

Prediction of Water-Level in the Urmia Lake Using the Extreme Learning Machine Approach

Jalal Shiri¹ · Shahaboddin Shamshirband² ·
Ozgur Kisi³ · Sepideh Karimi¹ · Seyyed M Bateni⁴ ·
Seyed Hossein Hosseini Nezhad⁵ · Arsalan Hashemi⁶

Received: 17 November 2015 / Accepted: 23 August 2016 /

Published online: 6 September 2016

© Springer Science+Business Media Dordrecht 2016

Abstract Predicting the dynamics of water-level in lakes plays a vital role in navigation, water resources planning and catchment management. In this paper, the Extreme Learning Machine (ELM) approach was used to predict the daily water-level in the Urmia Lake. Daily water-level data from the Urmia Lake in northwest of Iran were used to train, test and validate the employed models. Results showed that the ELM approach can accurately forecast the water-level in the Urmia Lake. Outcomes from the ELM model were also compared with those of genetic programming (GP) and artificial neural networks (ANNs). It was found that the ELM technique outperforms GP and ANN in predicting water-level in the Urmia Lake. It also can learn the relation between the water-level and its influential variables much faster than the GP and ANN. Overall, the results show that the ELM approach can be used to predict dynamics of water-level in lakes.

Keywords Extreme Learning Machine (ELM) · Water level · Time series · Urmia Lake

✉ Jalal Shiri
j_shiri2005@yahoo.com

✉ Shahaboddin Shamshirband
shamshirband1396@gmail.com

¹ Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

² Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

³ Civil Engineering Department, Architectural and Engineering Faculty, Canik Basari University, Samsun, Turkey

⁴ Department of Civil and Environmental Engineering & Water Resources Research Center, University of Hawaii, Manoa, HI, USA

⁵ Department of Computer Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran

⁶ Department of Civil Engineering, University of Tabriz, Tabriz, Iran

1 Introduction

Lakes typically provide water for various domestic, industrial and agricultural applications (Vuglinskiy 2009). Forecasting the water-level in lakes is necessary for water resources planning and management, lake navigation, management of tidal irrigation and agricultural drainage canals, etc. The water-level in lakes is a complex phenomenon, which is mainly controlled by the natural water exchange between the lake and its watershed, and thus the level reflects the hydrological changes in the watershed (Karimi et al. 2012; Altunkaynak 2007). For many practical applications, it is necessary to have a model to predict the water-level fluctuations based on previously recorded levels (Karimi et al. 2012). Numerous studies have been performed to predict water level fluctuations in lakes using neural networks, neuro-fuzzy, and genetic programming models (Kisi et al. 2015; Buyukyildiz et al. 2014; Karimi et al. 2012, 2013; Kisi et al. 2012; Shiri et al. 2011; Sulaiman et al. 2011).

Using the measured data from the Urmia lake in northwest of Iran, we use the Extreme Learning Machine (ELM) approach to predict its daily water-level for different prediction intervals.

Soft computing approaches have been used in many disciplines. Neural network (NN), as a well-known soft computing approach, has been employed in many water resources engineering problems. NN is able to solve complex nonlinear problems which may be difficultly solved by classic parametric approaches. There are many algorithms for training NNs, e.g. back propagation (BP), support vector machine (SVM), and hidden Markov model (HMM). One of the shortcomings of NN is its relatively large learning time. Huang et al. (2004) introduced an algorithm (which is known as Extreme Learning Machine (ELM)) for training the single layer feed forward NN. This algorithm is able to solve problems caused by ANNs' gradient descent based algorithms e.g. back propagation. ELM significantly reduces the training time of NN and generates accurate results (Huang et al. 2003). A number of studies have successfully employed the ELM algorithm in different problems (e.g. Yu et al. 2014; Wang and Han 2014; Ghouti et al. 2013; Sajjadi et al. 2016; Nian et al. 2014; Shamshirband et al. 2015).

In general, ELM is a robust algorithm with faster learning speed and a better performance compared to traditional algorithms such as back-propagation (BP), where it tries to get the smallest training error and norm of weights. To the our knowledge, there is not any published paper that applies the ELM for lake level prediction.

