20.

Table 3. Jackknife Validation Results for Summer Low-Flow Quantile Estimation[a]

	Single ANN Model			Ensemble ANN Model			Regression Model		
	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$
NASH	**0.97**	0.96	0.95	**0.97**	**0.97**	**0.96**	0.93	0.94	0.92
RMSE	31.12	38.57	31.02	**27.95**	**35.90**	**27.33**	45.66	51.61	41.63
RMSEr(%)	39.77	40.11	42.69	**31.02**	**31.41**	**36.17**	54.37	52.15	68.10
BIAS	3.84	5.49	4.80	0.94	5.47	2.69	7.87	9.56	8.31
BIASr(%)	−6.54	−4.66	−7.05	**−3.08**	**−1.65**	**−3.17**	−10.55	−9.90	−15.33

[a]Bold denotes the best performing approach.

while the level 1 training uses a relative performance function which can help the model performance in the *RMSEr* criterion.

[40] Based on the *BIAS* index, all the approaches tend to overestimate low-flow quantiles, and the EANN and SANN approaches represent the least biased models for summer and winter low-flow estimation, respectively. However, the analysis based on the *BIASr* index suggests that all models underestimate low-flow quantiles, and the EANN approach is the least biased model for both summer and winter low-flow estimation. The contradictory signal sent by the *BIAS* and *BIASr* indices is mainly caused by scale effects. Sites with larger quantile values have more influence on the *BIAS* index than the smaller sites, while the *BIASr* index treats each site in the study area equally.

[41] The regional estimates of the summer low-flow quantiles using the SANN, EANN, and regression approaches are shown in Figures 4, 5, and 6, respectively. The single and ensemble ANN models show a better overall performance according to these figures compared to the regression approach. For the sites with local estimates of the summer low-flow quantiles $Q_{5,30}$, $Q_{2,7}$ and $Q_{10,7}$ in the ranges of (0, 500), (0, 500) and (0, 400), respectively, the EANN model shows a significantly better performance than the SANN model. The ANN based approaches also provide less biased estimates at these sites compared to the regression method. At the five sites with quantiles $Q_{5,30}$, $Q_{2,7}$ and $Q_{10,7}$ in the ranges of (500, 1400), (600, 1600) and (400, 1100) respectively, EANN and SANN models show a comparable performance. There is one noticeable outlier (site 076601) which is identified by the regression method. Compared to other sites with a similar catchment area in the database, this site has an unusual low mean number of days (only 1.1) for which the temperature exceeds 27°C and a small curve number (27.0). Unlike the regression method, both single and ensemble ANN models lead to very good estimates at site 076601. There is another noticeable outlier (site 090601) which is identified by all the approaches.

Compared to other sites with a similar catchment area in the database, this site has a very high percentage of the basin that is covered by forest (94%), a very low percentage of the basin area that is covered by lakes (5%) and a small value of the curve number (26.7).

[42] The regional estimates of the winter low-flow quantiles using the SANN, EANN, and regression approaches are shown in Figures 7, 8, and 9, respectively. All the approaches have good performances at catchments with local estimates of the winter low-flow quantiles $Q_{5,30}$, $Q_{2,7}$ and $Q_{10,7}$ in the ranges of (0, 120), (0, 130) and (0, 100), respectively, although ANN based approaches show a better performance at these catchments. For the catchments with local estimates of the quantiles $Q_{5,30}$, $Q_{2,7}$ and $Q_{10,7}$ in the ranges of (120, 400), (130, 450) and (100, 350), respectively, the regression approach shows a slightly better performance than the ANN based approaches. There are two major outliers: sites 081007 and 081002. Site 081007 has a very large percentage of the area that is covered by lakes (22%), and the mean annual degree days below 0°C at the site is also very large (2351). Site 081002 shows also large values of these two characteristics, with the percentage of the area covered by lakes and the mean annual degree days below 0°C being 20% and 2320, respectively.

[43] The majority of the studies [e.g., *Dingman and Lawlor*, 1995; *Reed and Robson*, 1999; *Shu and Burn*, 2004; *Shu and Ouarda*, 2008; *Tasker*, 1987; *Vogel and Kroll*, 1990, 1992] for regional flood and low-flow estimation have been carried out in the original flow domain. However, there have been several studies, such as *Chokmani and Ouarda* [2004], which used specific (or normalized) flow quantiles. Specific quantiles for a given site are computed by dividing the site's quantiles with its drainage area; regression equations are then developed to model the relationship between the specific quantiles and the physiographical variables. Specific quantiles can be used to account for scale issues [*Chokmani and Ouarda*, 2004] in flow estimation and the high correlation between the

Table 4. Jackknife Validation Results for Winter Low-Flow Quantile Estimation[a]

