CrossMark

ORIGINAL ARTICLE

A novel hybrid artificial intelligent approach based on neural fuzzy inference model and particle swarm optimization for horizontal displacement modeling of hydropower dam

Kien-Trinh Thi Bui[1,2] · **Dieu Tien Bui**[3] 🄳 · **Jingui Zou**[1] · **Chinh Van Doan**[4] · **Inge Revhaug**[5]

Abstract Horizontal displacement of hydropower dams is a typical nonlinear time-varying behavior that is difficult to forecast with high accuracy. This paper proposes a novel hybrid artificial intelligent approach, namely swarm optimized neural fuzzy inference system (SONFIS), for modeling and forecasting of the horizontal displacement of hydropower dams. In the proposed model, neural fuzzy inference system is used to create a regression model whereas Particle swarm optimization is employed to search the best parameters for the model. In this work, time series monitoring data (horizontal displacement, air temperature, upstream reservoir water level, and dam aging) measured for 11 years (1999–2010) of the Hoa Binh hydropower dam were selected as a case study. The data were then split into a ratio of 70:30 for developing and validating the hybrid model. The performance of the resulting model was assessed using RMSE, MAE, and R^2. Experimental results show that the proposed SONFIS model performed well on both the training and validation datasets. The results were then compared with those derived from current state-of-the-art benchmark methods using the same data, such as support vector regression, multilayer perceptron neural networks, Gaussian processes, and Random forests. In addition, results from a Different evolution-based neural fuzzy model are included. Since the performance of the SONFIS model outperforms these benchmark models with the monitoring data at hand, the proposed model, therefore, is a promising tool for modeling horizontal displacement of hydropower dams.

Keywords Horizontal displacement · Hydropower dam · Neural fuzzy · Hoa Binh · Vietnam

1 Introduction

Safety of hydropower dams requires comprehensive understanding of the mechanism of their deformation processes; therefore, monitoring models for forecasting dam behavior are important. These are considered to be key components of dam safety systems that help in the day-to-day operation and long-term assessment of dams [1]. However, the displacement of hydropower dams is a typical nonlinear and complex process that is influenced by many factors such as seismic load, sediment pressure, water pressure, air temperature, and rock deformability. Therefore, it is not easy to forecast dam behavior with high accuracy [2].

Various methods and techniques have been proposed for dam displacement modeling over the last decade and they can be grouped in three categories such as deterministic, statistical, and machine learning methods [3–5]. Deterministic methods (i.e., the finite element method (FEM) [6] and the boundary element method (BEM) [7]) use information on material properties, acting loads, and stress–strain laws to establish displacement functions. These functions are then

✉ Dieu Tien Bui
Dieu.T.Bui@hit.no; BuiTienDieu@gmail.com

1 School of Geodesy and Geomatics, Wuhan University, No. 129 Luo Yu Road, Wuhan 430072, Hubei, China

2 Geomatics Center, Water Resources University, No. 175 Tay Son Street, Hanoi, Vietnam

3 Geographic Information System Group, University College of Southeast Norway, Hallvard Eikas Plass1, 3800 Bø i Telemark, Norway

4 Faculty of Surveying and Mapping, Le Quy Don Technical University, 100 Hoang Quoc Viet Street, Hanoi, Vietnam

5 Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003 IMT, 1432 Aas, Norway

used to forecast displacements in different scenarios. Thus, deterministic methods are considered as the most widely used in dam deformation modeling. They are especially useful in the design phase, the filling phase, and the early stage of dam operation where no or only short time series monitoring data are available. However, due to many uncertainties, i.e., imperfect knowledge of the material properties, the estimated displacements from the deterministic methods and the observed values may be significantly different [8]. Therefore, statistical and machine learning methods have been proposed.

When long time series of monitoring data are available, statistical methods could be used for deformation analysis. Statistical methods have many advantages such as simplicity of formulation and speed of execution [9]; however, dam deformation is a typical nonlinear time-varying behavior process; therefore, statistical methods require collection of a large number of samples to produce reliable results, which is difficult to obtain in many cases [10]. Nevertheless, it is still difficult to forecast hydropower dam deformation with high accuracy [2].

Due to the criticality issues of the dam deformation, machine learning techniques have recently been explored and employed for the dam displacement modeling. Of these methods, artificial neural network models are well suited to deal with nonlinear and complex interactions of input–output relations for dams [1]. Literature review shows that they are the most used methods for dam modeling, i.e., in Seyedpoor et al. [11], Karimi et al. [12], Mata [4], Ranković et al. [13], and Kao and Loh [14]. In more recent years, support vector regression has been explored for accurate forecasting of dam displacement with promising results in many studies such as in Zheng et al. [15], Ranković et al. [16], Su et al. [17], and Salazar et al. [2]. Further models have also been investigated for dam deformation analysis, such as Random forests, boosted regression tree [18], and multivariate adaptive regression splines [2]. In general, machine learning methods are effective alternative tools for modeling of dam displacement [14].

Among the machine learning approaches, neural fuzzy is accepted to be an effective tool for modeling nonlinear time-varying behavior of a dam. Seyedpoor et al. [11] used an adaptive neural fuzzy inference system for the selection of input variables for finding the optimal shape of arch dams, with the conclusion that the neural fuzzy model is an efficient tool. Ranković et al. [13] proposed a neural fuzzy model to predict the radial displacement of the arch dam where the best model parameters were found using the traditional gradient and the least squares methods. They concluded that the neural fuzzy model was an effective tool for modeling of behaviors of the arch dam.

It is clear that the performance of a neural fuzzy model is strongly influenced by the premise and consequent parameters; therefore, they must be carefully selected. Thus, searching for optimal parameter values becomes an optimization problem in soft computing [19–21]. Literature reviews shows that various newly metaheuristic optimization techniques have been proposed such as Differential evolution [22, 23], Genetic algorithm [24, 25], Simulated annealing [26], Ant colony optimization [27], and Artificial bee colony [28, 29]. However, an integration of metaheuristic optimizations and neural fuzzy models for displacement modeling of hydropower dams has not been investigated.

We fill this gap in the literature by proposing a hybrid artificial intelligence approach based on a neural fuzzy inference system with metaheuristic optimization (namely SONFIS) for dam deformation modeling. The Hoa Binh hydropower dam (Vietnam) is selected as a case study. For this purpose, Particle swarm optimization (PSO) [30], which is the most promising and powerful metaheuristic optimization technique [31, 32], was selected to search the best parameters for the SONFIS model. In addition, the usability of the proposed model is assessed through comparisons with other methods such as Support vector regression (SVR), Multilayer perceptron neural networks (MLP Neural Nets), Gaussian processes, Random forest, and Different evolution-based neural fuzzy inference system (DE-FIS) using the same data.

In this research, the time series data preprocessing and visualization were carried out using Microsoft Excel®2013. The neural fuzzy algorithm is implemented in the Fuzzy Logic Toolbox™ in Matlab®2014. The proposed SONFIS model that combines the neural fuzzy algorithm and PSO is programed by the authors in Matlab®2014. The modeling process using SVR, MLP Neural Nets, Gaussian processes, and Random forests was carried out using WEKA®3.7.10 (The University of Waikato, Hamilton City, New Zealand). The rest of the paper is organized as follows: the second section describes briefly the theoretical background of the ANFIS algorithm and PSO; the third section provides description of the study area and the collected data; the next section depicts the proposed hybrid artificial intelligence approach (SONFIS); the results, discussion, and comparison are presented in the fifth section. Concluding remarks of this research are stated in the final section.