In this study, the ELM algorithm is used to predict the daily water-level in the Urmia Lake. Results indicate that the proposed algorithm can adequately predict the water level in the Urmia Lake for different prediction intervals. ELM results were also compared with those of the genetic programming (GP) and ANN-BP techniques.

This paper is organized as follows: Section 2 explains the Urmia Lake and the collected data. Section 3 presents the description of ELM algorithm and GP and ANN used as benchmark methods. The obtained results are presented in Section 4. Finally, the conclusions are given in Section 5.

2 The Urmia Lake

The Urmia Lake is the second largest saline lake in the world. Daily water-level data in the Urmia Lake in northwest of Iran were used in this study. Table 1 shows the statistical metrics

of daily water-level (*WL*) in the Urmia lake. It can be seen from Table 1 that the train and test data present a more skewed distribution than the validation data.

3 Soft Computing Algorithms

3.1 Extreme Learning Machine

(ELM was introduced by Huang et al. (2004) for the single layer feed-forward neural networks (SLFN) (Annema et al. 1994; Huang et al. 2006). This algorithm chooses the input weights of SLFN randomly, but determines the output weights analytically. The ELM algorithm is most robust with faster learning speed. ELM does not involve too much human intervention, because it determines all the network parameters analytically, and can be run much faster than the conventional algorithms.

3.1.1 Single Hidden Layer Feed-Forward Neural Network

The SLFN function with L hidden nodes might be presented as a network incorporating both additive and RBF hidden nodes in an unified way given (Huang et al. 2006; Liang et al. 2006):

$$f_L(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x), \quad x \in R^n, \quad a_i \in R^n \tag{1}$$

where a_i and b_i denote the learning parameters of hidden nodes, and β_i stands for the weight connecting the i th hidden node to the output node. The output of the i th hidden node with respect to the input x may be shown by $G(a_i, b_i, x)$. The additive hidden node with the activation function $g(x): R \rightarrow R$ (e.g., sigmoid and threshold), $G(a_i, b_i, x)$ is (Huang et al. 2006):

$$G(a_i, b_i, x) = g(a_i \cdot x + b_i), \quad b_i \in R \tag{2}$$

where x is the inner product of vector a_i and x in R^n . $G(a_i, b_i, x)$ can be determined for RBF hidden node with activation function $g(x): R \rightarrow R$ (e.g., Gaussian), $G(a_i, b_i, x)$ as (Huang et al. 2006):

$$G(a_i, b_i, x) = g(b_i \|x - a_i\|), \quad b_i \in R^+ \tag{3}$$

Table 1 Statistical metrics of the utilized data during the study period

	WL _{mean}	WL _{max}	WL _{min}	Standard deviation	Coefficient of variation	Skewness coefficient
Training set	1275.5	1277.0	1273.5	0.93	0.001	-0.93
Testing set	1275.7	1277.0	1275.0	0.40	0.001	0.89
Validation set	1276.6	1278.4	1275.0	0.98	0.001	0.07
whole dataset	1276.0	1278.4	1273.5	0.94	0.001	-0.13

a_i and b_i represent the center and impact factor of i th RBF node. The set of all positive real values is indicated by R^+ . Further details about RBF might be found in e.g. Huang et al. (2006).

3.1.2 Principle of ELM

Theorem 1: (Liang et al. 2006) Let an SLFN with L additive or RBF hidden nodes and an activation function $g(x)$ which is infinitely differentiable in any interval of R be given. Then for arbitrary L distinct input vectors $\{x_i | x_i \in R^n, i = 1, \dots, L\}$ and $\{(a_i, b_i)\}_{i=1}^L$ randomly produced by any continuous probability distribution, respectively, the hidden layer output matrix is invertible with probability one, the hidden layer output matrix H of the SLFN is invertible and $\|H\beta - T\| = 0$.

Theorem 2: (Liang et al. 2006) Given any small positive value $\varepsilon > 0$ and activation function $g(x) : R \rightarrow R$ which is infinitely differentiable in any interval, there exists $L \leq N$ such that for N arbitrary distinct input vectors $\{x_i | x_i \in R^n, i = 1, \dots, L\}$ for any $\{(a_i, b_i)\}_{i=1}^L$ randomly produced based upon any continuous probability distribution $\|H_{N \times L} \beta_{L \times m} - T_{N \times m}\| < \varepsilon$ with probability one.