	Single ANN Model			Ensemble ANN Model			Regression Model		
	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$
NASH	0.91	0.92	0.91	**0.92**	**0.93**	**0.92**	0.89	0.89	0.90
RMSE	16.51	18.11	15.04	**15.84**	**16.59**	**13.91**	18.18	20.30	16.38
RMSEr(%)	41.90	40.65	49.68	**34.87**	**33.13**	**42.92**	36.80	34.01	44.22
BIAS	**1.27**	**1.32**	**0.92**	1.61	1.66	1.10	2.13	2.32	1.75
BIASr(%)	−7.38	−6.43	−8.09	**−5.13**	**−4.55**	**−6.87**	−5.58	−5.00	−7.66

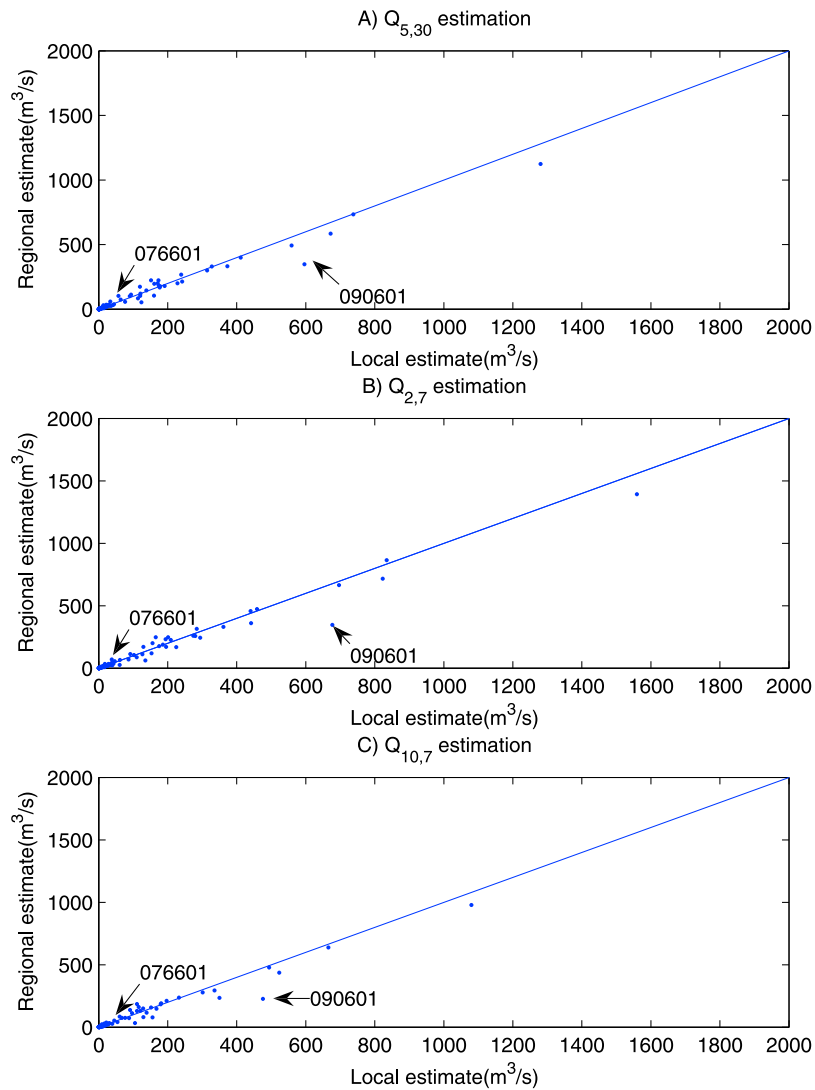[a]Bold denotes the best performing approach.

Figure 4. Jackknife estimation of the summer low-flow quantiles using the SANN approach.

drainage area and flow quantiles. To investigate the impact of using specific quantiles, single ANN models with specific quantiles as outputs and the same sets of input variables as described in section 4.2 are evaluated. For convenience, such ANN models are called specific ANNs. The cross-validation procedure described at the beginning of section 5 is used to determine the optimal numbers of hidden neurons in a specific ANN, and as a result, seven and ten hidden neurons are used in the specific ANNs for summer and winter-specific low-flow quantile estimation, respectively. The performance indices for the specific ANNs for summer and winter low-flow estimation are shown in Table 5. For summer low-flow quantile estimation, we can observe that specific ANN performs better than the regression model for all indices except BIASr. The single ANN performs better than the specific ANNs for all indices except BIAS. For winter low-flow estimation, specific ANNs shows a performance that is similar to the regression model for most performance indices. Single ANNs outperform specific ANNs for the NASH, RMSE and BIASr, while specific ANNs are superior in terms of RMSEr and BIAS. The results indicate that the use of specific quantile

shows no advantage over the estimation of quantiles in the original flow domain. Since ANN is used as a universal approximator in the regional estimation, and the improvement over the linear regression is in the nonlinear part of the relationship, normalizing the flow quantile should not materially change the essence of the modeling task. However, the assumption of a linear relationship between the quantile and drainage area (as implied in the normalization) certainly works against the ANN's nonlinear approximation capability.