2 Theoretical background of the method used

2.1 Neural fuzzy inference system

Adaptive neural fuzzy inference system proposed by Jang [33] is a combination of artificial neural networks and the Sugeno-type fuzzy inference system. Structurally, this is a feedforward multilayer neural network designed with five layers: fuzzy layer, rule layer, normalization layer,
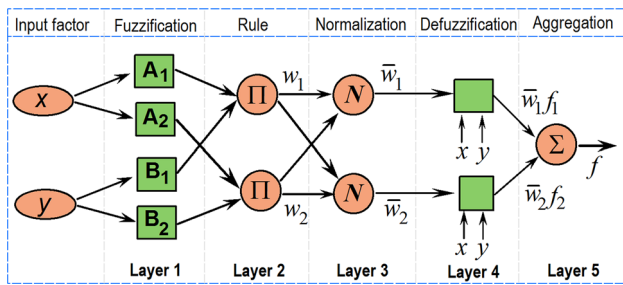
Fig. 1 A general architecture of an adaptive neural fuzzy inference system

defuzzification layer, and aggregation layer. Figure 1 illustrates a general ANFIS architecture (with two input factors, one output, and two rules) in which circles and squares are fixed nodes and adaptive node functions, respectively.

Layer 1: This layer is called the fuzzification layer where fuzzy membership values for input factors are generated using a membership function as follows:

$$O_{1,i} = \mu_{A_i(x)}; \quad O_{2,i} = \mu_{B_i(y)} \quad \text{for} \quad i = 1, 2 \tag{1}$$

where A and B are the linguistic variables, specified by the membership functions, $\mu_{A_i(x)}$ and $\mu_{B_i(y)}$. In this study, Gaussian membership function was used (Eq. 2).

$$\mu_{A_i(x)} = e^{\frac{-(x-c)^2}{2\delta^2}} \tag{2}$$

where δ and c are called antecedent parameters that control the shape of the Gaussian membership function.

Layer 2: This layer is the rule layer where each node in this layer has one fuzzy rule with firing strength (ω_i). This firing strength is calculated as the product of all incoming signals.

$$O_{2,i} = \omega_i = \mu_{A_i(y)} \cdot \mu_{B_i(y)}; \quad with \quad i = 1, 2 \tag{3}$$

Layer 3: This layer is called the normalization layer where the output firing strength ($\bar{\omega}_i$) for each node is normalized using an equation as follows:

$$O_{3,i} = \bar{\omega}_i = \omega_i / (\omega_1 + \omega_2); \quad i = 1, 2 \tag{4}$$

Layer 4: This layer is called the defuzzification layer where the value of the rule consequence for each node is calculated as follows:

$$O_{4,i} = \bar{\omega}_i \cdot f_i = \bar{\omega}_i \cdot (p_{0i} + p_{1i}x + p_{2i}y); \quad i = 1, 2 \tag{5}$$

where p_{0i}, p_{1i}, and p_{2i} are the consequent parameters of the output function f_i.

Layer 5: This layer is called the aggregation layer that transforms fuzzy values into a crisp output using an equation as follows.

$$O_{5,i} = \sum_i \bar{\omega}_i f_i = \sum_i \omega_i f_i \Big/ \sum_i \omega_i; \quad i = 1, 2 \tag{6}$$

2.2 Particle swarm optimization

Particle swarm optimization (PSO) is a relatively new swarm intelligence technique proposed by Kennedy and Eberhart [30]. This technique was developed based on the social behavior of bird flocks for solving complex optimization problems. Using a population (swarm), PSO scatter a number of individuals (particles) of the swarm in the domain space, to find the best position which is the optimized solution of an optimization problem [34].

Assume that D is the dimension of the search space with m is the population size. x_i and v_i ($i = 1,…, m$) are position and velocity vectors for particle i, respectively. *pbest* is the best position of an individual particle in the swarm, and *gbest* is the best position of all particles in the swarm. The evolutionary process of the PSO algorithm is summarized as follows:

Step 1: Initialization. An initial population (swarm) with random positions and velocities was generated in the D dimension of the search space.
Step 2: Fitness evaluation. A fitness function must be determined, and in this study, RMSE is used as the fitness function. RMSE for each particle in the swarm is evaluated to determine *pbest,* and the particle with the lowest RMSE is used as *gbest.*
Step 3: Change and update. The position and velocity of each particle in the swarm are changed using Eqs. 7, and 8 and then *pbest* and *gbest* are updated.

$$v_{i+1} = \omega v_i + ac_1 r_1(pbest - x_i) + ac_2 r_2(gbest - x_i) \tag{7}$$

$$x_{i+1} = x_i + v_{i+1} \tag{8}$$

where ω is the inertia weight; ac_1 and ac_2 are the acceleration coefficients; r_1 and r_2 are two randomly generated numbers in the range [0, 1].

Step 4: Termination. Stop the algorithm if the number of iteration reaches the predetermined maximum number of iterations; or return to step 2 otherwise.

2.3 Accuracy assessment

The accuracy of the models is measured using the root mean square error (*RMSE*), the mean absolute error (MAE), and the correlation coefficient (R^2) [1] as follows:

$$RMSE = \text{Sqrt} \sum_{i=1}^{n} \left((pr_i - y_i)^2 / n \right);$$

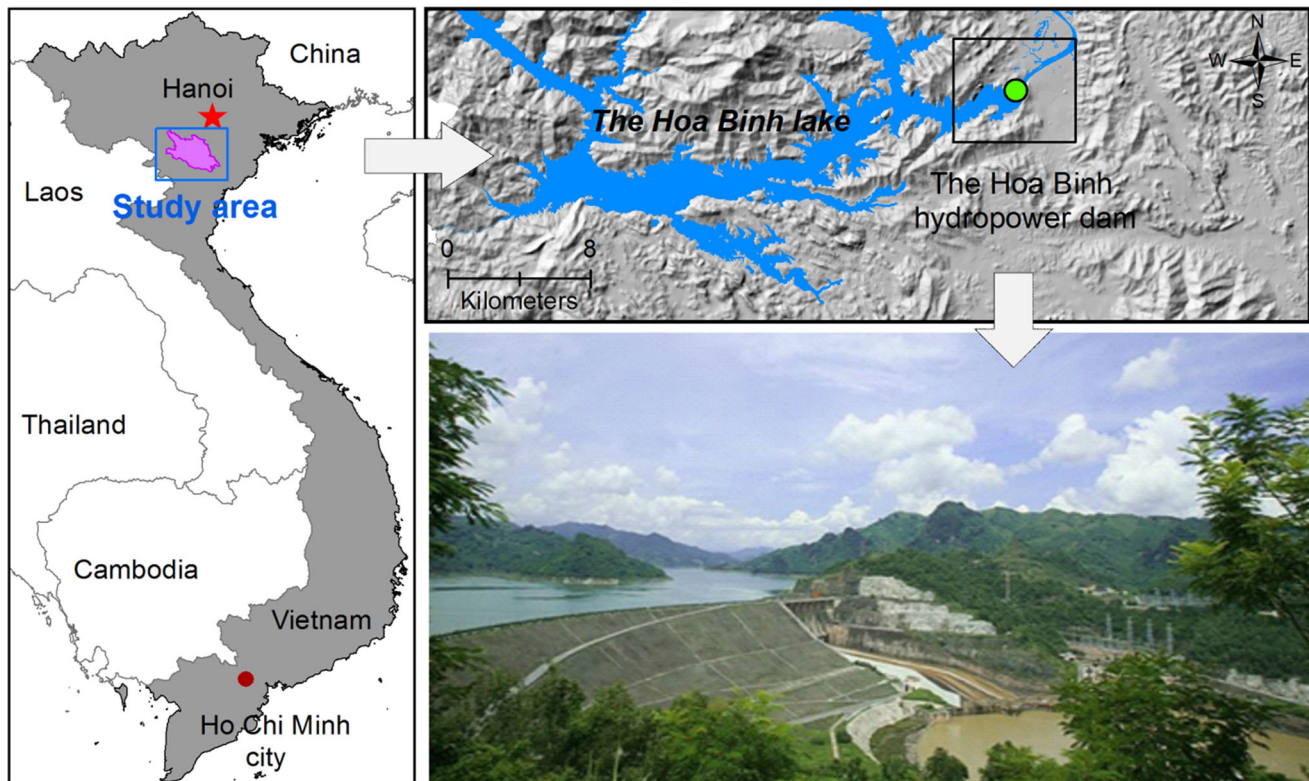$$MAE = \frac{\sum_{i=1}^{n} |pr_i - y_i|}{n} \tag{9}$$