Since the hidden node parameters of ELM should not be tuned throughout training and because they are easily assigned with random values, Eq. (5) becomes a linear system and the output weights can be estimated as (Huang et al. 2006):

$$\beta = H^+ T \quad (4)$$

where H^+ is the Moore-Penrose generalized inverse (Singh and Balasundaram 2007) of the hidden layer output matrix H which can be computed via several approaches consisting orthogonal projection, orthogonalization, iterative, singular value decomposition (SVD), etc. (Singh and Balasundaram 2007). The orthogonal projection method can be utilized only when $H^T T$ is non-singular and $H^+ = (H^T T)^{-1} H^T$. Owing to the use of searching and iterations, orthogonalization method and iterative method have limitations. Implementations of ELM uses SVD to compute the Moore-Penrose generalized inverse of H , because it can be utilized in all situations. ELM is thus a batch learning method.

3.2 Artificial Neural Networks

The multilayer feed forward network with a backpropagation learning algorithm is one of the most popular neural network architectures; it has been deeply studied and widely used in many fields (Yu et al. 2014). One of the most commonly used functions is the sigmoid function, which is monotonic increasing and ranges from zero to one. Details on ANNs can be found in e.g. Karunanithi et al. (1994). The parameters used for ANN are determined iteratively (see Table 2).

3.3 Genetic Programming

GP is an evolutionary algorithm based on Darwinian theories of natural selection and survival to approximate the equation, in symbolic form, that best describes how the output can relate to the input variables. The algorithm considers an initial population of

Table 2 User-defined parameters for ANN

ANN parameters				
Learning rate	Momentum	Hidden node	Number of iteration	Activation function
0.2	0.1	3,6,10	1000	Continuous Log-Sigmoid Function

randomly generated programs (equations), derived from the random combination of input variables, and random numbers and functions. The population of potential solutions is then subjected to an evolutionary process, then the 'fitness' (a measure of to which degree they can solve the problem) of the evolved programs is assessed. Then, the individual programs which produce the most accurate fit of data are selected from the initial population. The programs which gives the best fit are then selected to exchange part of the information between them to produce better programs through 'crossover' and 'mutation', which mimics the natural world's reproduction process. More details on GP can be found in e.g. Koza (1992). Table 3 summarizes the parameters used per run of GP.

4 Results and Discussion

Predictive models of water levels were established for 1- as well as 7-days ahead time intervals. The utilized input parameters consist of previous water levels from 1 to 6 days in the past. When developing multi-step ahead prediction models, two approaches are possible:

- Iterated prediction approach,
- Direct prediction approach.

Table 3 Parameters employed in GP modelling

Population size	512
Function set	+, -, *, /, $\sqrt{\quad}$, x^2 , $\ln(x)$, e^x , a^x
Head size	5–9
Chromosomes	20–30
Linking functions	Addition, 2subtraction, arithmetic, Trigonometric, Multiplication
Number of genes	2–3
Mutation rate	91.46
One-point recombination rate	0.2
Two-point recombination rate	0.2
Homologues crossover rate	98.46
Crossover rate	30.56
Fitness function error type	RMSE
Inversion rate	108.53
Gene transposition rate	0.1
Gene recombination rate	0.1

In the iterated prediction scheme, one step ahead prediction is used for building the subsequent predictions (i.e. for p steps ahead), while in direct prediction approach, separate (direct) prediction models are developed for each prediction horizon.

The advantages/shortcomings of each approach have been extensively discussed in literature (e.g. Martin 1989; Farmer et al. 1987; Marcellino et al. 2006). In the present paper, direct approach is employed. The main advantage of this approach is its intuitive manner (Chevillon and Hendry 2005), where there is no accumulation of forecasting error, in contrary to the iterated scheme. Nevertheless, the superiority of direct approach has been confirmed in electricity load forecasting (McSharry et al. 2005; Ramanathan et al. 1997; Fan and Hyndman 2012), where the domain is quite similar to one discussed here.

