Estimation of ANN prediction bounds for the suspended sediment load modeling

V Nourani, E Sharghi*, N J Paknejad

1,2,3 Department of Water Resources Engineering, Faculty of Civil Engineering University of Tabriz, Tabriz, Iran.

*Corresponding author – Email: sharghi@tabrizu.ac.ir

Abstract. In this paper, the point prediction of the Artificial Neural Network (ANN) for the suspended sediment load modeling was evaluated for the Lighvanchai River located in Iran, in monthly and daily scales. Since point prediction of ANN convey no information about the accuracy of prediction, so prediction intervals (PIs) were constructed by the Bootstrap method as a most frequently used technique for assessing the uncertainty of ANN. In this way, the accuracy of PIs was quantified by coverage and width criteria. The results showed that the ANN-based modeling in daily scale had better performance compared to that in monthly scale and Nash Sutcliff efficiency was 32% higher in daily scale compared to monthly. Moreover, the width and coverage of the constructed PIs in daily scale were 14% and 24%, lower and higher compared to that in monthly scale and the Bootstrap method could appropriately capture the target values.

1. Introduction

The need for accurate modeling of suspended sediment load has rapidly grown during the past decades in water resources and environmental engineering. Suspended load refers to the sediment that is pendant by the upward components of turbulent currents and remains in suspension for an appreciable length of time. Erosion and sediment transport phenomena in watersheds and rivers are complex hydrological and environmental problems [1]. Many models have been provided to simulate these phenomena [2,3]. Due to the large number of obscure parameters involved in this phenomenon, the theoretical governing equations may not be of much advantage in gaining knowledge of the overall process. Studies have been conducted to reduce the complexities of the problem in terms of developing practical techniques that do not require much algorithm and theory.

Nowadays, Artificial Neural Network (ANN), as self-learning and self-adaptive approximator function, has shown great ability in modeling and forecasting nonlinear time series. Capability of ANNs to establish nonlinear links between inputs and outputs makes them useful tools for modeling hydraulic and hydrological phenomena. ANN models have been successfully applied to many tasks in environment and hydrology engineering [4-6]. ANN's employment in suspended sediment estimation and prediction has recently been worked out in some previous studies [7, 8].

Despite the acceptable level of accuracy in the modeling, the major criticism that ANNs encounter is that they convey no information about the accuracy of the prediction. So some measures were
needed to quantify the accuracy. Prediction Interval (PI) is one of the well-known measures for this purpose. Several methods have been proposed for construction of PIs and assessment of the NN outcome uncertainty. Chryssolouris et al. [9] developed the delta technique through the nonlinear regression representation of the ANNs. The Bayesian technique is another method for construction of NN-based PIs. Training ANNs using the Bayesian technique allows error bars to be assigned to the predicted values of the network [10]. The method suffers from the massive computational burden and requires calculation of the Hessian matrix of the cost function for construction of PIs. Bootstrap is one of the most frequently used techniques in the literature for the construction of PIs for ANN point forecasts [11]. The main advantage of this method is its simplicity and ease of implementation. It does not require calculation of complex derivatives and the Hessian matrix involved in the delta and Bayesian techniques. This method has been used in some previous studies in hydrology. Srivastava et al. [12] used Bootstrap method to analysis the uncertainty associated to prediction and parameters in ANN hydrologic models, it is assumed that Bootstrap method is capable of applying to ANN-based hydrologic model and obviously describes the strengths and weaknesses of the ANN models. Kasiviswanathan and Sudheer [13] used the Bootstrap method and applied quantitative measures for assessing the uncertainty of predicting the river flow by the ANN. In Tiwari and Chatterjee [14], the influence of the length of training datasets in the Bootstrap method was analyzed. Moreover, it was illustrated that the Bootstrap method can solve the issue of over-fitting and under-fitting in training the sub-sets of the Bootstrap method. Singh and Panda [15] attempted to develop the unbiased ANN models with readily available climatic variables using shorter length of training dataset, for estimation of the daily sediment yield from a small agricultural watershed and its sub-watersheds in Eastern India. The bootstrap technique was used to avoid the neural network over-fitting for shorter length of training dataset. The Bootstrap ANN was applied to estimate the daily sediment yield by relating it to readily available climatic variables such as: rainfall and temperature. The study also reveals that the Bootstrap ANN models can provide more stable solution as compared to ANN models and improves the estimation accuracy for shorter length of training dataset. It was expressed that the most prominent feature of Bootstrap ANN is that it can provide more reasonable estimation for extremely high and low values, because of the different realisations of the training datasets and the developed Bootstrap ANN models have the potential of filling missing data for a daily sediment yield time series and for assessment of hydrological responses to climate change under data scarcity condition. It is assumed in previous studies that Bootstrap method is capable of applying to an ANN-based hydrologic model and obviously describes the strengths and weaknesses of the ANN models and the Bootstrap method can solve the issue of over-fitting and under-fitting in training the sub-sets of the Bootstrap method. Therefore, in this study Bootstrap method was used to quantify the PIs of ANN-based suspended sediment load modeling.

