

## Article

# An Improved Autoencoder and Partial Least Squares Regression-Based Extreme Learning Machine Model for Pump Turbine Characteristics

Chu Zhang <sup>1,2</sup>, Tian Peng <sup>1,\*</sup>, Jianzhong Zhou <sup>2</sup>, Jie Ji <sup>1</sup> and Xiaolu Wang <sup>3</sup>

<sup>1</sup> College of Automation, Huaiyin Institute of Technology, Huaian 223003, China; zhangchuhust@foxmail.com (C.Z.); j.ji2@newcastle.ac.uk (J.J.)

<sup>2</sup> School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; jz.zhou@hust.edu.cn

<sup>3</sup> Division of Informatization Construction & Management, Huaiyin Institute of Technology, Huaian 223003, China; wangxiaolu@hyit.edu.cn

\* Correspondence: husthydropt@126.com; Tel.: +86-182-7140-1257

Received: 20 August 2019; Accepted: 20 September 2019; Published: 23 September 2019



**Featured Application:** Authors are encouraged to provide a concise description of the specific application or a potential application of the work. This section is not mandatory.

**Abstract:** Complete characteristic curves of a pump turbine are fundamental for improving the modeling accuracy of the pump turbine in a pump turbine governing system. In view of the difficulty in modeling the “S” characteristic region of the complete characteristic curves in the pump turbine, a novel Autoencoder and partial least squares regression based extreme learning machine model (AE-PLS-ELM) was proposed to describe the pump turbine characteristics. First, a mathematical model was formulated to describe the flow and moment characteristic curves. The improved Suter transformation was employed to transfer the original curves into WH and WM curves. Second, the ELM-Autoencoder technique and the partial least squares regression (PLSR) method were introduced to the architecture of the original ELM network. The ELM-Autoencoder technique was employed to obtain the initial weights of the Autoencoder based extreme learning machine (AE-ELM) model. The PLS method was exploited to avoid the multicollinearity problem of the Moore-Penrose generalized inverse. Lastly, the effectiveness of the proposed AE-PLS-ELM model has been verified using real data from a pumped storage unit in China. The results demonstrated that the AE-PLS-ELM model can obtain better modeling accuracy and generalization performance than the traditional models and, thus, can be exploited as an effective and sufficient approach for the modeling of pump turbine characteristics.

**Keywords:** pump turbine; complete characteristic curves; Autoencoder; partial least squares regression; extreme learning machine

## 1. Introduction

As the demand for electricity and the requirements for developing a low-carbon economy continues, the driving force for energy development will gradually shift to renewable and clean energy such as photovoltaic power and wind power [1–3]. Due to its fast start-stop speed and flexible working conditions, the pumped storage units (PSUs) can quickly respond to the requirements of the power system for frequency and phase modulation, peak load shifting, rotation, and accident reserve, which can enhance the grid’s ability to absorb wind power and photoelectricity [4,5]. The pump turbine governing system (PTGS) is a complex hydraulic-mechanical-electrical-magnetic coupling system,

which is time-varying, stochastic, and nonlinear [6,7]. The precise modeling of the PTGS is of great significance in analyzing the dynamic response of PSUs under complex operational conditions [8,9].

As a crucial part of PSU, an accurate pump turbine model is the key to the accurate modeling and simulation of PTGS [10]. In the current research work, the modeling of a pump turbine is mainly based on complete characteristic curves where different operational conditions of this model can be described [11–13]. The complete curves of pump turbine include the flow characteristic curve and the moment characteristic curve. These curves contain the inherent nonlinear characteristics of the pump turbine under various normal operational conditions or transient working conditions [14,15]. However, due to the existence of the “S” characteristic and hump area, the estimation of complete characteristic curves is highly complex [16,17]. It is easy for the complete characteristic curves to gather, cross, and twist in the pumps and anti-pump regions, which cause a huge interpolation error and non-convergence problem to the model when calculating the transient parameters of the runner boundary of the pump turbine. An in-depth and extensive research study has been conducted to extend the complete characteristic curves in recent years. The common methods to estimate the complete curves of pumped turbine mainly include the assistant mesh processing [18], the Suter transformation, its improved versions [19,20], and the 3D surface fitting [14].

The assistant mesh processing method divides the flow characteristic curve or moment characteristic curve into multiple sections by introducing the idea of piecewise linearization, and then the original curve is approximated using small line segments. An assistant mesh line, which is approximately orthogonal to the opening line, is drawn to facilitate calculation. The characteristic curves of the pump turbine have been obtained and the modeling accuracy has been improved through the assistant mesh processing technique [8]. However, the above method requires prior artificial meshing of which the workload is heavy and the piecewise approximation also causes errors in the orthogonal curvilinear grid.

The Suter transformation proposed by Suter [19] transformed the original complete characteristic curves into  $WH(x, y)$  and  $WM(x, y)$  curves using dimensionless parameters. The abscissa of the curves is determined according to the relative flow and the relative speed. Because the Suter transformation stretches the flow and moment characteristic curves to both sides of the coordinate axis, the difficulties of the interpolation calculation caused by curve bending, torsion, anti-S, and multi-value characteristics can be alleviated effectively. However, scholars have found that the complete curves using Suter transformation or improved Suter transformation still have the following drawbacks: the length of the opening lines is different, some opening lines have singularities and are not conductive, and the density of data points on the opening lines is different. These non-uniform curves will affect the interpolation accuracy and computational efficiency of the model. Therefore, it is necessary to establish neural network models for curve extension, bad point correction, and encryption processing of opening lines for the complete characteristic curves.

According to whether the flow and moment characteristic curves are pre-transformed, the 3D surface fitting technique can be mainly divided into two categories: the first category is based on the original complete curves of the pump turbine. The original complete curves are fitted using the least squares or neural network models directly. The second category is based on the pre-processed curves using Suter transformation. The pre-processed curves are then fitted using the least squares or neural network models. Li et al. [14] applied a backpropagation (BP) neural network to process the synthetic characteristic curve for the Francis turbine. Actual engineering applications validate that the proposed method can provide not only higher precision data for the transition process but also better investigation on the real hydraulic unit. Zhang et al. [21] applied a radial basis function (RBF) to process the synthetic characteristic curve of the Kaplan turbine. The results have demonstrated that the nonlinear model based on RBF networks can reflect the nonlinear operating characteristics of the Kaplan turbine with higher accuracy. Liu et al. [20] first employed a modified Suter transformation to pre-process the complete characteristic curve and then proposed an Adaboost-BP neural network ensemble model optimized by particle swarm optimization to describe the WH and WM characteristics

of the pump turbine. The results show that the proposed model can obtain higher fitting accuracy and better generalization performance than a single BP neural network model.