4.1 Input Variables for Model Building

The whole data set, comprising the observations for the period between 1972 and 2003 was divided into three blocks, including train set (50 % of observations), test set (25 % of observations), and validation test (25 % of observations). Figure 1 illustrates the daily water levels for the study period. A linear trend line was also fitted to each data set as depicted in the figure.

4.2 Evaluating Accuracy of Proposed Models

The performances of the proposed models were assessed using root means square error (RMSE), coefficient of determination (R^2) and Pearson correlation coefficient (r), defined as:

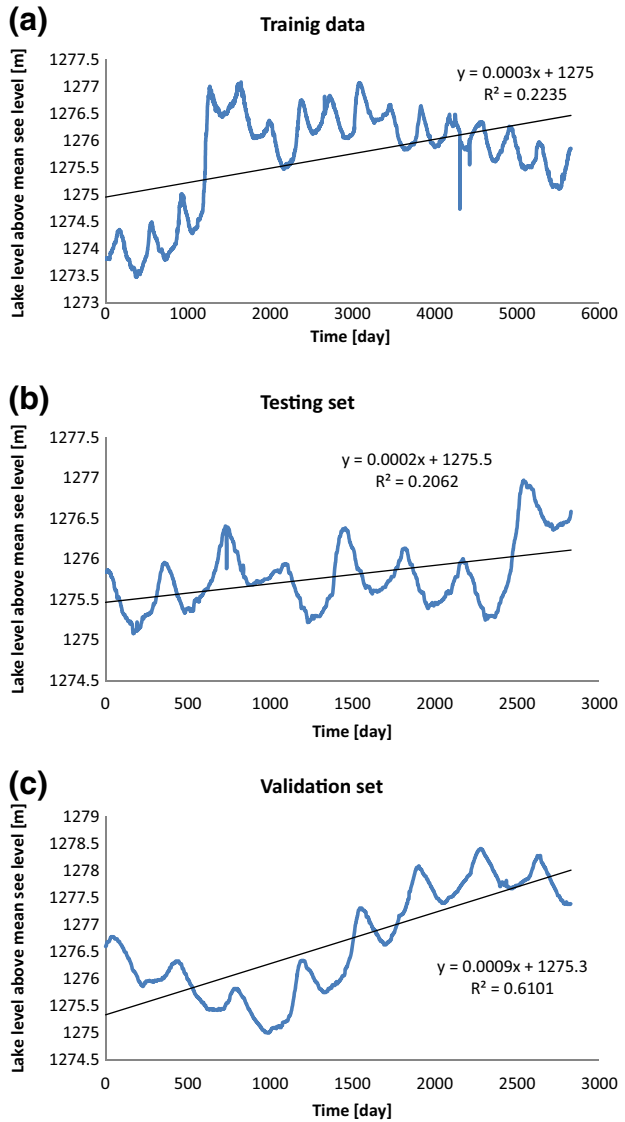
- 1) Root-mean-square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}, \quad (5)$$

- 2) Pearson correlation coefficient (r)

$$r = \frac{n \left(\sum_{i=1}^n O_i \cdot P_i \right) - \left(\sum_{i=1}^n O_i \right) \cdot \left(\sum_{i=1}^n P_i \right)}{\sqrt{\left(n \sum_{i=1}^n O_i^2 - \left(\sum_{i=1}^n O_i \right)^2 \right) \cdot \left(n \sum_{i=1}^n P_i^2 - \left(\sum_{i=1}^n P_i \right)^2 \right)}} \quad (6)$$

Fig. 1 Time series of the observed water level for **a** train set, **b** test set and **c** validation set



3) Coefficient of determination (R^2)

$$R^2 = \frac{\left[\sum_{i=1}^n (O_i - \bar{O}_i) \cdot (P_i - \bar{P}_i) \right]^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2 \cdot \sum_{i=1}^n (P_i - \bar{P}_i)^2} \tag{7}$$

where P_i and O_i are the observed and forecasted values of water level, respectively, and n is the total number of data patterns.