Fig. 2 Location of the Hoa Binh hydropower dam

$$R^2 = \frac{S_{AP}}{\sqrt{S_A S_P}}; \; S_{AP} = \frac{\sum_{i=1}^{n} (pr_i - \overline{pr})(y_i - \overline{y})}{n-1}; \; S_A$$
$$= \frac{\sum_{i=1}^{n} (pr_i - \overline{pr})^2}{(n-1)}; \; S_P = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{(n-1)} \quad (10)$$

where y_i and $\overline{y}$ are the measure and the mean values of the horizontal displacement, respectively; pr_i and $\overline{pr}$ are predicted values and mean predicted value from the model, respectively.

3 Study area and data used

3.1 General description of the Hoa Binh hydropower dam

The Hoa Binh hydropower dam (Fig. 2) is situated on the Da River (a tributary of the Red River) in the Hoa Binh Province, about 75 km west of Hanoi city. This is an earth–rock fill dam that was constructed in an arch where the slopes of the banks are from 20° to 40°. The construction of the dam was started in 1981 and was completed in 1990 [35]. The total length of the dam is 970 m, and the maximum height of the dam crest is 128 m. Other characteristics of the dam are shown in Table 1.

Table 1 Characteristics of the Hoa Binh hydropower dam

No	Characteristics	Value (m)
1	The top height of the dam crest	128.0
2	Length of the dam crest	600.0
3	Width of the dam crest	20.0
4	Width of the core foundation	50.0
5	Width of the core crest	10.0

The Hoa Binh hydropower dam is the second largest dam in Vietnam with the installed capacity of 1920 MW. Currently, the hydropower produces around 10 billion kWh electricity per year. The total water can be held by the dam is around 9.45 billion m³. The minimum water level for the operation of the hydropower is 80 m whereas the maximum water level is 120 m [36]. Other detailed descriptions of the dam can be found in Vladimirov et al. [35].

3.2 Time series of monitoring data

The geodetic monitoring system of the Hoa Binh hydropower dam was established with 12 stations on the downstream side of the dam at the heights of 75 and 123 m. The aim is to monitor and assess horizontal and vertical
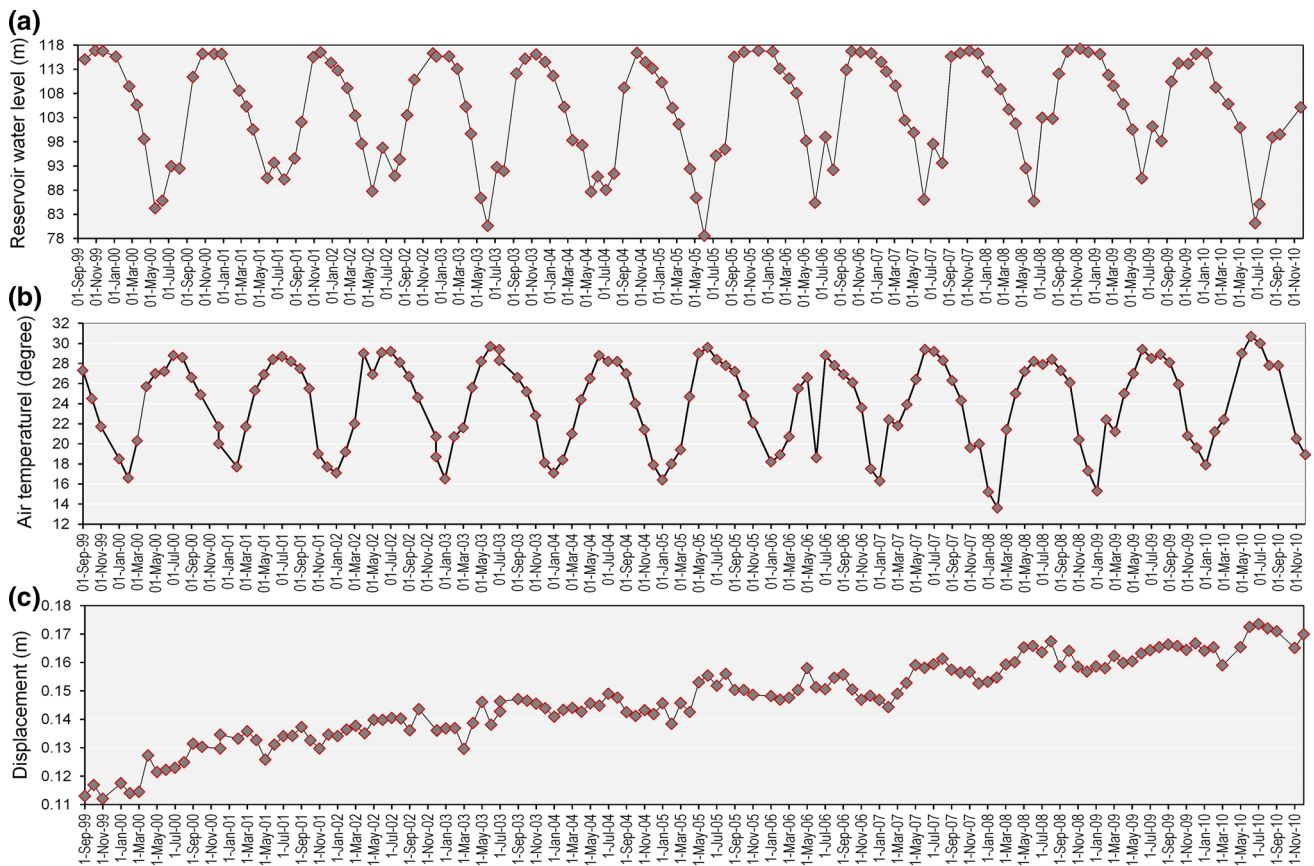
Fig. 3 Reservoir water level, air temperatures, and horizontal displacement recorded at the PVM12 station

displacements of the dam. Geodetic raw data were recorded in detailed hand-written reports for each epoch, including the date of measurement, horizontal and vertical deflections. In addition, other measured quantities were included such as rainfall, reservoir water level, and air temperature.