Although the neural network model can fit and extend the complete characteristic curves and facilitate the calculation of the flow and moment characteristic curves for the pump turbine, the convergence speed of the traditional neural network model is slow and it is easy to fall into local minimum. As a novel single hidden layer feedforward neural network, the extreme learning machine (ELM) obtained the input weights and hidden layer biases through random initialization [22]. The output weights are directly obtained by calculating the generalized inverse matrix of the hidden layer output matrix [23,24]. The convergence rate is far faster than the traditional BP neural network. The ELM has been widely used in pattern recognition, statistical prediction, and classification and regression [25]. However, because of the random initialization strategy of input weights, the common ELM model fails to make full use of the inherent characteristics of the training data. In addition, the Moore-Penrose generalized inverse used in the ELM model may produce pathological solutions. Therefore, there exists multicollinearity in the output matrix, which affects the fitting and generalization performance of the model [26]. Therefore, there still exists much room to improve in describing pump turbine characteristics using ELM.

To improve the modeling accuracy of the pump turbine in the simulation of PTGS, an Autoencoder and partial least squares regression based extreme learning machine model (AE-PLS-ELM) is proposed to model the pumped turbine of a PSU. With the strong fitting ability of the AE-PLS-ELM model, the non-linear mapping relationship of the full characteristic curves of pump turbines is represented. The flow and moment characteristic curves of pump turbines are transformed into neural network models, which can be used for a real-time simulation. On the basis of curve pretreatment with improved Suter transformation, two AE-PLS-ELM models are used to model the characteristic curves. The automatic encoder technique (AE) and partial least squares regression algorithm (PLSR) are introduced to improve the performance of the ELM model. The rest of this paper is arranged as follows. Section 2 describes the model of pump turbine based on characteristic curves, Section 3 proposes an AE-PLS-ELM model, Section 4 provides the specific modeling process of the pump turbine characteristics based on the proposed AE-PLS-ELM model, Section 5 employs a numerical example to verify the performance of AE-PLS-ELM, Section 6 provided an additional test problem, and Section 7 gives the conclusions.

## 2. Nonlinear Modeling of the Pump Turbine

The most common method for the pump turbine nonlinear modeling is through the complete characteristic curves [14]. The main idea of the nonlinear modeling based on complete curves is to first extract certain discrete data points from the practical curves, and then the extracted data points are fitted or extended to obtain the modeling curves [14]. The mathematical model of the flow and moment characteristics to express the pump turbine characteristics is as follows [2].

$$\begin{cases} M_{11} = f_M(a, N_{11}) \\ Q_{11} = f_Q(a, N_{11}) \end{cases} \quad (1)$$

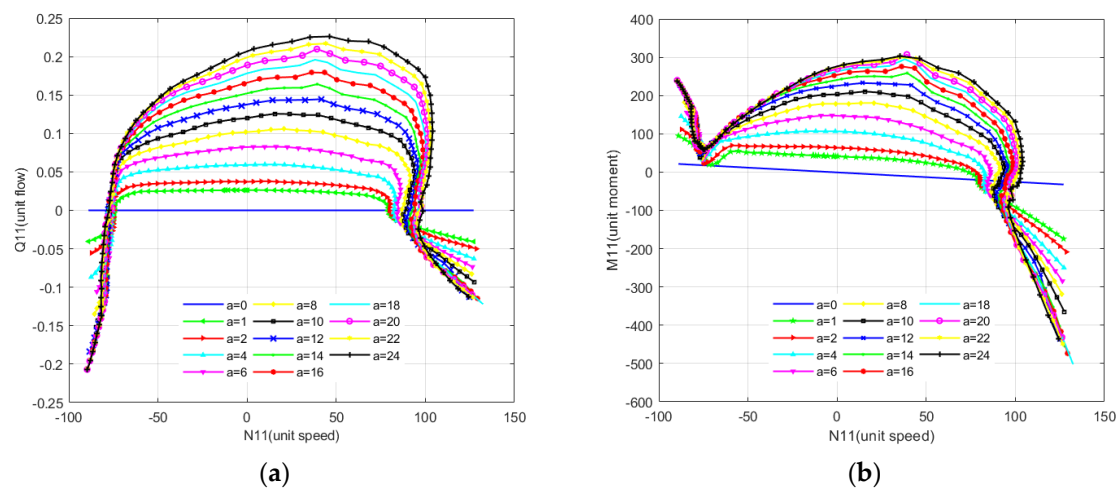
where  $M_{11}$  represents the unit moment,  $Q_{11}$  denotes the unit flow,  $N_{11}$  is the unit speed, and  $a$  denotes the guide vane opening. In this study, the pump turbine complete characteristic curves of a hydropower station in China are employed as a case study. The practical complete characteristic curves are shown in Figure 1 as follows [2].

As can be seen in Figure 1, the original completed curves still have a significant twist curl when the unit speed is bigger than 80, which causes the multi-value phenomenon in the “S” characteristic area. For example, three flow (Figure 1a) and moment (Figure 1b) data points appear when the value of unit speed is 90 r/min and the value of the guide vane opening is 10. The interpolation error is large and the derivative is discontinuous when multiple values exist, which may cause an iteration error to occur in the PTGS. To avoid the multi-value phenomenon of the original curves, this study introduces

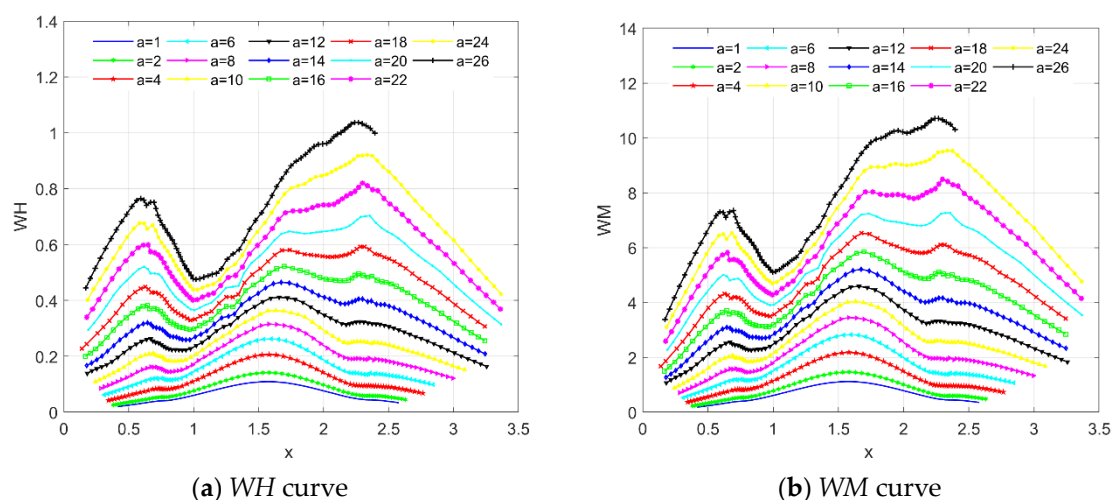
the improved Suter transformation [20] to pre-process the original complete curves. The original flow and moment characteristic curves are changed to WH and WM curves, respectively, through the improved Suter transformation. The converting equations of the improved Suter transformation are expressed as follows.