2. Materials and methods

2.1. Case study

In this paper is Lighvanchai watershed, located in northwest Iran at Azerbaijan province. Lighvanchai watershed is located between 37° 43′ and 37°50′ North latitude and 46°22′ and 46°28′ East longitude in the northern slope of Sahand Mountain (northwestern Iran). Watershed altitude is varying between 1263m and 3679m. The length of the Lighvanchai River is 28.5 km. The watershed area is approximately 142 km² (Figure.1). Watershed altitude varies between 1,263m and 3,679m above sea level. The rainfall peaks in winter and spring. The watershed contains medium vegetative land cover as a rural region. The topography is steep with an average slope of 11%. Consequently, the soils are disposed to erosion to some extent. The time series data of 28 years, from 1987 to 2015 for Lighvanchai River were employed in the modeling process, these records were divided into two sub-sets: the first 75% and the rest 25% were used for training and validation purposes, respectively.
2.2. Inputs selection for ANN modeling

Input selection analysis was required to identify the most informative inputs of the ANN models. One of the popular input selection methods is cross correlation analysis (CCA) which has been widely used in data-driven models [16, 17]. But according to the complex behavior of the hydrological time series, it has been criticized by Nourani et al. [18] which for modeling non-linear process via a non-linear model it is to have weak linear relation between input and output but a strong non-linear relation. Hence it is recommended to use a non-linear measure. Mutual Information (MI) is a non-linear measure for input selection and was applied in some previous studies (e.g. [19]).

Therefore, in this study in order to select the most important and dominant inputs for the ANN modeling, nonlinear MI measure between time series of suspended sediment load and runoff (with different lag times) was used. Entropy or information content is a statistical measure of the randomness or uncertainty in terms of probability distribution.

The MI of two random variables is a measure of the mutual dependence between two variables. MI between two random variables of X and Y is defined as [20]

\[
MI(X,Y) = H(X) + H(Y) - H(X,Y)
\]

Where H(X) and H(Y) are the entropies of X and Y, respectively. The entropy for random variable X with length of n with values X1, X2, ..., Xn and the corresponding probabilities P1, P2, ..., Pn can be computed as [21]

\[
H(X) = -\sum_{i=1}^{n} P(X_i) \log[P(X_i)]
\]

In Eq. 1, H(X,Y) is the joint entropy of X and Y which can be calculated by Gao et al.[22]:

\[
H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{n} P(X_i,Y_j) \log[P(X_i,Y_j)]
\]