For this work, the available time series data of the PVM12 station are used for modeling. These data cover a period of 11 years (from 24/9/1999 to 14/12/2010 with 131 epochs) and consist of reservoir water level, air temperatures, and horizontal displacement (Fig. 3).

4 Proposed hybrid artificial intelligence approach for horizontal displacement modeling at the Hoa Binh hydropower dam (Vietnam)

This section describes the hybrid artificial intelligence approach based on a neural fuzzy inference system and PSO (SONFIS). SONFIS is used for the horizontal displacement modeling of the Hoa Binh hydropower dam. It is noted that the hybrid SONFIS model is programed by the authors in the Matlab$^{@}$2014 environment. The structure of the proposed hybrid SONFIS approach for horizontal displacement modeling is shown in Fig. 4. First, an

initial SONFIS model is generated for the study area using the training dataset. Then, PSO is adopted to optimize the model by searching the best parameter values of the premise and consequent parameters. Once the optimized parameters are found, the hybrid SONFIS regression model is derived and can be used to forecast dam displacements.

4.1 Data preparation

Horizontal displacements of hydropower dams have strong interactions with environmental factors (i.e., the temperature of the concrete), hydraulic factors (i.e., the reservoir water level), and aging; therefore, these factors should be taken into account for displacement analyses [3, 9, 10, 37]. In this study, reservoir water level (H), air temperature (T), and aging (t), measured at the Hoa Binh hydropower dam from 24/9/1999 to 14/12/2010 including 131 periods, were used.

The reservoir water level (H) is considered to be a reversible effect of the load and can be modeled in the form of a third-order polynomial with terms of H, H^2, and H^3 [10, 17], whereas the air temperature (T) influences horizontal displacements in the form of delayed actions [2]. Therefore, lagged variables should be used. In this study,
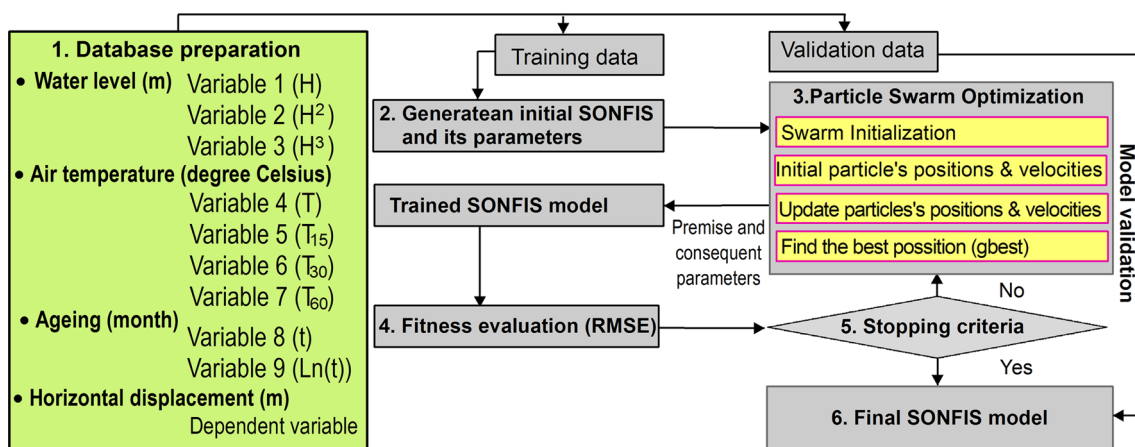
Fig. 4 Structure of the proposed SONFIS regression model for the horizontal displacement modeling of the Hoa Binh hydropower dam

Table 2 Descriptive statistics of the time series monitoring data (131 samples) in this study

Descriptive statistics	Reservoir water level (m)	Air temperature (°C)	Aging (month)	Horizontal displacement (m)
Lowest value	78.56	13.60	112.77	0.1124
Highest value	117.25	30.70	249.40	0.1739
Mean	104.43	23.99	180.79	0.1470
Median	105.62	25.00	180.43	0.1470
Standard deviation	10.71	4.32	39.19	0.0141

the air temperature (T) components used are (T, T_{15}, T_{30}, and T_{60}) where 15, 30, and 60 indicate number of days prior to the measurement. Aging (t) refers to the evolution of the hydropower dam over time; therefore, the horizontal displacement influenced by aging could be apportioned as t and $\ln(t)$ [9], where t is the cumulative number of months from the beginning date of the displacement measurement (20/6/1990). Consequently, a total of nine variables (H, H^2, H^3, T, T_{15}, T_{30}, T_{60}, t, $\ln(t)$) were used as input factors for the SONFIS model, whereas the horizontal displacement was the output.

The validation of predictive models is an important component in the analysis, and without this task, the models will be useless and have no scientific significance [38]. Therefore, the time series data were split into two subsets in a ratio of 70/30 [39], the first one is a training dataset that consists of 91 samples (from 24/9/1999 to 8/6/2007) and was used to train the SONFIS regression model whereas the second one with 40 samples (from 9/7/2007 to 14/12/2010) is a validation dataset that used to validate and confirm the forecasting accuracy of the model. Since the fuzzy membership values of the SONFIS model are in the range of [0, 1] [39], and the values of the nine input factors were rescaled into the above range. Descriptive statistics of the time series monitoring data in this study are shown in Table 2.

4.2 Model configuration

In this step, an initial SONFIS model is generated from the training dataset, the antecedent and the consequent parameters of the initial model are not optimized. Since the performance of neural fuzzy models can be enhanced if the training data are represented more concisely [39], the fuzzy c-means clustering [40] was used to transfer 91 samples into ten clusters. Ten clusters were selected based on a trial-and-error test between numbers of clusters vs. RMSE of the SONFIS model. Each cluster is used to generate a *fuzzy If–Then* rule of the model and the best values for the antecedent and the consequent parameters of the rule are obtained through the optimization process in the next step. It is noted that the Gaussian fuzzy membership function was used. The structure of the SONFIS model for this study is shown in Fig. 5, including nine input variables, one output, and ten rules.