$$\begin{cases} WH(x, y) = \frac{h}{a^2 + q^2 + C_h \cdot h} (y + C_y)^2 \\ WM(x, y) = \frac{(m + s_2 h)}{a^2 + q^2 + C_h \cdot h} (y + C_y)^2 \\ x = \arctan [(q + s_1 \sqrt{h}) / a] \quad a \geq 0 \\ x = \pi + \arctan [(q + s_1 \sqrt{h}) / a] \quad a < 0 \end{cases} \quad (2)$$

where  $a$ ,  $q$ ,  $h$ , and  $m$  denote the relative speed, relative flow, relative water head, and relative moment, respectively,  $x$  and  $y$  denote the relative flow angle and relative opening, respectively.  $s_2 > |M_{11\max}|/M_{11r}$ ,  $s_1 = 0.5 \sim 1.2$ ,  $C_y = 0.1 \sim 0.3$ , and  $C_h = 0.4 \sim 0.6$ . The  $WH(x, y)$  and  $WM(x, y)$  curves based on an improved Suter transformation are given in Figure 2 as follows [2].



**Figure 1.** The characteristic curves of a pump turbine under different openings. (a) Flow characteristic curve and (b) moment characteristic curve.



**Figure 2.** WH and WM characteristic curves based on improved Suter transformation under different relative speed.

### 3. An Autoencoder and Partial Least Squares Regression Based Extreme Learning Machine Model

An accurate pump-turbine model is the key to the modeling and simulation of PTGS. In this study, an autoencoder and partial least squares regression-based extreme learning machine model (AE-PLS-ELM) is proposed for the nonlinear modeling of pump turbine characteristics. The AE-PLS-ELM model is introduced in this section.

#### 3.1. Extreme Learning Machine

ELM is a new type of single hidden layer forward neural network. It randomly generates the connection weights and biases between the input layer and the hidden layer. The connection weights and biases do not need to be adjusted during the training process. Once the number of hidden neurons is determined, the optimal solution can be obtained. Compared with the traditional training methods, ELM has the advantages of fast learning speed and good generalization performance [27].

Suppose that there are  $N$  samples  $(x_k, y_k), k = 1, 2, \dots, N$ , where  $x_k \in \mathbb{R}^p$  and  $y_k \in \mathbb{R}^q$ . The mathematical expression for an ELM model with  $L$  hidden neurons is as follows.

$$f_L = \sum_{i=1}^L \beta_i g(a_i \cdot x_k + b_i) = \hat{y}_k \quad k = 1, 2, \dots, N \quad (3)$$

where  $\hat{y}_k$  denotes the simulated output of the  $k$ th sample,  $a_i$  and  $\beta_i$  are the connection weights between the  $i$ th hidden neuron and the input layer and hidden layer, respectively,  $b_i$  is the bias of the  $i$ th hidden neuron, and  $g(\cdot)$  is the activation function. Equation (3) can be reformulated below.

$$H\beta = \hat{Y} \quad (4)$$

where

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g(a_1 \cdot x_1 + b_1) & \cdots & g(a_L \cdot x_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(a_1 \cdot x_N + b_1) & \cdots & g(a_L \cdot x_N + b_L) \end{bmatrix}_{N \times L} \quad (5)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m} \quad (6)$$

where  $H$  is known as the hidden layer output matrix, and  $\beta$  is the output weight matrix.

Based on the Moore-Penrose generalized inverse matrix theory,  $\beta$  can be calculated as:

$$\hat{\beta} = H^\dagger Y = (H^T H)^{-1} H^T Y \quad (7)$$

where  $H^\dagger$  is the generalized inverse matrix of  $H$ . To improve the stability and the generalization of the ELM network, Huang et al. [28] added a positive constant  $1/C$  to the diagonal of  $H^T H$  (or  $HH^T$ ).  $\beta$  can then be calculated as:

$$\beta^* = \left( \frac{1}{C} + H^T H \right)^{-1} H^T Y \quad (8)$$

#### 3.2. The ELM-Autoencoder Technique

The convergence speed of ELM is fast and the generalization capability is well compared with the traditional BP neural networks. However, the initial parameters of ELM are independent of the modeling data. Thus, the characteristics and internal relations of the modeling data cannot be effectively reflected. To obtain better initial parameters of ELM, the Autoencoder technology, which has been widely employed in deep learning, is introduced to ELM for modeling [29]. The traditional

Autoencoder technology developed by Rumelhart et al. in 1986 is an unsupervised learning method based on a BP algorithm. The purpose of the Autoencoder technology is to approximate an identity function such that the output data is the same as the input data [30]. The ELM-Autoencoder technology based on the ELM algorithm is introduced in this study to avoid repeated iterative training of the BP network [31]. The Autoencoder based extreme learning machine (AE-ELM) proposed in this study can be implemented mainly in two stages. First, the ELM-Autoencoder technology is employed to establish the mapping relationship between  $X$  to  $X$  ( $X$  is the input data) using the ELM algorithm. The output weights of the ELM-Autoencoder is taken as the initial weights of the AE-ELM model. Second, the AE-ELM model, which takes  $X$  and  $Y$  ( $Y$  is the output data) as training data, is trained with the initial weights taken from the first stage.