6. Conclusions and Future Work

[44] A methodology for the use of ANN models for low-flow estimation at ungauged sites is presented in this paper. Two ensemble techniques, bagging and stacking, used for creating and combining ensemble members, respectively, are adapted in this paper to build the ANN ensemble models. The application of the ANN-based models and the traditional regression method in the study area for low-flow estimation for both the summer and the winter seasons shows that superior results can be achieved by
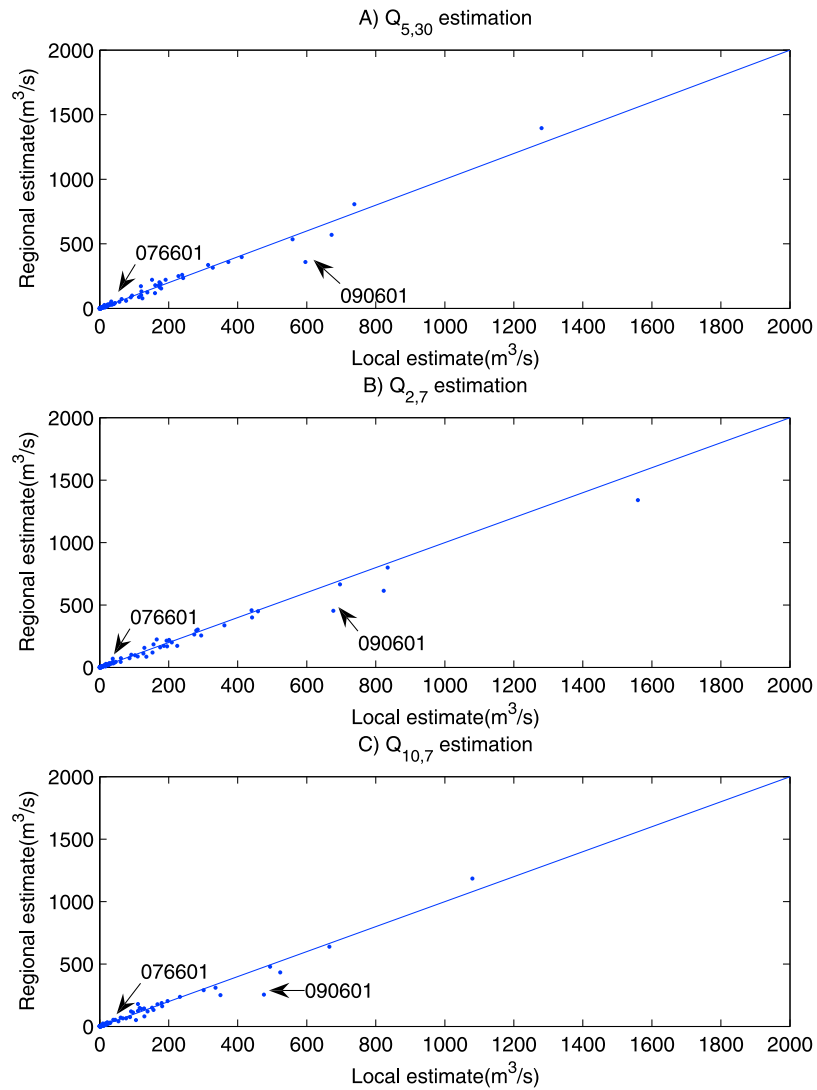
Figure 5. Jackknife estimation of the summer low-flow quantiles using the EANN approach.
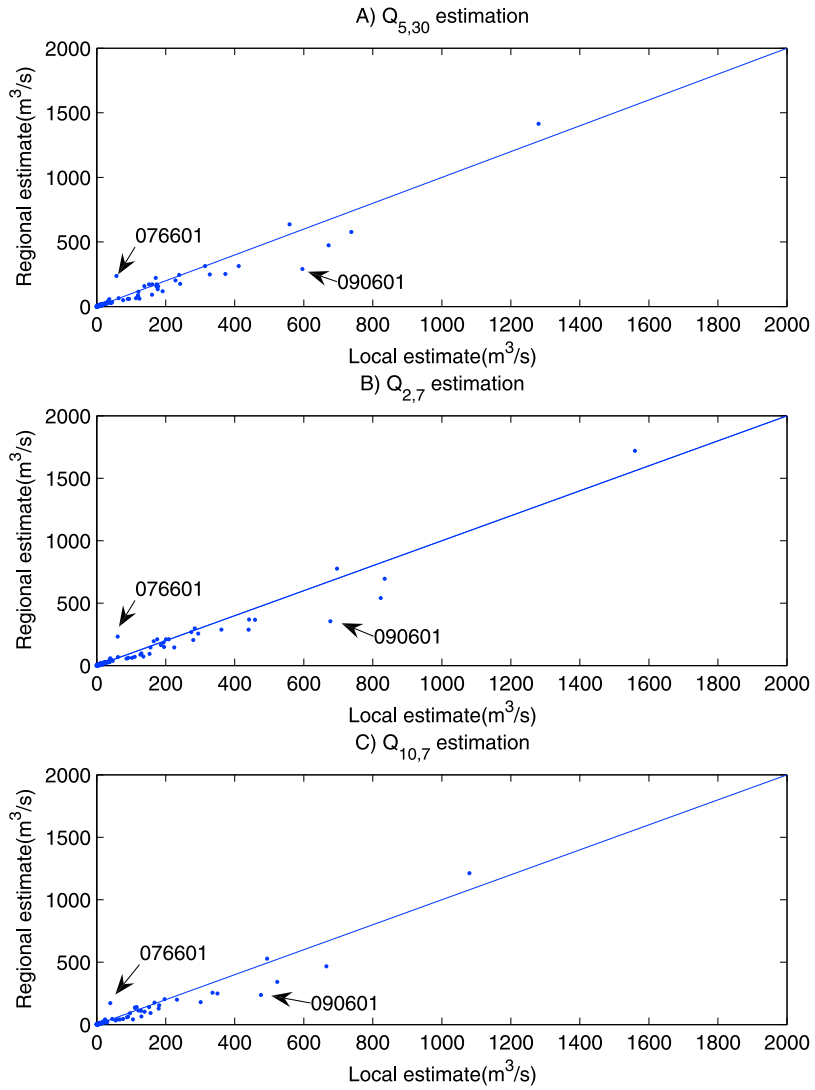
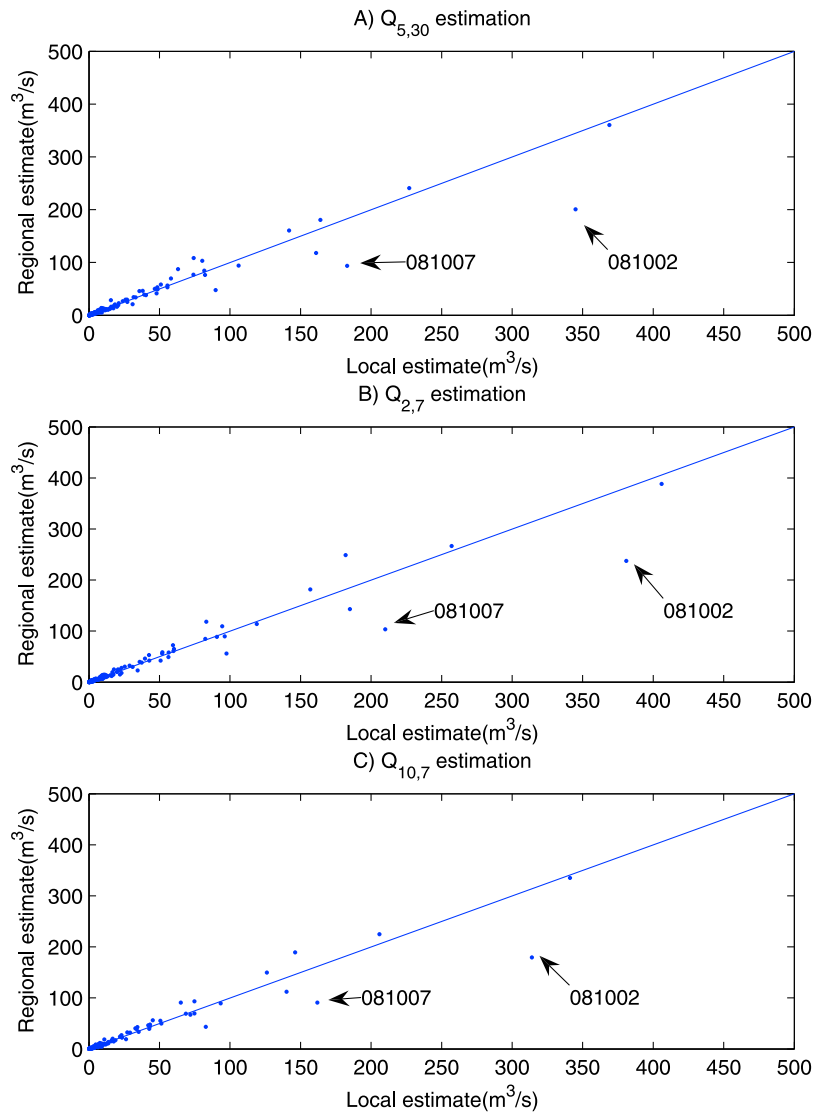**Figure 6.** Jackknife estimation of the summer low-flow quantiles using the regression approach.