Fig. 2 Scatter plots of the observed vs. predicted water levels using ELM method for 1 day ahead: **a** train set, **b** test set, and **c** validation set

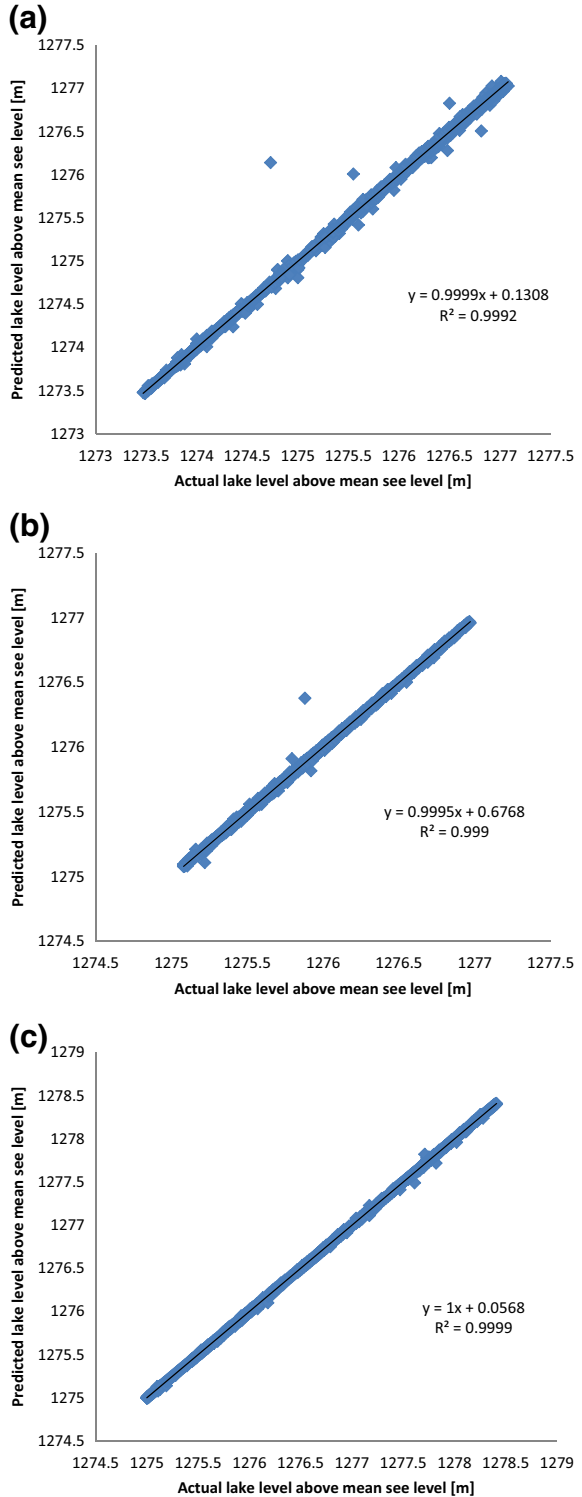
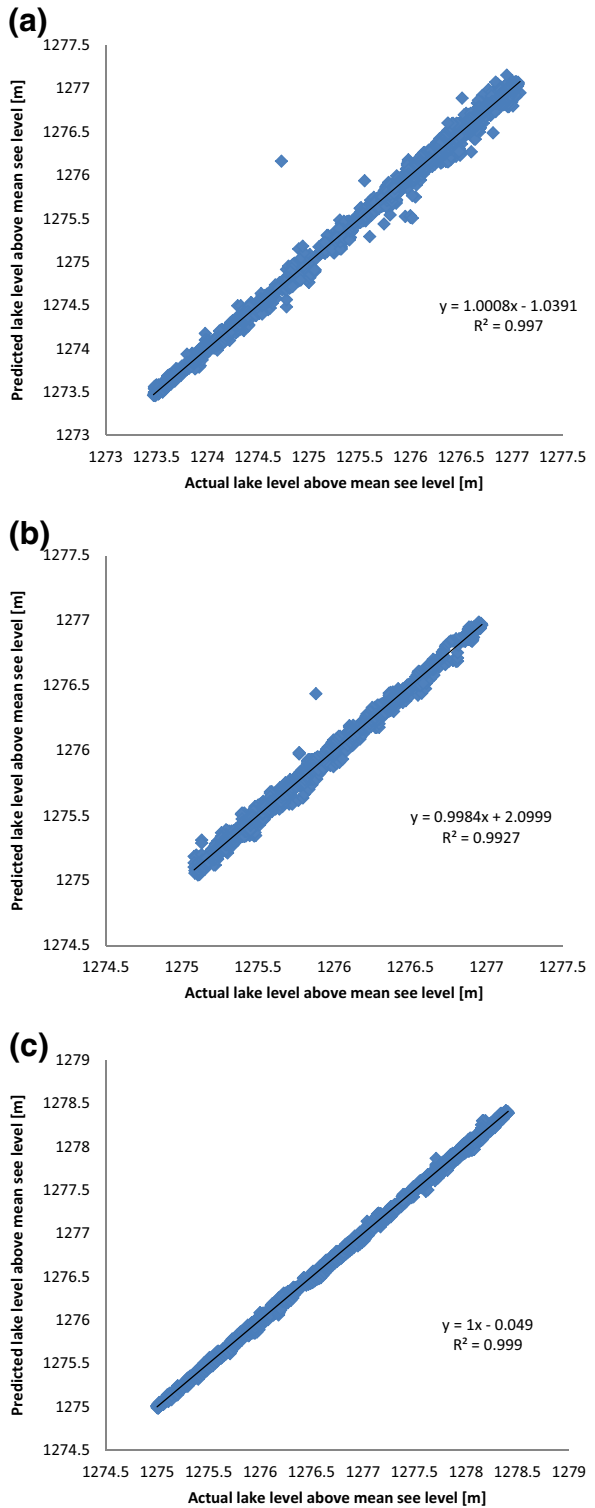