The ELM-Autoencoder is a type of unsupervised network and the input weights and hidden biases should be orthogonal. Assume the number of input neurons is  $N_p$ , and the number of the hidden neurons is  $N_L$ . In this study, the number of input neurons is smaller than the number of hidden neurons. Therefore, a sparse ELM-Autoencoder architecture is adopted [32]. Given a set of  $N$  data samples, i.e.,  $(x_k, y_k)$  for  $k = 1, 2, \dots, N$ , the hidden layer output of the ELM-Autoencoder can be expressed as:

$$h(x_i) = g(a^T x_i + b) \quad i = 1, 2, \dots, N \quad (9)$$

where  $h(x_i) \in R^{N_L}$  denotes the hidden layer output with respect to the  $i$ th input,  $a^T a = I$ ,  $b^T b = I$ ,  $I$  is a unit matrix,  $a = [a_1, a_2, \dots, a_N]$  denotes the orthogonal weights connecting the input layer and the hidden layer,  $b = [b_1, b_2, \dots, b_N]$  denotes the orthogonal biases of the hidden nodes, and  $g(\cdot)$  is the activation function.

The output weight  $\beta$  is optimized to minimize the squared loss of the training error. The optimization problem can be expressed as the equation below.

$$\min_{\beta} O_{\beta} = \min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \|X - H\beta\|^2 \quad (10)$$

where  $X$  denotes the input data,  $H$  denotes the hidden layer outputs, and  $C$  denotes a penalty factor on the training error.

The first derivative of  $O_{\beta}$  with respect to  $\beta$  can be denoted as:

$$\Delta_{\beta} = \beta - CH^T(X - H\beta) = 0 \quad (11)$$

The final hidden layer output weights can be calculated as:

$$\beta = (I/C + H^T H)^{-1} H^T X \quad (12)$$

### 3.3. Partial Least Squares Regression

PLSR developed by Wold is a multivariate statistical analysis method that takes advantage of both principal component analysis and the least square method [33]. Compared with the least square method, the PLSR method can deal with the multicollinearity problem. The PLSR method allows for regression modeling when the number of samples is smaller than the number of independent variables. In addition, the PLSR model considers information of both the independent and responsible variables, which makes it easier to identify system information and noises.

Given a set of  $N$  observed data samples, which were composed of  $p$  input and  $q$  output variables, i.e.,  $S = (x_i, y_i)$  for  $i = 1, 2, \dots, N$ , where  $x_i \in R^p$  denotes the independent variable and  $y_k \in R^q$  denotes the responsible variable. The objective of PLSR is modeling the linear relationship between the  $p$  independent variables and  $q$  response variables. The modeling process of PLSR can be denoted as follows [26].



First, the independent matrix  $X = (x_1, x_2, \dots, x_p)_{n \times p}$  and response matrix  $Y = (y_1, y_2, \dots, y_q)_{n \times q}$  are normalized into a zero mean and one variance. The normalized matrix of  $X$  and  $Y$  are denoted as  $E_0$  and  $F_0$ , respectively. Second, the PLSR method is applied to  $X$  and  $Y$  to extract the first pair of score vectors  $u_1$  and  $v_1$ , respectively. The score vectors  $u_1$  and  $v_1$  are the combination of the independent variable and responsible variables, respectively, and should contain the maximum variation information of them. The relationship between  $u_1$  and  $v_1$  should be as maximum as possible. Then the regression model between  $Y$  and  $u_1$  is deduced. Third, the residual matrix  $E_1$  and  $F_1$  are calculated to replace  $E_0$  and  $F_0$ , respectively, to skip to the next step iteration until the residual matrix meets the stopping criteria.

### 3.4. The Proposed AE-PLS-ELM Model

In the Autoencoder based extreme learning machine (AE-ELM) model, the Moore-Penrose generalized inverse based on least square is exploited to calculate the output weights, which makes it possible to apply the PLSR method by replacing the least square method. The employment of PLSR in the AE-ELM model can avoid the multicollinearity problem when applying the Moore-Penrose generalized inverse especially when the hidden layer output matrix is highly correlated and contains noises [34]. Based on the description of AE-ELM and PLSR, the key to establish the AE-PLS-ELM model is to model the linear relationship between the hidden layer output matrix  $H$  and the output layer matrix  $Y$  using the partial least squares regression (PLSR) method. As has been described in Section 3.2, the hidden layer output matrix  $H$  is an  $N \times L$  dimension matrix with  $L$  hidden layer output variables. The output layer matrix  $Y$  is a  $N \times q$  dimension matrix. The detailed modeling process of the AE-PLS-ELM model can be given as follows.

Step 1: Randomly assign the input weights  $a_i, i = 1, 2, \dots, n$  and hidden layer bias  $b_i, i = 1, 2, \dots, L$ .

Step 2: Calculate the hidden layer output matrix of the training data  $H$ , according to Equation (3).

Step 3: The hidden layer output matrix  $H$  is taken as the independent matrix of the PLSR model.

Let  $E_0 = H$  and  $F_0 = Y$ .

Step 4: Extract the first pair of score vectors of  $E_0$  and  $F_0$ . The score vectors are denoted as  $u_1$  and  $v_1$ , respectively.  $u_1$  and  $v_1$  can be denoted below.

$$\begin{aligned} u_1 &= E_0 \omega_1 \\ v_1 &= F_0 c_1 \end{aligned} \quad (13)$$

where  $\omega_1 = [\omega_{11}, \omega_{12}, \dots, \omega_{1L}]^T$  is the load factor of  $E_0$  and  $\|\omega_1\| = 1$ , and  $c_1 = [c_{11}, c_{12}, \dots, c_{1L}]^T$  is the load factor of  $F_0$  and  $\|c_1\| = 1$ .

The relevance between  $u_1$  and  $F_0$  should be as maximum as possible. Then  $u_1$  and  $v_1$  should satisfy the following two criterions.

(1)  $u_1$  and  $v_1$  should contain the maximum variation information of  $E_0$  and  $F_0$ .

(2)  $u_1$  and  $v_1$  should have maximum relevance.

The deduction of  $\omega_1$  and  $c_1$  can be transformed to the following optimization problem.

$$\begin{aligned} \max (u_1, v_1) &= (E_0 \omega_1, F_0 c_1) = \omega_1^T E_0^T F_0 c_1 \\ \text{s.t. } \begin{cases} \omega_1^T \omega_1 = \|\omega_1\|^2 = 1 \\ c_1^T c_1 = \|c_1\|^2 = 1 \end{cases} \end{aligned} \quad (14)$$

According to the Lagrange multiplier method, the following equation is true.

$$L = \omega_1^T E_0^T F_0 c_1 - \lambda_1 (\omega_1^T \omega_1 - 1) - \lambda_2 (c_1^T c_1 - 1) \quad (15)$$

The partial derivatives of  $L$  on  $\omega_1, c_1, \lambda_1$ , and  $\lambda_2$  can be expressed by the equation below.