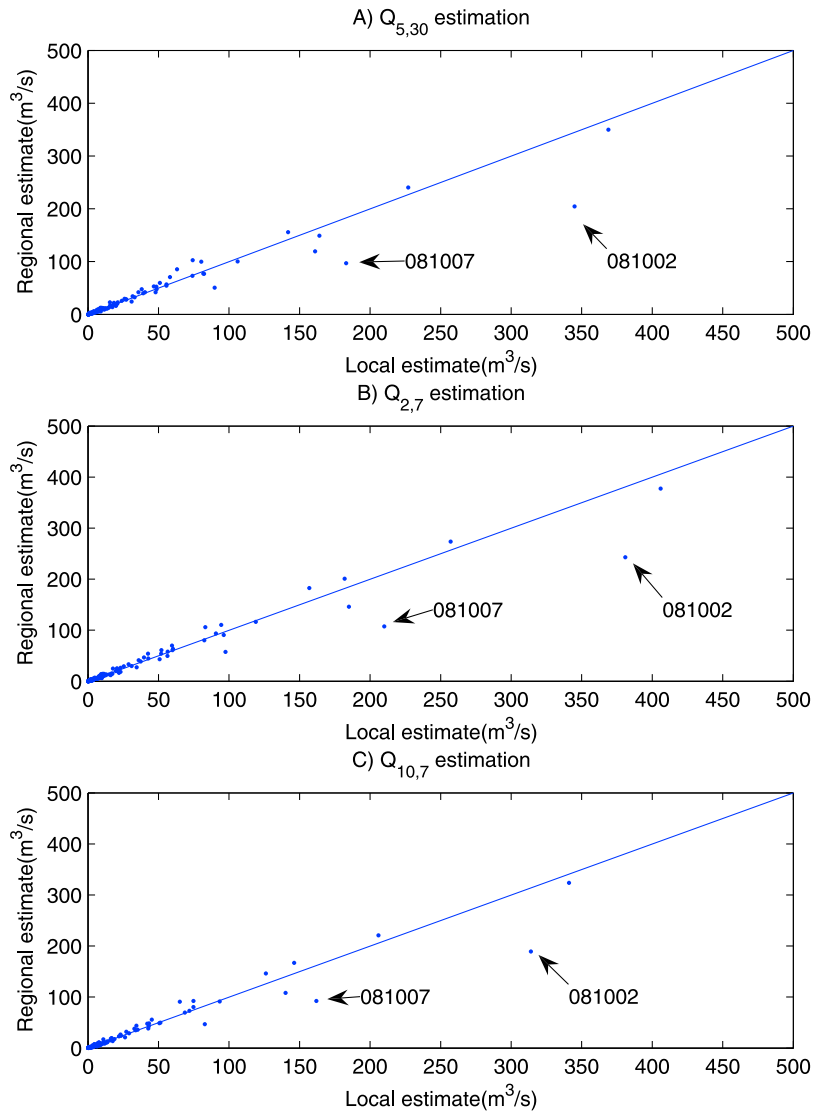**Figure 7.** Jackknife estimation of the winter low-flow quantiles using the SANN approach.