Fig. 3 Scatter plots of actual and predicted lake level values using ELM method for 7-days ahead for **a** training set, **b** testing set and **c** validation set



4.3 Performance Evaluation of the Proposed ELM Algorithms

Figure 2 presents the scatterplots of the observed vs. predicted water levels by ELM during the train, test and validation periods for the 1-day ahead prediction interval. From the scatters, it is seen that the most of the points are located along the diagonal line, so, it follows that the prediction results of ELM are in very good agreement with the corresponding measured water levels. The corresponded R^2 values (in the scatters) also confirm this statement.

The scatter plots of the observed and 7-days ahead predicted water levels have been presented in Fig. 3. Analyzing the scatterplots it is clear that the ELM algorithm give promising results in predicting 7-days ahead water levels, but, expectedly, the prediction accuracy has been reduced (to a less extent) compared to 1-day ahead predictions.

Figure 4 shows the time series variations of the water level predictions (using ELM) and corresponding observed values for the validation period. Although clear differences can be seen between the results of two prediction intervals (where the outcomes of 1-day ahead prediction is much better than those of the 7-days ahead prediction), ELM is able to predict the water levels in both the prediction intervals rigorously.

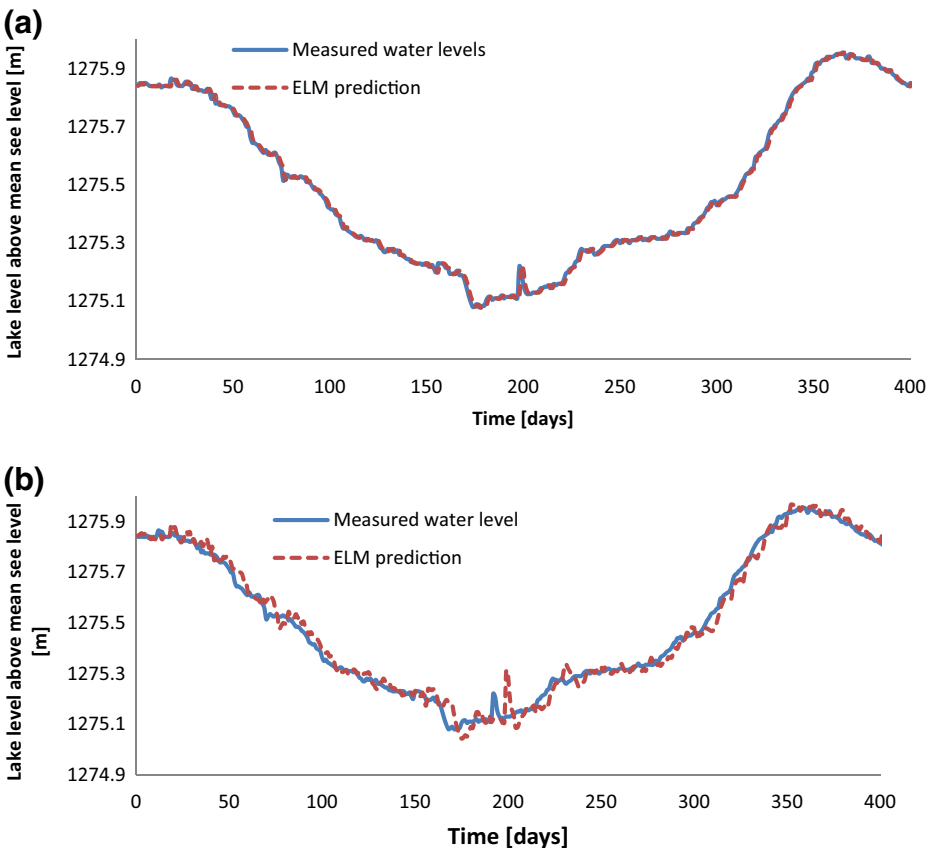


Fig. 4 Predicted water levels (suing ELM) against corresponding measured data for a 1-day ahead for b 7-days ahead

Table 4 Statistical parameters of the ELM, ANN and GP models

Prediction horizon	ELM			ANN			GP		
	RMSE	R ²	r	RMSE	R ²	r	RMSE	R ²	r
	1 day ahead	Train 0.02573	0.9992	0.999613	0.028244	0.9991	0.999535	0.028234	0.9991