$$\begin{cases} \frac{\partial L}{\partial \omega_1} = E_0^T F_0 c_1 - 2\lambda_1 \omega_1 = 0 \\ \frac{\partial L}{\partial c_1} = F_0^T E_0 \omega_1 - 2\lambda_2 c_1 = 0 \\ \frac{\partial L}{\partial \lambda_1} = -(\omega_1^T \omega_1 - 1) = 0 \\ \frac{\partial L}{\partial \lambda_2} = -(c_1^T c_1 - 1) = 0 \end{cases} \quad (16)$$

From Equation (16), it can be deduced that:

$$2\lambda_1 = 2\lambda_2 = \omega_1^T E_0^T F_0 c_1 = \langle E_0 \omega_1, F_0 c_1 \rangle \quad (17)$$

Note that:

$$\theta_1 = 2\lambda_1 = 2\lambda_2 = \omega_1^T E_0^T F_0 c_1 \quad (18)$$

$\theta_1$  is the objective function of the optimization problem, and Equations (19) and (20) are true.

$$E_0^T F_0 c_1 = \theta_1 \omega_1 \quad (19)$$

$$F_0^T E_0 \omega_1 = \theta_1 c_1 \quad (20)$$

From Equation (19) and Equation (20), it can be deduced that:

$$E_0^T F_0 F_0^T E_0 \omega_1 = \theta_1^2 \omega_1 \quad (21)$$

where  $\omega_1$  is the eigenvector of  $E_0^T F_0 F_0^T E_0$ ,  $\theta_1^2$  is the eigenvalue, and  $\omega_1$  and  $c_1$  can be obtained according to Equation (19) and Equation (20). After obtaining  $\omega_1$  and  $c_1$ , the score vectors  $u_1$  and  $v_1$  can be calculated according to Equation (13).

Step 5: Establish the linear regression model between  $E_0$  and  $u_1$ , and  $F_0$  and  $v_1$  according to the least square method.

$$\begin{cases} E_0 = u_1 \alpha_1^T + E_1 \\ F_0 = v_1 \gamma_1^T + F_1 \end{cases} \quad (22)$$

where  $\alpha_1 = [\alpha_{11}, \alpha_{12}, \dots, \alpha_{1L}]$  and  $\gamma_1 = [\gamma_{11}, \gamma_{12}, \dots, \gamma_{1L}]$  are regression coefficients, and  $E_1$  and  $F_1$  are the residual matrix. The least square estimation of  $\alpha_1$  and  $\gamma_1$  can be denoted below.

$$\begin{cases} \alpha_1 = \frac{E_0^T u_1}{\|u_1\|^2}, \\ \gamma_1 = \frac{F_0^T v_1}{\|v_1\|^2}. \end{cases} \quad (23)$$

Step 6: If  $F_1$  satisfy the stopping criteria, Equation (22) is the final regression model and the iteration stops. Otherwise, replace  $E_0$  and  $F_0$  with  $E_1$  and  $F_1$ , respectively, and skip to Step 3 to get the second pair of score vectors.

$$\begin{aligned} u_2 &= E_1 \omega_2 \\ v_2 &= F_1 c_2 \end{aligned} \quad (24)$$

$$\begin{cases} E_0 = u_1 \alpha_1^T + u_2 \alpha_2^T + E_2 \\ F_0 = v_1 \gamma_1^T + v_2 \gamma_2^T + F_2 \end{cases} \quad (25)$$

where  $\alpha_2$  and  $\gamma_2$  are regression vectors and can be denoted as:

$$\begin{cases} \alpha_2 = \frac{E_1^T u_2}{\|u_2\|^2}, \\ \gamma_2 = \frac{F_1^T v_2}{\|v_2\|^2}. \end{cases} \quad (26)$$



Step 7: Repeat Steps 4–5 until  $r$  principal components are calculated. The remaining  $m - r$  components are small and are considered as noises. The residuals  $E_r$  and  $F_r$  are very small.  $E_0$  and  $F_0$  can be expressed by the equation below.

$$\begin{cases} E_0 = u_1\alpha_1^T + u_2\alpha_2^T + \dots + u_r\alpha_r^T + E_r = U\alpha^T + E_r, \\ F_0 = v_1\gamma_1^T + v_2\gamma_2^T + \dots + v_r\gamma_r^T + F_r = V\gamma^T + F_r \end{cases} \quad (27)$$

The relationship between  $u_k$  and  $v_k$  can be expressed by the equation below [1].

$$v_k = u_k b_1, k = 1, 2, \dots, r. \quad (28)$$

Then  $F_0$  can be translated into:

$$F_0 = V\gamma^T + F_r = \sum_{i=1}^r u_i b_1 \gamma_1^T + u_2 b_2 \gamma_2^T + \dots + u_r b_r \gamma_r^T + F_r = UB\gamma^T + F_r \quad (29)$$

where  $\hat{U} = E_0 W$ , then the regression equation can be denoted as:

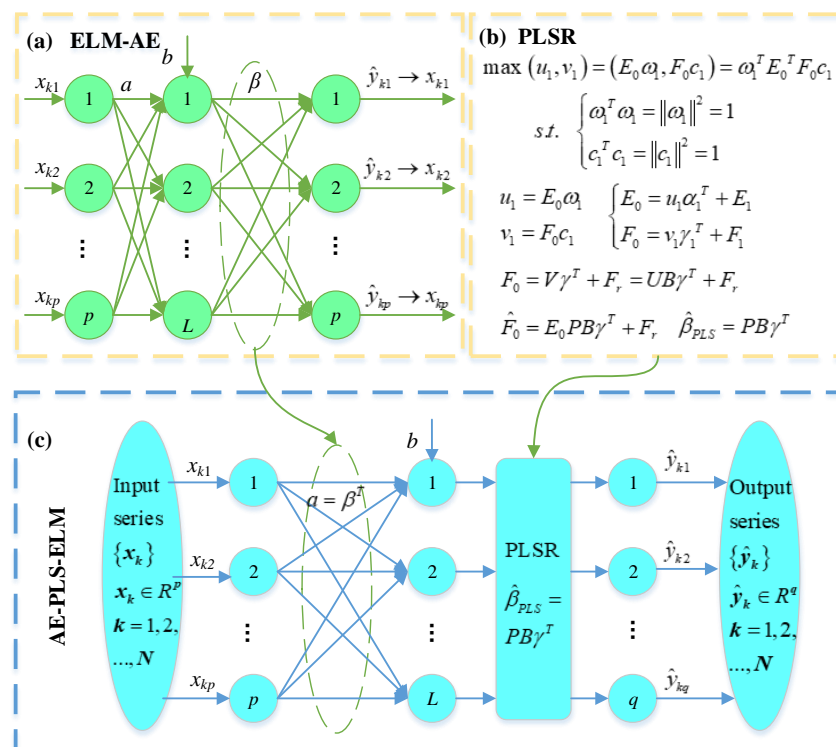
$$\hat{F}_0 = E_0 WB\gamma^T + F_r \quad (30)$$

According to the description above, the hidden layer output weight vector can be expressed by the equation below.

$$\hat{\beta}_{PLS} = WB\gamma^T \quad (31)$$

where  $W$  denotes the component matrix and  $B$  denotes the diagonal matrix.