4.3 Training the SONFIS model using PSO

The aim of this step is to find the best antecedent parameter and consequent parameters. When an initial swarm is generated, initial position and velocity of each particle in the swarm are determined. We used the inertia weight of 0.9 as stated by Poli et al. [41] due to
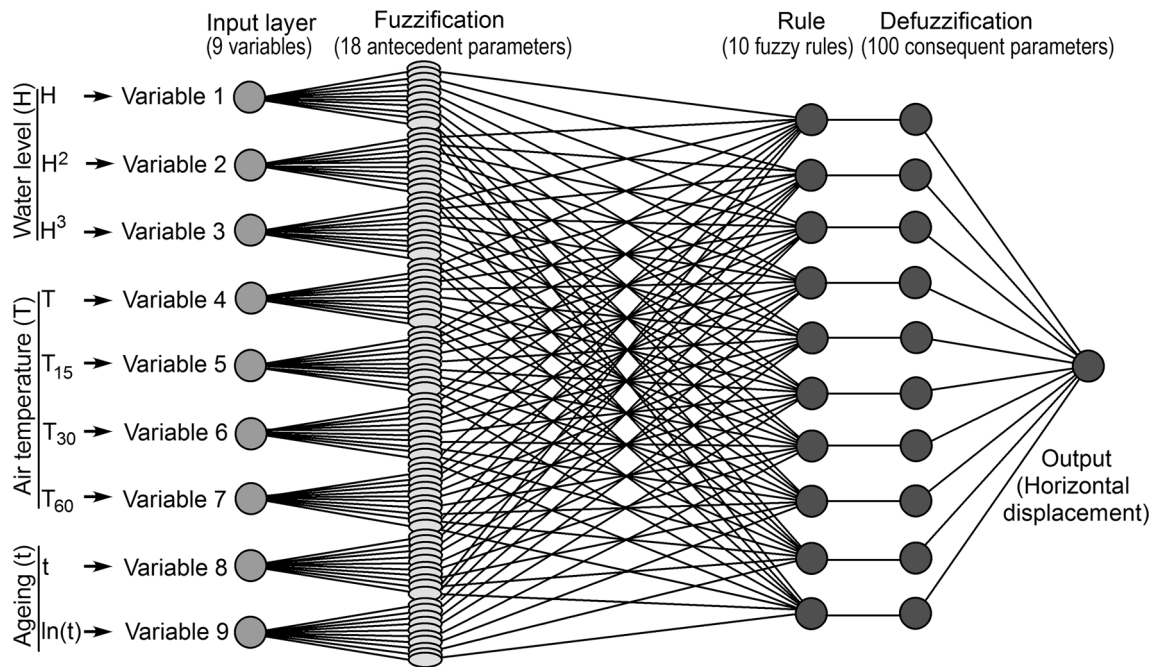
Fig. 5 SONFIS model for this study

the ability to get better performance. In the next step, all particles in the swarm are scattered and then their positions and their velocities are updated to find the best position. Accordingly, various combinations of antecedent and consequent parameters were explored by the PSO algorithm. For each iteration, the fitness evaluation for each particle was performed based on RMSE (Eq. 9). And then *pbest* for all particles were updated and compared to obtain the optimized position (*gbest*) of the swarm.

4.4 Stopping criteria

Maximum number of iterations is used as the stopping criteria and when the optimization process is terminated, the best position of the swarm is determined. With the best position, *RMSE* is the smallest; therefore, the optimized values for the antecedent and consequent parameters are determined for the SONFIS regression model. The final model is then validated using the validation dataset to confirm accuracy and then use for forecasting horizontal displacements of the dam.

5 Results and discussion

Using the Gaussian membership function, the SONFIS model was trained with 1000 epochs. The size of the swarm influences the diversity of the population; therefore, a trial-and-error test was used to determine the swarm size, and as result, a

population with 22 particles is the best for this study. The structure of the SONFIS model with ten rules is shown in Fig. 5. It could be seen that the model consists of 18 antecedent parameters and 100 consequence parameters (see Eqs. 2, 5) that have been optimized by the PSO algorithm. The optimized fuzzy membership curves obtained from the optimization process for this study are shown in Fig. 6, whereas the best values for the antecedent and consequent parameters are shown in Tables 3 and 4, respectively.

The training result (Table 5) shows that RMSEs of the SONFIS model are 0.306 cm in the training dataset and 0.294 cm in the validation dataset, respectively (Table 5). These values are significantly smaller than the standard deviation (1.249) that estimated from the measured displacement values indicating high performance of the model with these data. Since RMSE that shows overall information on the error distribution is a quadratic scoring index, therefore, RMSE is sensitive to some few large errors and outliers [42]. For this reason, MAE is a better choice [1] and is used interchangeably for estimating the model error in this study.

In addition, the difference between RMSE and MAE could be used to diagnose the variation of the model errors. The results (Table 5) show that MAEs of the SONFIS model are 0.238 cm and 0.237 cm for the training dataset and the validation dataset, respectively, indicating that the SONFIS model performs well. The difference of MAEs and RMSEs of the SONFIS model on the two datasets is 0.68 mm and 0.57 mm indicating a low variance of individual errors. R^2 for the training dataset is 0.960 and for the
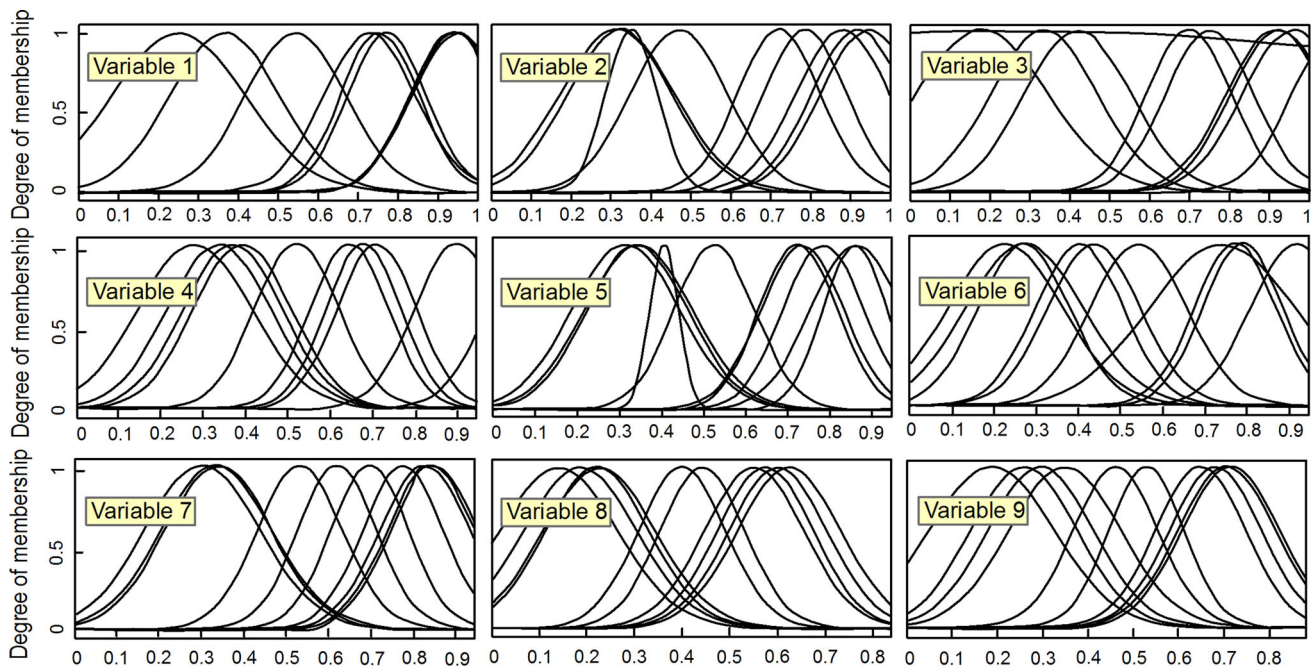
Fig. 6 Optimized fuzzy membership curves obtained from the PSO for this study

Table 3 Optimized fuzzy parameters for the nine input variables in the SONFIS model

Input variable	Gaussian fuzzy membership function parameter	
	δ	c
Variable 1 (H)	0.16479682	0.05420686
Variable 2 (H^2)	0.11341594	0.94604036
Variable 3 (H^3)	0.11083054	0.93398094
Variable 4 (T)	0.13378548	0.35756191
Variable 5 (T_{15})	0.10854647	0.54531438
Variable 6 (T_{30})	0.11002275	0.53855110
Variable 7 (T_{60})	0.10155999	0.85528916
Variable 8 ($t/100$)	0.11591654	0.95298775
Variable 9 ($\ln(t/100)$)	0.09753939	0.35592464

validation dataset is 0.869 indicating satisfactory results. Horizontal displacement and residual plots of this study are shown in Fig. 7.