	Test 0.009254	0.999	0.999742	0.002979	0.9986	0.999974	0.002981	0.9986	0.999974
	validation 0.000947	0.9999	0.999	0.00093	0.9999	1	0.001347	0.9999	0.999999
7 dayss ahead	Train 0.05052	0.997	0.998509	0.053904	0.9967	0.998503	0.053484	0.9966	0.99833
	Test 0.035006	0.9927	0.996349	0.036516	0.9921	0.996035	0.036861	0.9919	0.995945
	validation 0.03112	0.999	0.999505	0.029405	0.9991	0.999559	0.032507	0.9989	0.99946

5 Performance Comparison of ELM, ANN and GP

In order to assess the proposed ELM approach with respect to the GP and conventional ANN methods, a comparison was also made among the results of these models. Table 4 summarizes the prediction accuracy of each model for train, test and validation periods.

For training data set, all three methods provide similar results. In case of 1 day ahead water level forecasting, ELM performs better than the ANN and GP models in validation period. In case of 7 days ahead prediction, however, the ELM shows better performance than GP. The ANN seems to be slightly better than the ELM model in this case.

6 Conclusion

The ability of ELM model was investigated in prediction of daily water levels in Urmia Lake. Observations collected in Urmia Lake (Northwestern Iran) during from 1972 to 2003 were utilized for model building and evaluation. Two predictive models, for different prediction horizons (1 day and 7 days ahead), were created using the ELM method. A comparison of ELM method with GP and ANN was performed in order to assess the prediction accuracy. Comparison of the results in terms of RMSE, r and R^2 , indicated that the ELM model is superior to the GP and ANN approaches in 1 day ahead water level prediction. The results suggested that the proposed ELM model can be successfully used in water level forecasting in lakes.