Based on the above modeling process, the structure of the proposed AE-PLS-ELM model is shown in Figure 3 as follows.



**Figure 3.** Structure of the proposed Autoencoder and partial least squares regression based extreme learning machine (AE-PLS-ELM) model. (a) ELM-AE, (b) PLSR, (c) AE-PLS-ELM.

#### 4. Modeling Process of the Pump Turbine Based on AE-PLS-ELM

The proposed AE-PLS-ELM model is used to model the pump turbine characteristics of a PSU. The flow and moment characteristic curves of the pump turbine are preprocessed using an improved Suter transformation method. Two independent AE-PLS-ELM models are used to model the preprocessed complete curves. The preprocessed complete curves are then converted into target variables to construct a neural network model that can be used for real-time simulation. The input of the AE-PLS-ELM model is the relative flow angle  $x$  and the relative vane opening  $y$ . The full characteristic curves based on improved Suter transformation (Figure 2) are employed as data samples.

The specific steps of the modeling process of the pump turbine characteristics based on AE-PLS-ELM are as follows.

Step 1: Apply the improved Suter transformation to the complete characteristic curves of the pump turbine provided by the power station and the corresponding preprocessed WH and WM curves are obtained.

Step 2: Extract data points from the curves. Convert the relative flow angle  $x$ , the relative vane opening  $y$ , and the extracted data points into input and output sample pairs.

Step 3: Divide the above sample data into training data and test data. Since the dimensions and magnitudes of the data samples are different, the input and output data are normalized to facilitate the modeling and calculation process.

Step 4: Set the Sigmoid function as the activation function of the hidden layer and determine the range of the number of hidden layer nodes, according to the Kolmogorov empirical formula. The optimal number of hidden nodes is selected using a trial calculation modeling error.

Step 5: Import the normalized training data to the AE-PLS-ELM model for training, and a well-trained AE-PLS-ELM model is obtained.

Step 6: Import the normalized test data to the well-trained AE-PLS-ELM model and de-normalize the output data to obtain the test output.

#### 5. Numerical Experiments and Analysis

In this section, a PSU in China is used as the research object to carry out the nonlinear modeling of the pump turbine [8]. A total number of 1125 data points are extracted from the pre-processed WH and WM characteristic curves using the improved Suter transformation (Figure 2). Thus, 1125 data pairs are generated for constructing an AE-PLS-ELM model. In addition, 90% of the data pairs (1012 points) are employed as the training samples and the remaining 10% (112 points) as the test sample.

##### 5.1. Parameters Setting

To highlight the effectiveness of the proposed model, four conventional data-driven techniques named the Bagtree, the support vector regression (SVR), the BP neural network [35], and the ELM are employed as a control group to simulate the complete characteristic curve of the pump turbine. The Bagtree model is constructed using MATLAB's "Bag" function. The parameters of the SVR model including the penalty factor  $C$  and the kernel parameter  $\sigma$  are obtained using the grid search (GS) algorithm. The search range of  $C$  is set as  $[2^{-8}, 2^8]$ , and the search range of  $\sigma$  is set as  $[2^{-5}, 2^5]$ . The "trainlm" algorithm is employed in training the BP neural network. The maximum number of iterations is set as 500 and the target error is  $1e-5$ . The number of the hidden nodes is selected using a trial-and-error method.

## 5.2. Comparative Analysis of the Results

To evaluate the performance of different models, four evaluation indices were used to include the root mean square error (RMSE), the mean absolute error (MAE), the mean absolute percent error (MAPE) [23,36], and the Nash-Sutcliffe Efficiency (NSE) [37] are employed. The four indices are defined as follows.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_s(i) - f_o(i))^2} \quad (32)$$

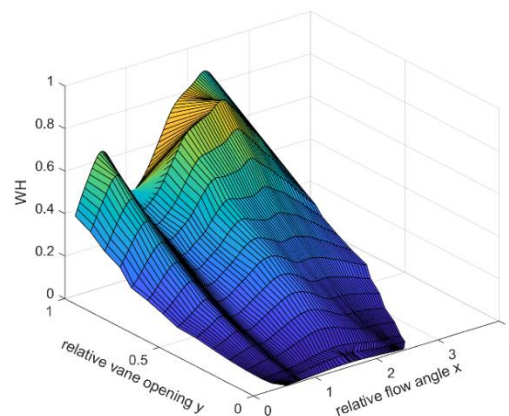
$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |f_s(i) - f_o(i)| \quad (33)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|f_s(i) - f_o(i)|}{f_o(i)} \times 100\% \quad (34)$$

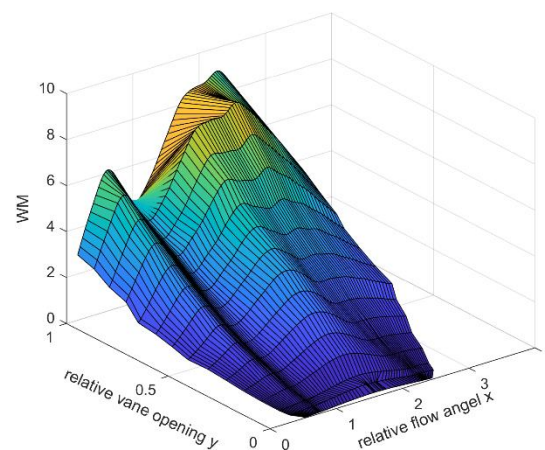
$$\text{NSE} = 1 - \frac{\sum_{i=1}^N (f_s(i) - f_o(i))^2}{\sum_{i=1}^N (f_o(i) - \overline{f_o(i)})^2} \quad (35)$$

where  $f_s(i)$  and  $f_o(i)$  denote the simulated and observed value of the  $i$ th sample point, respectively,  $\overline{f_o(i)}$  denotes the mean observed value, and  $N$  denotes the size of the data set.