2.3 Artificial neural network

There are several applications of ANN in suspended sediment load modeling. An ANN may be described as a network of interconnected neurons. The common structures of the ANNs consist of three layers. The first layer that connects with the input variables is named the input layer and the last layer relevant to the output variable is called the output layer. The layers in-between the input and output layers are the hidden layers, and there can be more than one hidden layer. Neurons that connect the layers have weights. The optimal set of weights is determined through learning process. The learning process consists of known input and output values, which train the network. While finding the known output using a known input, the best weights are calculated. The way that nodes in input and output layers are arranged and the direction that the data are processed in a network create networks with various characteristics. Feed-forward neural networks (FNNs) and recurrent neural networks
(RNNs) are the classifications for these networks. Among the applied neural networks, the FNN with the Back Propagation training algorithm is the most commonly used method in solving various engineering problems. There are several essential factors affecting the performance of ANN, including (I) predictors selection, (II) number of hidden neurons (network structure), and (III) specified training algorithm for connecting weights. There is no specific method to find the optimal number of layers and hidden neurons, except for the commonly used trial and error approach. The goal of an optimal ANN architecture is to minimize error between simulated output and the desired value with the most compact and simple structure possible.

2.4 Evaluation criteria

2.4.1 Point prediction evaluation

For the purpose of assessing efficiency of ANN-based modeling, criteria of root mean square error (RMSE, Eq. 4), Nash-Sutcliffe Efficiency (NSE Eq. 5), and CC (Eq. 6) were used in this study.

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (R_i - Z_i)^2}
\]

\[
\text{NSE} = 1 - \frac{\sum_{i=1}^{N} (R_i - Z_i)^2}{\sum_{i=1}^{N} (R_i - \bar{R})^2}
\]

\[
\text{CC} = \frac{\sum_{i=1}^{N} (R_i - \bar{R})(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^{N} (R_i - \bar{R})^2 \sum_{i=1}^{N} (Z_i - \bar{Z})^2}}
\]

In Eqs. 4, 5, 6 Ri and Zi are observed and output values, respectively. R and \( \bar{Z} \) are means of observed and target and finally N is the total number of observations.

2.4.2 PIs assessment measures

Two measures for the quantitative assessment of constructed PIs are PI coverage probability (PICP) and mean PI width (MPIW). PICP measures the percentage of observed values that fall within the prediction band. PICP as Eq. 7 is the measure that shows how good the PIs are [13, 23]:

\[
\text{PICP} = \frac{1}{N} \sum_{i=1}^{N} c_i
\]

if \( L(X_i) < x_i < U(X_i) \rightarrow c_i = 1 \); else \( c_i = 0 \)

Where \( U(X_i) \) and \( L(X_i) \) represent respectively the upper and lower bounds of PIs corresponding to the \( i \)th sample.

NMPIW is another criterion for quantification of the width of prediction bounds as [24]:

\[
\text{NMPIW} = \frac{1}{Rn} \sum_{i=1}^{n} L(X_i) - U(X_i)
\]

where R is the range of the underlying target. NMPIW is a dimensionless measure indicating how narrow PIs are.

2.5 PIs construction by Bootstrap method

Bootstrap method developed by Efron and Tibshirani which resamples the training datasets for the purpose of generating various models by training an individual network on each resampled instance of the original dataset (see Figure.2). This technique develops new sets of samples to present appropriate comprehension about the mean and variability of the main sample, independent of any knowledge about data distribution. Bootstrap is a simplified approach for quantification the uncertainty related to ANN prediction [13]. The positive point about the application of this technique is that it does not require complex derivatives of any non-linear function. If random samples are bootstrapped each time from the total available dataset, the simple arithmetic average of prediction can be considered as the model output corresponds to the \( i \)th input data point of ‘x’.
This process is performed by repeated sampling with replacement of the original dataset of size \( N \), to obtain \( B \) Bootstrap datasets, each with a size of \( N \).