Figure 8. Jackknife estimation of the winter low-flow quantiles using the EANN approach.
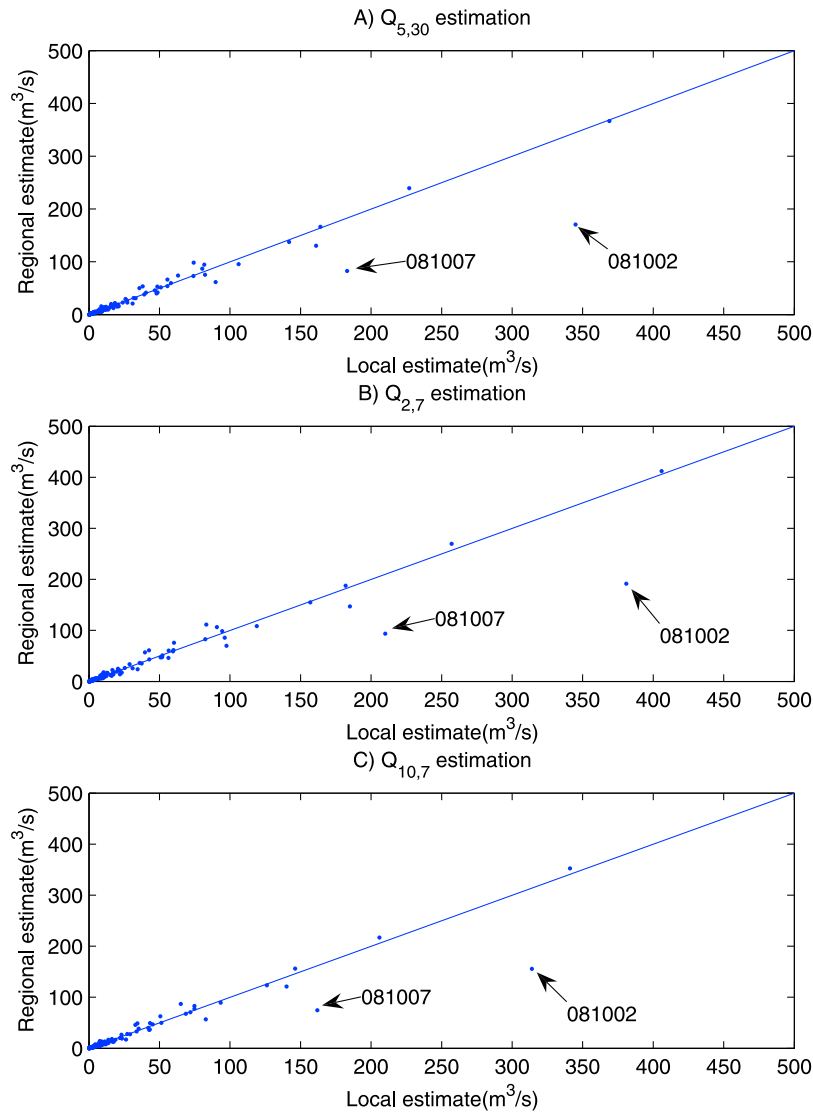
Figure 9. Jackknife estimation of the winter low-flow quantiles using the regression approach.