The 3D spatial surfaces for the WH and the WM characteristics based on the AE-PLS-ELM model are shown in Figures 4 and 5, respectively. As can be seen from Figures 4 and 5, the WH and WM spatial surfaces based on the AE-PLS-ELM model are smooth and uniform. In addition, the complete characteristic curves are continuous derivative, which make it easy to ensure the convergence of the water hammer calculation process. The AE-PLS-ELM model can also be used to encrypt and extend the WH and WM characteristic surfaces, according to the practical requirements for research and engineering applications. Therefore, the transition between different opening lines on the surface is smoother and it is convenient for operators to obtain the pump turbine characteristics of different working conditions.



**Figure 4.** 3D surface for WH characteristics based on the proposed AE-PLS-ELM model.



**Figure 5.** 3D surface for WM characteristics based on the proposed AE-PLS-ELM model.

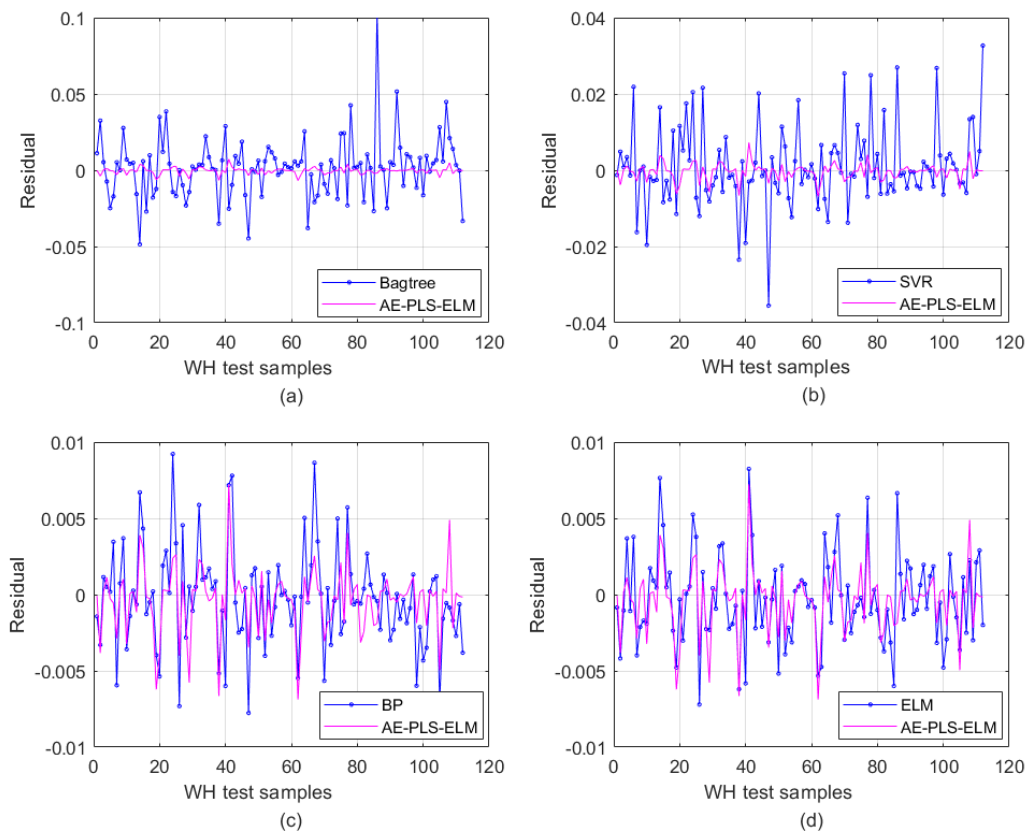
The results of the Bagtree, SVR, BP, ELM, and AE-PLS-ELM models for WH characteristics are shown in Table 1. It can be seen from the table that the five models have good training and test accuracy and can model the WH characteristics of the PSU accurately. The AE-PLS-ELM model performs the best in terms of the four indices in both the training and test period among the five models, which indicates that the AE-PLS-ELM model can enhance the modeling accuracy of pump turbine for the WH characteristic curve effectively. Taking the RMSE value of the test period as an example, the RMSE value of ELM is 0.00297, which is lower than that of the Bagtree, SVR, and BP models. The neural network models as BP and ELM perform better than the Bagtree and SVR models. The performance of ELM is slightly better than BP and the Bagtree model performs the worst.

**Table 1.** Statistics error indices for WH characteristics of different models.

Characteristics	Model	Training				Test			
		RMSE	MAE	MAPE	NSE	RMSE	MAE	MAPE	NSE
WH	Bagtree	0.01827	0.01214	0.13402	0.99232	0.02034	0.01404	0.16405	0.98938
	SVR	0.01019	0.00711	0.05071	0.99646	0.01069	0.00749	0.05963	0.99464
	BP	0.00317	0.00222	0.02438	0.99977	0.00324	0.00236	0.02524	0.99973
	ELM	0.00290	0.00207	0.02198	0.99981	0.00297	0.00234	0.02467	0.99977
	AE-PLS-ELM	0.00229	0.00130	0.00599	0.99988	0.00217	0.00142	0.00697	0.99988

Compare the training and test results between the ELM and AE-PLS-ELM models, it can be found that the performance of the AE-PLS-ELM model is significantly better than the ELM model. For the training samples, the RMSE, MAE, MAPE, and NSE values of the AE-PLS-ELM model were 0.00229, 0.00130, 0.00599, and 0.99988, respectively, which were improved by 21.03%, 37.20%, 72.75%, and 0.007% compared with the 0.00290, 0.00207, 0.02198, and 0.99981 obtained by the ELM model. For the test samples, the RMSE, MAE, MAPE, and NSE values of the AE-PLS-ELM model were 0.00217, 0.00142, 0.00697, and 0.99988, respectively, which were improved by 26.93%, 39.32%, 71.75%, and 0.011% when compared with the 0.00297, 0.00234, 0.02467, and 0.99977 obtained by the ELM model. In a word, the proposed AE-PLS-ELM model overcomes the instability and multi-collinearity of the single ELM model, and can improve the generalization ability and fitting accuracy of the ELM for modeling the pump turbine characteristics.

The comparison of the residuals for the WH characteristics between the AE-PLS-ELM model and the Bagtree, SVR, BP, and ELM models are shown in Figure 6a–d, respectively. It can be seen from Figure 6 that the prediction accuracy of the AE-PLS-ELM model is significantly better than that of the Bagtree and SVR models at all test sample points. In addition, the residual of the AE-PLS-ELM model is generally smaller than the BP and ELM models at most of the test points.