Therefore, \( B \) neural network models are developed where each Bootstrap dataset may differ slightly. The mean and variance of outputs of these \( B \) networks are calculated as follows to compute PIs.

\[
\hat{y}_{\text{boot}}(x) = \frac{1}{B} \sum_{b=1}^{B} f(x_i; \text{P}_b) \tag{9}
\]

\[
\hat{\sigma}_{\text{boot}}^2(x) = \frac{1}{B-1} \sum_{b=1}^{B}(f(x_i; \text{P}_b) - \hat{y}_{\text{boot}}(x))^2 \tag{10}
\]

where \( \text{P}_b \) denotes the parameter obtained from \( b \)th bootstrap sample and \( f \) denotes the functional form of ANN model.

Confidence intervals can be constructed from \( \hat{y}_{\text{boot}} \) and \( \hat{\sigma}_{\text{boot}} \) [24]. To construct PIs a prediction \([l, u]\) for a future observation \( X \) in a normal distribution \( N(\mu, \sigma^2) \) with known mean and standard derivation may be calculated from:

\[
l = P(l < X < u) = P\left(\frac{l - \hat{y}_{\text{boot}}}{\hat{\sigma}_{\text{boot}}} < \frac{X - \hat{y}_{\text{boot}}}{\hat{\sigma}_{\text{boot}}} < \frac{u - \hat{y}_{\text{boot}}}{\hat{\sigma}_{\text{boot}}}\right) \tag{11}
\]

Where \( Z = \frac{X - \hat{y}_{\text{boot}}}{\hat{\sigma}_{\text{boot}}} \), the standard score of \( X \), is distributed as standard normal. Hence

\[
\frac{l - \mu}{\sigma} = -Z ; \quad \frac{u - \mu}{\sigma} = Z \tag{12}
\]

Or

\[
l = \mu - z\sigma ; \quad u = \mu + z\sigma \tag{13}
\]

Different values of \( Z \) is tabulated in Table 1.

| PI   | \( z \) |
|------|------|
| 75%  | 1.15 |
| 90%  | 1.64 |
| 95%  | 1.96 |
| 99%  | 2.58 |

### Results and Discussion

#### 3.1 Point prediction of ANN

In this study for modeling suspended sediment load, the MI measure was calculated between target (suspended sediment load) and runoff time series and also suspended sediment load series(with
different lag times). Then time series with maximum values of MI were selected as inputs of the ANN. The selected time series with different lag times were tabulated in Table 2. As shown in Table 2, for modeling in daily scale, combination of runoff time series with 1 and 2 steps lag time and suspended sediment load time series up to 3 steps lag time were selected as inputs of modeling. For modeling in monthly scale runoff with 1 step lag time and suspended sediment load up to 2 steps lag time were considered as inputs of modeling. In order to calibrate and validate the ANNs, the data set was divided into two parts, 75% of data were used for training and the rest 25% were used for the validation purpose. Therefore, periods 1978-2005 were used for the training set, and periods 2005-2015 for the validation set. A three-layer FFNN with BP training algorithm of levenberg-Marquardt and the tangent sigmoid (Tansig) as the activation function was used in modeling. To distinguish the optimum ANN structure, the epoch and hidden neuron numbers were obtained through the procedure of trial and error. The best training epoch number and optimum hidden neurons number were 100, 90 and 3, 5, respectively for the modeling in monthly and daily scales. The performance of the modeling was assessed via evaluation measures mentioned in previous sections and were tabulated in Table 3.

Table 2 Point prediction results

| Scale    | input                                  | NSE train | NSE verification | CC train | CC verification | RMSE* train | RMSE* verification |
|----------|----------------------------------------|-----------|------------------|----------|-----------------|-------------|--------------------|
| Daily    | S(t-1), S(t-2), S(t-3), Q(t-1), Q(t-2) | 0.98      | 0.9              | 0.9      | 0.9             | 0.007       | 0.0098             |
| Monthly  | S(t-1), S(t-2), Q(t-1)                 | 0.66      | 0.41             | 0.7      | 0.65            | 0.05        | 0.08               |