using both single and ensemble ANN models. For summer low-flow estimation, the ensemble ANN leads to better estimation compared to single ANN for all the performance indices, and the improvement in the relative measures including RMSEr and BIASr is significant. For winter low-flow estimation, the ensemble ANN leads to better estimation compared to single ANN in four out of the five performance indices, and the improvement in RMSEr is significant.

[45] Due to the architecture of the staked generalization, two different types of performance functions are used in ANN ensemble models. The level 0 generalizer and the level 1 generalizer in an ANN ensemble use performance functions in absolute and relative scale, respectively. This kind of setup in an ANN ensemble makes the training process essentially a multiobjective search for the optimal parameters (weights). The advantage of this ANN ensemble structure is manifested in the application to the estimation of winter low-flow quantiles in the study area. Compared to the single ANN model which shows a good performance in the *RMSE* performance measure and a bad performance in

the *RMSEr* measure, the ANN ensemble model simultaneously improves the performances in both *RMSE* and *RMSEr* measures.

[46] The type of ANN selected in the present paper for regional low-flow analysis is MLP. Further studies are required to investigate the feasibility of other types of ANNs such as radial basis network and generalized regression network. Indeed, these types of ANNs use very different architectures and learning algorithms for network

Table 5. Jackknife Validation Results Using Specific ANN Models

	Summer Low Flow			Winter Low Flow		
	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$	$Q_{5,30}$	$Q_{2,7}$	$Q_{10,7}$
NASH	0.95	0.95	0.94	0.89	0.89	0.90
RMSE	37.68	38.64	36.28	18.07	20.11	16.12
RMSEr(%)	51.17	49.15	60.99	37.73	34.49	44.58
BIAS	−3.19	−5.05	−1.49	0.39	0.32	1.03
BIASr(%)	−11.20	−10.85	−15.15	−8.57	−7.88	−8.34

training and simulation. A multimodel ANN ensemble based on these types of ANNs could also be employed for regional low-flow estimation.

References

Bauer, E., and R. Kohavi (1999), An empirical comparison of voting classification algorithms: Bagging, boosting, and variants, *Mach. Learn.*, 36(1–2), 105–139, doi:10.1023/A:1007515423169.

Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Clarendon, New York.

Breiman, L. (1996a), Bagging predictors, *Mach. Learn.*, 26(2), 123–140.

Breiman, L. (1996b), Stacked regressions, *Mach. Learn.*, 24(1), 41–64.

Burney, S. M. A., T. A. Jilani, and C. Ardil (2004), A comparison of first and second order training algorithms for artificial neural networks, paper presented at International Conference on Computational Intelligence, Int. Comput. Intell. Soc., Istanbul.

Carney, J. G., and P. Cunningham (1999), The NeuralBAG algorithm: Optimizing generalization performance in bagged neural networks, in *Proceedings of the 7th European Symposium on Artificial Neural Networks*, edited by M. Verleysen, pp. 35–40, D-Facto Public, Brussels.

Chokmani, K., and T. B. M. J. Ouarda (2004), Physiographical space-based kriging for regional flood frequency estimation at ungauged sites, *Water Resour. Res.*, 40, W12514, doi:10.1029/2003WR002983.

Crespo, J. L., and E. Mora (1993), Drought estimation with neural networks, *Adv. Eng. Software*, 18, 167–170, doi:10.1016/0965-9978(93)90064-Z.

Dawson, C. W., R. J. Abrahart, A. Y. Shamseldin, and R. L. Wilby (2006), Flood estimation at ungauged sites using artificial neural networks, *J. Hydrol.*, 319(1–4), 391–409, doi:10.1016/j.jhydrol.2005.07.032.

Demuth, H., and M. Beale (2003), MATLAB neural network toolbox, version 4, The MathWorks, Natick, Mass.