The usability of the SONFIS model was further assessed through a comparison with those obtained from benchmark methods using the same data: Support vector regression (SVR) and MLP Neural Nets, and Random forest. These methods are selected because they outperformed conventional methods in various dam deformation studies [2, 4, 12, 15–17]. In addition, several state-of-the-art soft computing methods that have seldom been used for dam deformation analysis are included, such as the Gaussian

Processes, and the neural fuzzy model with Different evolution optimization (DE-FIS).

Since explanations of SVR, MLP Neural Nets, and Random forest algorithms for dam behavior modeling have been well documented, such as in Salazar et al. [2], therefore we only address briefly how these methods were used in this study. For the case of SVR, the radial basis kernel (RBF) function and the Sequential minimal optimization (SMO) algorithm were selected to solve the quadratic optimization problem [43]. The best parameters (the regularization = 6.6 and the kernel width = 0.0283) were found using the grid search method [44–46]. For the case of MLP Neural Nets, the structure with nine input neurons, one hidden layer with two neurons, and an output layer was determined using the method suggested in Tien Bui et al. [47]. Accordingly, the logistic sigmoid was selected as the activation function whereas the learning rate, the momentum, and the training epoch were used as 0.3, 0.2, and 500, respectively.

Regarding the Random forest, the model was built with 500 trees as suggested in Beck et al. [48]. Regarding the Gaussian processes, the method has been recently reported to be a powerful regression tool [49], and for the Hoa Binh dam displacement analysis, the RBF function was used, with the best kernel parameter that was found of 0.0056. Detailed explanations of the application of the Gaussian processes can be seen in Grbić et al. [50]. For the case of the DE-FIS, the same neural fuzzy model as in SONFIS was used, but the antecedent and consequent parameters were optimized using the DE technique.

Table 4 Optimized consequent parameters of ten rules of the SONFIS model

Rule	Output membership function parameters									
	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
1	51.97314	0.92915	−0.48175	−554.25736	1033.97950	−554.25973	0.00041	−0.10328	0.13884	−0.45023
2	51.97321	0.92942	−0.48176	−554.25764	1033.97260	−554.25600	0.00041	−0.10328	0.13884	−0.45023
3	51.97330	0.92946	−0.48176	−554.25984	1033.97488	−554.25591	0.00041	−0.10328	0.13884	−0.45022
4	51.97333	0.92946	−0.48176	−554.25957	1033.97236	−554.25600	0.00041	−0.10332	0.13884	−0.45023
5	51.97170	0.92946	−0.48176	−554.25712	1033.97503	−554.24352	0.00041	−0.10328	0.13246	−0.45023
6	51.96853	0.92945	−0.48176	−554.25276	1033.97128	−554.25596	0.00041	−0.10328	0.13884	−0.45022
7	51.97190	0.92946	−0.48036	−554.26661	1033.97153	−554.25222	0.00041	−0.10329	0.13884	−0.45023
8	51.97334	0.92946	−0.48175	−554.25893	1033.97240	−554.25372	0.00041	−0.10669	0.13884	−0.45020
9	51.97341	0.93114	−0.46637	−554.25838	1033.96748	−554.25604	0.00041	−0.10328	0.13884	−0.45019
10	51.97146	0.92946	−0.48175	−554.25552	1033.99201	−554.24706	0.00041	−0.10328	0.13884	−0.45557

Table 5 Performance comparison of the SONFIS model with the Support vector regression model, the MLP Neural Nets model, the Gaussian processes model, and Random forest model, and the DE-FIS model

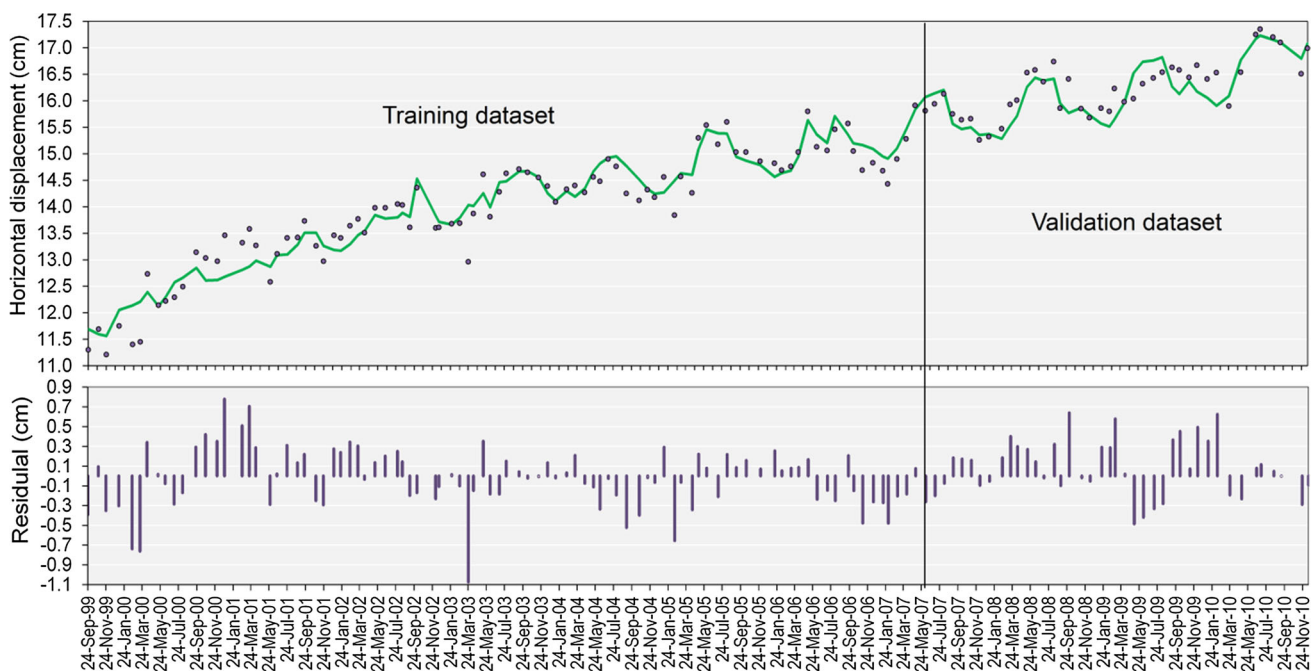| Regression model	Training dataset			Validation dataset		
	RMSE (cm)	MAE (cm)	R^2	RMSE (cm)	MAE (cm)	R^2
SONFIS	0.306	0.238	0.960	0.294	0.237	0.869
Support vector regression	0.404	0.292	0.939	0.528	0.443	0.826
MLP Neural Nets	0.596	0.516	0.925	0.616	0.563	0.875
Gaussian processes	0.419	0.328	0.927	0.365	0.302	0.883
Random forest	0.340	0.270	0.951	1.175	1.094	0.507
DE-FIS	0.377	0.301	0.939	0.323	0.258	0.849