References

- Altunkaynak A (2007) Forecasting surface water level fluctuations of Lake Van by artificial neural networks. *Water Resour Manag* 21(2):399–408
- Annema AJ, Hoen K, Wallinga H (1994) Precision requirements for single-layer feedforward neural networks. In: Fourth International Conference on Microelectronics for Neural Networks and Fuzzy Systems, 145–151
- Buyukyildiz M, Tezel G, Yilmaz V (2014) Estimation of the change in lake water level by artificial intelligence methods. *Water Resour Manag* 28(13):4747–4763
- Chevillon G, Hendry DF (2005) Non-parametric direct multi-step estimation for forecasting economic processes. *Int J Forecast* 21(2):201–218
- Fan S, Hyndman RJ (2012) Short-term load forecasting based on a semi-parametric additive model. *IEEE Trans Power Syst* 27(1):134–141
- Farmer JD, John J, Sidorowich S (1987) Predicting chaotic time series. *Phys Rev Lett* 59(8):45–848
- Ghouthi L, Sheltami TR, Alutaibi KS (2013) Mobility prediction in mobile Ad Hoc networks using extreme learning machines. *Procedia Comput Sci* 19:305–312
- Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. *Int Joint Conf Neural Netw* 2:985–990
- Huang GB, Zhu QY, Siew CK (2003) Real-time learning capability of neural networks, Technical Report ICIS/45/2003, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, April 2003
- Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70: 489–501
- Karimi S, Shiri J, Kisi O, Makarynskyy O (2012) Forecasting water level fluctuations of Urmieh Lake using gene expression programming and adaptive neuro- fuzzy inference system. *Int J Ocean Clim Syst* 3(2):109–125
- Karimi S, Kisi O, Shiri J, Makarynskyy O (2013) Neuro fuzzy and neural network techniques for forecasting sea level in Darwin Harbor, Australia. *Comput Geosci* 52:50–59
- Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. *J Comput Civ Eng* 8(2):201–220

- Kisi O, Shiri J, Nikoofar B (2012) Forecasting daily lake levels using artificial intelligence approaches. *Comput Geosci* 41:169–180
- Kisi O, Shiri J, Karimi S, Shamshirband S, Motamedi S, Petković D, Hashim R (2015) A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. *Appl Math Comput* 270:731–743
- Koza JR (1992) Genetic programming: on the programming of computers by natural selection. MIT Press, Cambridge, MA
- Liang NY, Huang GB, Rong HJ, Saratchandran P, Sundararajan N (2006) A fast and accurate on-line sequential learning algorithm for feedforward networks. *IEEE Trans Neural Netw* 17:1411–1423
- Marcellino M, Stock J, Watson MW (2006) A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *J Econ* 135:499–526
- Martin C (1989) Nonlinear prediction of chaotic time series. *Physica D* 35:335–356
- McShary PE, Bouwman S, Bloemhof G (2005) Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Trans Power Syst* 20:1166–1172
- Nian R, He B, Zheng B, Heeswijk MV, Yu Q, Míche Y (2014) Extreme learning machine towards dynamic model hypothesis in fish ethology research. *Neurocomputing* 128:273–284
- Ramanathan R, Engle RF, Granger CWJ, Vahid F, Brace C (1997) *J Forecast* 13:161–174
- Sajjadi S, Shamshirband S, Alizamir M, Yee L, Mansor Z, Manaf AA, Altameem TA, Mostafaeipour A (2016) Extreme learning machine for prediction of heat load in district heating systems. *Energy Build* 122:222–227
- Shamshirband S, Mohammadi K, Chen HL, Samy GN, Petković D, Ma C (2015) Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: a case study for Iran. *J Atmos Sol Terr Phys* 134:109–117
- Shiri J, Makarynsky O, Kisi O, Dierickx W, FakheriFard A (2011) Prediction of short term operational water levels using an adaptive neuro-fuzzy inference system. *J Waterw Port Coast Ocean Div ASCE* 137(6):344–354
- Singh R, Balasundaram S (2007) Application of extreme learning machine method for time series analysis. *Int J Intell Technol* 2:256–62
- Sulaiman M, El-Shafie A, Karim O, Basri H (2011) Improved water level forecasting performance by using optimal steepness coefficients in an artificial neural network. *Water Resour Manag* 25:2525–2541
- Vuglinskiy V (2009) Water Level: water level in lakes and reservoirs, water storage. Assessment of the status of the development of the standards for the terrestrial essential climate variables, Global Terrestrial Observing System (GTOS), Rome, Italy, pp 26
- Wang X, Han M (2014) Online sequential extreme learning machine with kernels for nonstationary time series prediction. *Neurocomputing* 145:90–97
- Yu Q, Míche Y, Séverin E, Lendasse A (2014) Bankruptcy prediction using extreme learning machine and financial expertise. *Neurocomputing* 128:296–302