*implies normalized RMSE values

As shown in Table 2 NSE for daily scale modeling was 32% higher than that in Monthly scale. Modeling in monthly scale didn’t have appropriate performance compared to modeling in daily scale. At the monthly scale, this less desirable performance. Perhaps the most important reason that can be justified the disability of ANN model regards to the non-stationary nature of the input time series of sediment. Another point that should be noted is the different performance of each model in daily and monthly scales. Clearly, ANN dealt with more samples of the input data in the daily scale rather than the monthly scale. This would make the network better in training and increase the accuracy of the model with regard to the monthly scale. The different nature of daily and monthly time series should also be maneuvered in visitation of the sediment modelling. The monthly time series not only contain fewer samples than the daily time series, their seasonal behaviour is much remarkable than the Markovian characteristic [8]. In some previous studies in the hydrology similar results were obtained for example see [18, 25].

Table 3. Constructed PIs results for watersheds via the Bootstrap method.

| Scale    | Inputs  | PICP | NMPIW |
|----------|---------|------|-------|
| daily    | S(t-1), S(t-2), S(t-3), Q(t-1), Q(t-2) | 0.82  | 0.05  |
| Monthly  | S(t-1), S(t-2), Q(t-1)               | 0.58  | 0.19  |

The validation time series for modeling were illustrated in Fig. 3. As shown in Fig. 3, despite the appropriate performance of the ANN point prediction, it was unable to capture some peak points,
which may be due to lack of samples of peak points in the training set of the ANN. It has been expressed by some researches that ANN models are unable to predict the peak points accurately [26]. As the point prediction of ANN convey no information about the accuracy of the prediction, the Bootstrap method has been used for quantification of the constructed PIs and they have been evaluated by mentioned measures in previous sections.

3.2 Bootstrap method
To construct the PIs with the Bootstrap method, an ANN was trained 80 times, each time for modeling by randomly selected resampled 5114, 166 samples among 10227, 335 samples respectively, for daily and monthly time series. Trained weights were then used to simulate the outputs of the ANN model for all samples. The results of the constructed PIs were tabulated in Table 3. As shown in Table 3, the constructed PIs for daily scale had higher value of PICP compared to monthly scale modeling and NMPIW in daily scale was remarkable lower than monthly scale, which indicates the more reliable and accurate results of daily scale modeling than that in monthly scale. So it was concluded that training multiple NNs from various random initial points provides a better coverage of the parameter space. As shown in Fig.3 Bootstrap method could appropriately capture the peak point and overcome to the weak point of point prediction weakness in the calculation of the peak points. But Bootstrap method overestimated some points which may be caused by the small number of Bootstrap models. So considering higher numbers of models for training may improve the performance of modeling.

4. Conclusion
In this study in order to modeling suspended sediment load by the ANN, MI measure was used to select dominant time series as inputs of the ANN in both daily and monthly scales. As the point prediction of the ANN, conveys no information about the accuracy and reliability of outputs, the Bootstrap as the simple and reliable method was used to quantify the constructed PIs. Both point prediction and PIs showed that the ANN-based modeling in the daily scale led to more accurate results than the monthly scale. PICP value for constructed PIs for daily scale modeling was 24% higher than
that in monthly modeling and value of NMPIW was up to 14% lower than that for monthly scale. The Bootstrap method could appropriately capture the peak points which was the weakness of the ANN models, therefore using the Bootstrap method can lead to accurate results and had better performance than the classic ANN.