Dietterich, T. G. (2000), An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization, *Mach. Learn.*, 40(2), 139–157, doi:10.1023/A:1007607513941.

Dingman, S. L., and S. C. Lawlor (1995), Estimating low flow quantiles from drainage-basin characteristics in New Hampshire and Vermont, *Water Resour. Bull.*, 31(2), 243–256.

Drucker, H. (1997), Improving regressors using boosting techniques, in *Machine Learning: Proceedings of the Fourteenth International Conference*, pp. 107–115, Morgan Kaufmann, Burlington, Mass.

Drucker, H. (1999), Boosting using neural networks, in *Combining Artical Neural Nets*, edited by A. J. C. Sharkey, pp. 51–78, Springer, London.

Efron, B., and T. J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Eng, K., and P. C. D. Milly (2007), Relating low-flow characteristics to the base flow recession time constant at partial record stream gauges, *Water Resour. Res.*, 43, W01201, doi:10.1029/2006WR005293.

Freund, Y., and R. E. Schapire (1996), Experiments with a new boosting algorithm, in *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148–156, Morgan Kaufmann, San Francisco, Calif.

Govindaraju, R. S., and A. R. Rao (Eds.) (2000), *Artificial Neural Networks in Hydrology*, 329 pp., Kluwer Acad., Dordrecht, Netherlands.

Gupta, H. V., K. Hsu, and S. Sorooshian (1997), Superior training of artificial neural networks using weight-space partitioning, in *Proceedings of the IEEE 1997 International Conference on Neural Networks*, vol. 3, pp. 1919–1923, Inst. of Electr. and Electron. Eng., New York.

Hagan, M. T., and M. Menhaj (1994), Training feedforward networks with the Marquardt algorithm, *IEEE Trans. Neural Networks*, 5(6), 989–993, doi:10.1109/72.329697.

Hansen, L., and P. Salamon (1990), Neural network ensembles, *IEEE Trans. Pattern Anal. Mach. Intell.*, 12, 993–1001, doi:10.1109/34.58871.

Haykin, S. (1994), *Neural Networks—A Comprehensive Foundation*, 852 pp., Macmillan, New York.

Hsu, K., X. Gao, S. Sorooshain, and H. V. Gupta (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *J. Appl. Meteorol.*, 36(9), 1176–1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

Hu, M. Y., and C. Tsoukalas (2003), Explaining consumer choice through neural networks: The stacked generalization approach, *Eur. J. Oper. Res.*, 146(3), 650–660, doi:10.1016/S0377-2217(02)00368-5.

Incerti, G., E. Feoli, L. Salvati, L. Salvati, A. Brunetti, and A. Giovacchini (2007), Analysis of bioclimatic time series and their neural network-based classification to characterise drought risk patterns in south Italy, *Int. J. Bioclimatol.*, 51, 253–263.

Kendall, M. G. (1975), *Rank Correlation Methods*, Griffin, London.

Krogh, A., and J. Vedelsby (1995), Neural network ensembles, cross validation and active learning, in *Advances in Neural Information Processing Systems*, vol. 7, edited by D. S. Touretzky, G. Tesauro, and T. K. Leen, pp. 231–238, MIT Press, Cambridge, Mass.

Kuligowski, R. J., and A. P. Barros (1998), Experiments in short-term precipitation forecasting using artificial neural networks, *Mon. Weather Rev.*, 126(2), 470–482, doi:10.1175/1520-0493(1998)126<0470:EISTPF>2.0.CO;2.

Matignon, R. (2007), *Neural Network Modeling Using SAS Enterprise Miner*, AuthorHouse, Bloomington, Indiana.

McCuen, R. H. (2005), Accuracy assessment of peak discharge models, *J. Hydrol. Eng.*, 10, 16–22, doi:10.1061/(ASCE)1084-0699(2005)10:1(16).

Miller, R. G., Jr. (1964), A trustworthy jackknife, *Ann. Math. Stat.*, 35, 1594–1605, doi:10.1214/aoms/1177700384.

Minns, A. W., and M. J. Hall (1996), Artificial neutral networks as rainfall-runoff models, *Hydrol. Sci. J.*, 41(3), 399–417.

Mishra, A. K., V. R. Desai, and V. P. Singh (2007), Drought forecasting using a hybrid stochastic and neural network model, *J. Hydrol. Eng.*, 12, 626–638, doi:10.1061/(ASCE)1084-0699(2007)12:6(626).

Moradkhani, H., K. Hsu, H. Gupta, and S. Sorooshian (2004), Improved streamflow forecasting using self-organizing radial basis function artificial neural networks, *J. Hydrol.*, 295(1–4), 246–262, doi:10.1016/j.jhydrol.2004.03.027.

Morid, S., V. Smakhtin, and K. Bagherzadeh (2007), Drought forecasting using artificial neural networks and time series of drought indices, *Int. J. Climatol.*, 27, 2103–2111, doi:10.1002/joc.1498.