Fig. 7 Measured and output horizontal displacement values derived from the SONFIS model using the training dataset and the validation dataset for this study

The performances of the five benchmark models on both the training and the validation datasets are shown in Table 5. It could be seen that RMSEs of the five models are higher than the SONFIS model on both the training dataset and the validation dataset (the SVR model (0.404, 0.528), the Gaussian processes model (0.419, 0.365), the MLP Neural Nets model (0.596, 0.606), the Random forest model (0.340, 1.175), and the DE-FIS model (0.377, 0.323). This is also the case for MAE (Table 5). The differences of RMSE and MAE representing the variance of individual errors show that the SONFIS model has the lowest one (RMSE-MAE = 0.068) in the training dataset, whereas in the validation dataset the variance is almost equal to the MLP Neural Nets model, but lower than the other models. Regarding to R^2, except the Random forests model, there is not much difference among the five models. Based on the above analysis, it could be concluded that the proposed SONFIS model performs better than the other models in this study.

6 Concluding remarks

This paper has proposed a novel hybrid artificial intelligence approach, namely SONFIS, for horizontal displacement modeling with a case study at the Hoa Binh hydropower dam (Vietnam). SONFIS is an integration of the PSO and the neural fuzzy inference system. The monitoring data with 131 periods cover a period of 11 years (1999–2010) and have been used to construct and validate the proposed model. The goodness-of-fit and prediction accuracy of the model were assessed using RMSE, MAE, and R^2. Five benchmark models (Support vector regression, MLP Neural Nets, Gaussian processes, Random forests, and DE-FIS) have been used for the comparison and confirmation of the usability of the proposed approach. The case study results show that the SONFIS is capable of providing prediction results with high accuracy. In other words, the accuracy of horizontal modeling of hydropower dams can be improved with the proposed method.

It is well known that model performance is dependent on the selection of input variables and it is common that the selection is based on engineering judgement [1]. In this research, the reservoir water level, the air temperature, and the aging are the measured data at the Hoa Binh dam site and therefore used as input data. Water level, air temperature, and aging are widely used for horizontal displacement modeling of dams. As a result, the high performance of the proposed method indicates that the selection, processing, and coding of the input variables have been carried out successfully.

One of the critical problems in building prediction models based on machine learning is that the models are prone to overfitting, such as in Ranković et al. [51] where MAE and MSE in the validation dataset were much higher compared to those in the training dataset. In general, it is not easy to eliminate the overfitting. In this study, both RMSE and MAE values of the proposed SONFIS model are almost equal in the training and validation datasets indicating that the overfitting is eliminated. On contrast, the Random forests model presents some degree of overfitting due to a large difference of MAE in the training and the validation data (Table 5). The reason is that the prediction of the Random forests model is made from the weighted average [52, 53] of the observed displacements in the training data, and as result, the prediction values were in the range of the observed displacements. In other words, the Random forests model has difficulties in extrapolation of values outside its known values.

Horizontal displacement modeling of hydropower dams is a real-world problem where time series monitoring data with sufficient length has seldom been available. Therefore, the size of the training and validation datasets should be considered properly. Although there is no thumb rule for defining the minimum amount of data, however, at least 5 years of data have been recommended as a minimum for training models in most cases in order to obtain high accuracy [2, 54, 55], whereas data of around 2–3 years of normal operations could be used for the model validation [56]. In this study, the time series spans around 7 years (from 24/9/1999 to 8/6/2007) were used to build models, whereas more than 3 years of data (9/7/2007 to 14/12/2010) were used for model validation. These indicate that the data for modeling in this study have a reasonable time span.

Overall, the major contributions of this study to the body of knowledge of dam displacement analysis are highlighted as follows: (1) the high performance of the SONFIS model, both on the training and validation datasets, implies that the SONFIS has successfully modeled a typical complex nonlinear problem of hydropower dam displacement; (2) the SONFIS model is constructed autonomously where the optimized antecedent and consequent parameters of the model were found autonomously with the use of the PSO algorithm; (3) overall, the SONFIS model outperforms the five benchmark methods; therefore, the proposed approach is a promising tool that could be an alternative method for modeling of horizontal displacements of hydropower dams.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

1. Salazar F, Morán R, Toledo M, Oñate E (2015) Data-based models for the prediction of dam behaviour: a review and some methodological considerations. Arch Comput Methods Eng. doi:10.1007/s11831-015-9157-9

2. Salazar F, Toledo MA, Oñate E, Morán R (2015) An empirical comparison of machine learning techniques for dam behaviour modelling. Struct Saf 56:9–17. doi:10.1016/j.strusafe.2015.05.001

3. De Sortis A, Paoliani P (2007) Statistical analysis and structural identification in concrete dam monitoring. Eng Struct 29(1):110–120

4. Mata J (2011) Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. Eng Struct 33(3):903–910. doi:10.1016/j.engstruct.2010.12.011

5. Bayrak T (2007) Modelling the relationship between water level and vertical displacements on the Yamula Dam, Turkey. Nat Hazards Earth Syst Sci 7(2):289–297

6. Areias P, Belytschko T (2005) Analysis of three-dimensional crack initiation and propagation using the extended finite element method. Int J Numer Meth Eng 63(5):760–788

7. Antes H, Von Estorff O (1987) Analysis of absorption effects on the dynamic response of dam reservoir systems by boundary element methods. Earthq Eng Struct Dyn 15(8):1023–1036

8. Vanatwerp R (1994) Engineering and design: deformation monitoring and control surveying. Engineer manual—US Army corps of engineering EM:1110-1111

9. Stojanovic B, Milivojevic M, Ivanovic M, Milivojevic N, Divac D (2013) Adaptive system for dam behavior modeling based on linear regression and genetic algorithms. Adv Eng Softw 65:182–190. doi:10.1016/j.advengsoft.2013.06.019

10. Xu C, Yue D, Deng C (2012) Hybrid GA/SIMPLS as alternative regression model in dam deformation analysis. Eng Appl Artif Intell 25(3):468–475. doi:10.1016/j.engappai.2011.09.020

11. Seyedpoor S, Salajegheh J, Salajegheh E, Gholizadeh S (2009) Optimum shape design of arch dams for earthquake loading using a fuzzy inference system and wavelet neural networks. Eng Optim 41(5):473–493

12. Karimi I, Khaji N, Ahmadi M, Mirzayee M (2010) System identification of concrete gravity dams using artificial neural networks based on a hybrid finite element–boundary element approach. Eng Struct 32(11):3583–3591

13. Ranković V, Grujović N, Divac D, Milivojević N, Novaković A (2012) Modelling of dam behaviour based on neuro-fuzzy identification. Eng Struct 35:107–113

14. Kao CY, Loh CH (2013) Monitoring of long-term static deformation data of Fei-Tsui arch dam using artificial neural network-based approaches. Struct Control Health Monit 20(3):282–303