References
[1] Mirbagheri, S. A., Tanji, K. K., & Krone, R. B. (1988). Sediment characterization and transport in Colusa Basin Drain. Journal of Environmental Engineering, 114(6), 1257–1273.
[2] Jang, J.-S., & Sun, C.-T. (1995). Neuro-fuzzy modeling and control. Proceedings of the IEEE, 83(3), 378–406.
[3] Verstraeten, G., & Poesen, J. (2001). Factors controlling sediment yield from small intensively cultivated catchments in a temperate humid climate. Geomorphology, 40(1–2), 123–144.
[4] Anmala, J., Zhang, B., & Govindaraju, R. S. (2000). Comparison of ANNs and empirical approaches for predicting watershed runoff. Journal of Water Resources Planning and Management, 126(3), 156–166.
[5] Sahoo, G. B., Ray, C., Mehnert, E., & Keefer, D. A. (2006). Application of artificial neural networks to assess pesticide contamination in shallow groundwater. Science of the Total Environment, 367(1), 234–251.
[6] Nourani, V., Komasi, M., & Mano, A. (2009). A multivariate ANN-wavelet approach for rainfall–runoff modeling. Water Resources Management, 23(4), 2877.
[7] Jain, A., & Srinivasulu, S. (2006). Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. Journal of Hydrology, 317(3–4), 291–306.
[8] Sharghi, E., Nourani, V., Najafi, H., & Gokcekus, H. (2019). Conjunction of a newly proposed emotional ANN (EANN) and wavelet transform for suspended sediment load modeling. Water Supply. Image and Vision Computing (Vol. 26). https://doi.org/10.1016/j.imavis.2006.08.002
[9] Chryssolouris G., Lee M & Ramsey, A (1996) Confidence interval prediction for neural network models, in IEEE Transactions on Neural Networks, 7(1), 229–232
[10] MacKay, D. J. C. (1992). A practical Bayesian framework for back propagation networks. Neural Computation, 4(3), 448–472.
[11] Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.
[12] Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). Unsupervised learning of video representations using lstms. International Conference on Machine Learning.843–852.
[13] Kasiviswanathan, K. S., & Sudheer, K. P. (2013). Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. Stochastic Environmental Research and Risk Assessment, 27(1), 137–146.
[14] Tiwari, M. K., & Chatterjee, C. (2011). A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. Journal of Hydroinformatics, 13(3), 500–519.
[15] Singh, G., & Panda, R. K. (2015). Bootstrap-based artificial neural network analysis for estimation of daily sediment yield from a small agricultural watershed. International Journal of Hydrology Science and Technology, 5(4), 333–348.
[16] Brion, G. M., Neelakantan, T. R., & Lingireddy, S. (2001). Using neural networks to predict peak Cryptosporidium concentrations. Journal-American Water Works Association, 93(1), 99–105.
[17] Coulibaly, P., Ancilt, F., & Bobee, B. (2000). Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. Journal of Hydrology, 230(3–4), 244–257.
[18] Nourani, V., Kisi, Ö., & Komasi, M. (2011). Twohybridartificialintelligenceapproachesformodelingrainfall–runoff process. Journal of Hydrology, 402(1–2), 41–59.
[19] Elshorbagy, A., Corzo, G., Srinivasulu, S., & Solomatine, D. P. (2010). Experimental investigation of the predictive capabilities of datadrivenmodellingtechniquesinhydrology-Part1: Concepts and methodology. Hydrology and Earth System Sciences, 14(10), 1931–1941.
[20] Yang, H. H., Van Vuuren, S., Sharma, S., & Hermansky, H. (2000). Relevance of time–frequency features for phonetic and speaker-channel classification. Speech Communication, 31(1), 35–50.
[21] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

[22] Gao, Z., Gu, B., & Lin, J. (2008). Monomodal image registration using mutual information based methods.

[23] Khosravi, A., Nahavandi, S., & Creighton, D. (2010). A prediction interval-based approach to determine optimal structures of neural network metamodels. *Expert Systems with Applications*, 37(3), 2377–2387.

[24] Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks*, 22(9), 1341–1356.

[25] Campolo, M., Andreussi, P., & Soldati, A. (1999). River flood forecasting with a neural network model. *Water Resources Research*, 35(4), 1191–1197.

[26] Sudheer, K. P., Gosain, A. K., & Ramasastri, K. S. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes*, 16(6), 1325–1330.