Ochoa-Rivera, J. C. (2008), Prospecting droughts with stochastic artificial neural networks, *J. Hydrol.*, 352(1–2), 174–180, doi:10.1016/j.jhydrol.2008.01.006.

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

Ouarda, T. B. M. J., V. Jourdain, N. Gignac, H. Gingras, H. Herrera, and B. Bobée (2005), Development of a hydrological model for the regional estimation of low-flows in the province of Quebec (in French), *Res. Rep. R-684-f1*, 174 pp., Eau, Terre, et Environ., Inst. Nat. de Rech. Sci., Sainte-Foy, Que., Canada.

Ouarda, T. B. M. J., C. Charron, and A. St-Hilaire (2008), Statistical models and the estimation of low flows, *Can. Water Resour. J.*, 33(2), 195–206.

Perrone, M. P., and L. N. Cooper (1993), When networks disagree: Ensemble methods for hybrid neural networks, in *Artificial Neural Networks for Speech and Vision*, edited by R. J. Mammone, pp. 126–142, Chapman and Hall, London.

Peters, R., G. Schmitz, and J. Cullmann (2006), Flood routing modelling with artificial neural networks, *Adv. Geosci.*, 9, 131–136.

Reed, D. W., and A. J. Robson (1999), *Flood Estimation Handbook*, vol. 3, Inst. of Hydrol., Wallingford, U. K.

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

Schapire, R. E. (1990), The strength of weak learnability, *Mach. Learn.*, 5(2), 197–227.

Seidou, O., T. B. M. J. Ouarda, L. Bilodeau, M. Hessami, A. St-Hilaire, and P. Bruneau (2006), Modeling ice growth on Canadian lakes using artificial neural networks, *Water Resour. Res.*, 42, W11407, doi:10.1029/2005WR004622.

Shao, J., and D. Tu (1995), *The Jackknife and Bootstrap*, Springer, Amsterdam.

Sharkey, A. J. C. (Ed.) (1999), *Combining Artificial Neural Nets: Ensemble and Modular Multi-net Systems*, Springer, London.

Shu, C., and D. H. Burn (2004), Artificial neural network ensembles and their application in pooled flood frequency analysis, *Water Resour. Res.*, 40, W09301, doi:10.1029/2003WR002816.

Shu, C., and T. B. M. J. Ouarda (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.

Shu, C., and T. B. M. J. Ouarda (2008), Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system, *J. Hydrol.*, *349*(1−2), 31−43, doi:10.1016/j.jhydrol.2007.10.050.

Smakhtin, V. U. (2001), Low flow hydrology: A review, *J. Hydrol.*, *240*(3−4), 147−186, doi:10.1016/S0022-1694(00)00340-1.

Smakhtin, V. U., D. A. Hughes, and E. Creuse-Naudine (1997), Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa, *Hydrol. Sci. J.*, *42*(6), 919−936.

Tasker, G. D. (1987), A comparison of methods for estimating low flow characteristics of streams, *Water Resour. Bull.*, *23*(6), 1077−1083.

The ASCE Task Committee on Artificial Neural Networks in Hydrology (2000a), Artificial neural networks in hydrology. I: Preliminary concepts, *J. Hydrol. Eng.*, *5*, 115−123, doi:10.1061/(ASCE)1084-0699(2000)5:2(115).

The ASCE Task Committee on Artificial Neural Networks in Hydrology (2000b), Artificial neural networks in hydrology. II: Hydrologic applications, *J. Hydrol. Eng.*, *5*, 124−137, doi:10.1061/(ASCE)1084-0699(2000)5:2(124).

Thomas, D. M., and M. A. Benson (1970), Generalization of streamflow characteristics from drainage-basin characteristics, *U.S. Geol. Surv. Water Supply Pap.*, *1975*.

Vogel, R. M., and C. N. Kroll (1990), Generalized low-flow frequency relationships for ungaged sites, *Water Resour. Bull.*, *26*(2), 241−253.

Vogel, R. M., and C. N. Kroll (1992), Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, *Water Resour. Res.*, *28*(9), 2451−2458, doi:10.1029/92WR01007.

Wald, A., and J. Wolfowitz (1943), An exact test for randomness in the nonparametric case based on serial correlation, *Ann. Math. Stat.*, *14*, 378, doi:10.1214/aoms/1177731358.

Wolpert, D. H. (1992), Stacked generalization, *Neural Networks*, *5*, 241−259, doi:10.1016/S0893-6080(05)80023-1.

Zhang, J. (1999), Developing robust non-linear models through bootstrap aggregated neural networks, *Neurocomputing*, *25*(1−3), 93−113, doi:10.1016/S0925-2312(99)00054-5.

————————————

T. B. M. J. Ouarda and C. Shu, INRS-ETE, University of Quebec, 490 de la Couronne, Quebec, QC G1K 9A9, Canada. (taha_ouarda@ete.inrs.ca; chang_shu@ete.inrs.ca)