15. Zheng D, Cheng L, Bao T, Lv B (2013) Integrated parameter inversion analysis method of a CFRD based on multi-output support vector machines and the clonal selection algorithm. Comput Geotech 47:68–77. doi:10.1016/j.compgeo.2012.07.006

16. Ranković V, Grujović N, Divac D, Milivojević N (2014) Development of support vector regression identification model for prediction of dam structural behaviour. Struct Saf 48:33–39

17. Su H, Wen Z, Sun X, Li H (2015) Rough set-support vector machine-based real-time monitoring model of safety status during dangerous dam reinforcement. Int J Damage Mech. doi:10.1177/1056789515616448

18. Salazar F, Toledo MÁ, Oñate E, Suárez B (2016) Interpretation of dam deformation and leakage with boosted regression trees. Eng Struct 119:230–251. doi:10.1016/j.engstruct.2016.04.012

19. Tien Bui D, Nguyen Q-P, Hoang N-D, Klempe H (2016) A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS. Landslides. doi:10.1007/s10346-016-0708-4

20. Hoang N-D, Tien Bui D (2016) A novel relevance vector machine classifier with cuckoo search optimization for spatial prediction of landslides. J Comput Civ Eng. doi:10.1061/(ASCE)CP.1943-5487.0000557

21. Hoang N-D, Tien Bui D, Liao K-W (2016) Groutability estimation of grouting processes with cement grouts using differential flower pollination optimized support vector machine. Appl Soft Comput 45:173–186. doi:10.1016/j.asoc.2016.04.031

22. Price K, Storn RM, Lampinen JA (2006) Differential evolution: a practical approach to global optimization. Springer, New York

23. Tien Bui D, Pham TB, Nguyen Q-P, Hoang N-D (2016) Spatial Prediction of rainfall-induced shallow landslides using hybrid integration approach of least squares support vector machines and differential evolution optimization: a case study in Central Vietnam. Int J Digit Earth. doi:10.1080/17538947.20161169561

24. Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. Mach Learn 3(2):95–99

25. Tien Bui D, Pradhan B, Nampak H, Quang Bui T, Tran Q-A, Nguyen QP (2016) Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modelling in a high-frequency tropical cyclone area using GIS. J Hydrol 540:317–330. doi:10.1016/j.jhydrol.2016.06.027

26. Ingber L (1993) Simulated annealing: practice versus theory. Math Comput Model 18(11):29–57

27. Parpinelli RS, Lopes HS, Freitas A (2002) Data mining with an ant colony optimization algorithm. Evol Comput IEEE Trans 6(4):321–332

28. Tien Bui D, Anh Tuan T, Hoang N-D, Quoc Thanh N, Nguyen BD, Van Liem N, Pradhan B (2016) Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a novel hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides. doi:10.1007/s10346-016-0711-9

29. Karaboga D, Gorkemli B, Ozturk C, Karaboga N (2014) A comprehensive survey: artificial bee colony (ABC) algorithm and applications. Artif Intell Rev 42(1):21–57

30. Kennedy J, Eberhart R (1995) Proceedings of IEEE international conference on neural networks. Perth, Australia

31. Song S, Kong L, Gan Y, Su R (2008) Hybrid particle swarm cooperative optimization algorithm and its application to MBC in alumina production. Prog Nat Sci 18(11):1423–1428. doi:10.1016/j.pnsc.2008.04.008

32. Voglis C, Parsopoulos KE, Papageorgiou DG, Lagaris IE, Vrahatis MN (2012) Mempsode: A global optimization software based on hybridization of population-based algorithms and local searches. Comput Phys Commun 183(5):1139–1154

33. Jang JSR (1993) ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685

34. Tien Bui D, Bui Q-T, Nguyen Q-P, Pradhan B, Nampak H, Trinh PT (2016) A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. doi:10.1016/j.agrformet.2016.11.002

35. Vladimirov VB, Zaretskii YK, Orekhov VB (2003) A mathematical model for monitoring the rock-earthen dam of the Hoa Binh hydraulic power system. Power Technol Eng 37(3):161–166. doi:10.1023/A:1025682101823

36. Nguyen TT, Pham VD, Tenhunen J (2013) Linking regional land use and payments for forest hydrological services: a case study of Hoa Binh Reservoir in Vietnam. Land Use Policy 33:130–140. doi:10.1016/j.landusepol.2012.12.015

37. Oro S, Mafioleti T, Chaves Neto A, Garcia S, Neumann Júnior C (2016) Study of the influence of temperature and water level of the reservoir about the displacement of a concrete dam. Int J Appl Mech Eng 21(1):107–120

38. Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick O (2013) Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. Nat Hazards 66(2):707–730

39. Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. Comput Geosci 45:199–211. doi:10.1016/j.cageo.2011.10.031

40. Abdulshahed AM, Longstaff AP, Fletcher S (2015) The application of ANFIS prediction models for thermal error compensation on CNC machine tools. Appl Soft Comput 27:158–168. doi:10.1016/j.asoc.2014.11.012

41. Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. Swarm Intell 1(1):33–57

42. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci Model Dev 7(3):1247–1250

43. Were K, Tien Bui D, Dick ØB, Singh BR (2015) A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol Indic 52:394–403

44. Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in Vietnam using Support vector machines, Decision tree and Naïve Bayes models. Math Probl Eng 2012:1–26

45. Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Application of support vector machines in landslide susceptibility assessment for the Hoa Binh province (Vietnam) with kernel functions analysis. In: Seppelt R, Voinov AA, Lange S, Bankamp D (eds) Proceedings of the iEMSs sixth biennial meeting: international congress on environmental modelling and software (iEMSs 2012). International Environmental Modelling and Software Society, Leipzig

46. Hong H, Pradhan B, Xu C, Tien Bui D (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. CATENA 133:266–281. doi:10.1016/j.catena.2015.05.019

47. Tien Bui D, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13:361–378. doi:10.1007/s10346-015-0557-6

48. Beck PS, Goetz SJ, Mack MC, Alexander HD, Jin Y, Randerson JT, Loranty M (2011) The impacts and implications of an intensifying fire regime on Alaskan boreal forest composition and albedo. Glob Change Biol 17(9):2853–2866

49. Verrelst J, Muñoz J, Alonso L, Delegido J, Rivera JP, Camps-Valls G, Moreno J (2012) Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. Remote Sens Environ 118:127–139. doi:10.1016/j.rse.2011.11.002

50. Grbić R, Kurtagić D, Slišković D (2013) Stream water temperature prediction based on Gaussian process regression. Expert Syst Appl 40(18):7407–7414

51. Ranković V, Novaković A, Grujović N, Divac D, Milivojević N (2014) Predicting piezometric water level in dams via artificial neural networks. Neural Comput Appl 24(5):1115–1121

52. Francke T, López-Tarazón J, Schröder B (2008) Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. Hydrol Process 22(25):4892–4904

53. Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H (2014) Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. Am J Epidemiol 179(6):764–774

54. Chouinard L, Roy V (2006) Performance of statistical models for dam monitoring data. Paper presented at the joint international conference on computing and decision making in civil and building engineering, Montreal

55. Swiss Committee on Dams (2003) Methods of analysis for the prediction and the verification of dam behaviour. Tech. rep, ICOLD

56. International Commission on Large Dams (2012) Dam surveillance guide. Tech. Rep. B-158, ICOLD