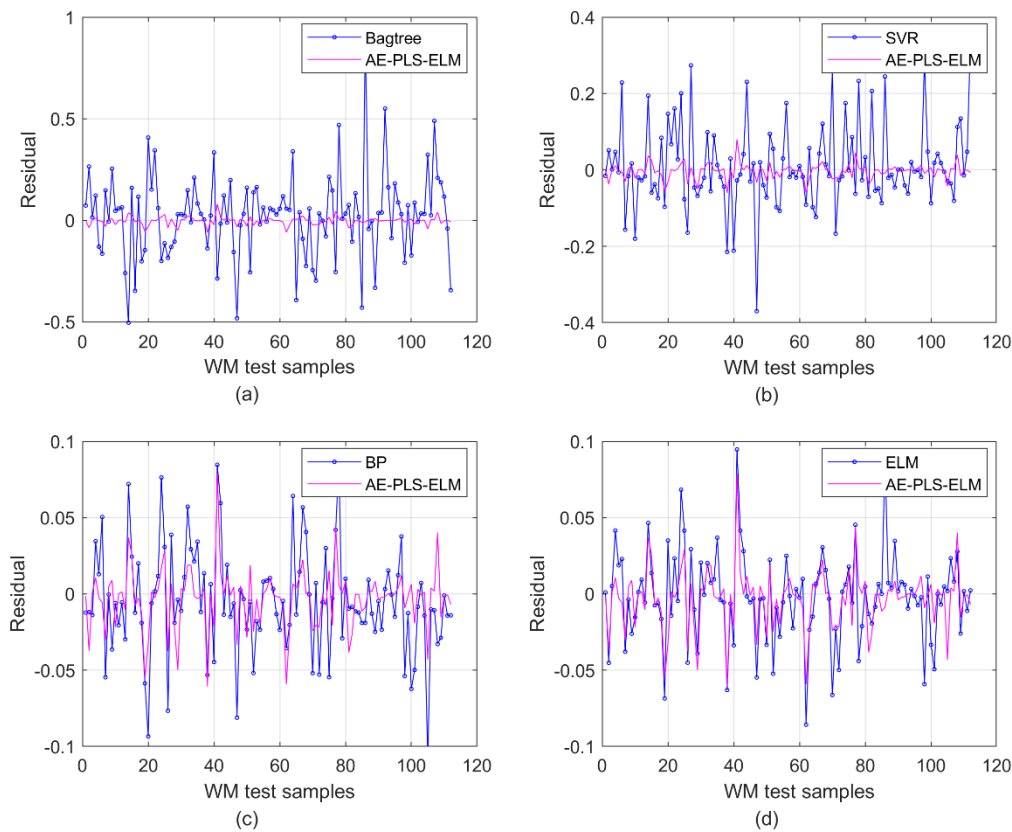


**Figure 6.** Comparison of the test residuals for WH characteristics between Bagtree, SVR, BP, ELM, and AE-PLS-ELM. (a) Bagtree vs. AE-PLS-ELM, (b) SVR vs. AE-PLS-ELM, (c) BP vs. AE-PLS-ELM, (d) ELM vs. AE-PLS-ELM.

The training and test results of the Bagtree, SVR, BP, ELM, and AE-PLS-ELM models for the WM characteristic are shown in Table 2. The comparison of the residuals for the WH characteristics between the AE-PLS-ELM model and the Bagtree, SVR, BP, and ELM models are shown in Figure 7a–d, respectively. The results obtained from Table 2 and Figure 7 are consistent with Table 1 and Figure 7. The neural network models that BP, ELM, and AE-PLS-ELM performed better than the Bagtree and the SVR models. The AE-PLS-ELM model can overcome the instability and multi-collinearity of the single ELM model, and obtain higher modeling accuracy.

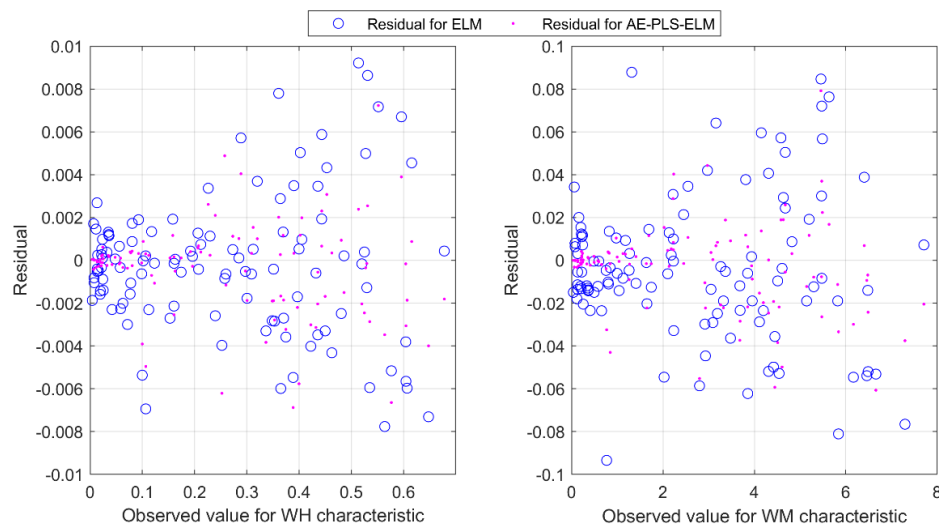
**Table 2.** Statistics error indices for WM characteristics of different models.

Characteristics	Model	Training				Test			
		RMSE	MAE	MAPE	NSE	RMSE	MAE	MAPE	NSE
WM	Bagtree	0.18222	0.12422	0.14329	0.99315	0.21125	0.15082	0.18732	0.99001
	SVR	0.10730	0.07476	0.05411	0.99763	0.11193	0.07852	0.06372	0.99719
	BP	0.03227	0.02232	0.02680	0.99979	0.03508	0.02601	0.03050	0.99972
	ELM	0.02656	0.01713	0.01582	0.99985	0.02950	0.02081	0.02110	0.99981
	AE-PLS-ELM	0.02315	0.01269	0.00539	0.99989	0.02027	0.01306	0.00615	0.99991



**Figure 7.** Comparison of the test residuals for WM characteristics between Bagtree, SVR, BP, ELM and AE-PLS-ELM. (a) Bagtree vs. AE-PLS-ELM, (b) SVR vs. AE-PLS-ELM, (c) BP vs. AE-PLS-ELM, (d) ELM vs. AE-PLS-ELM.

The regression analysis scatter diagram of the residual and actual value for the WH and WM characteristics is shown in Figure 8. It can be seen from Figure 8 that the scatter plot of the single ELM model is more divergent around the axis, and the points of the AE-PLS-ELM model distributes more closely, which demonstrates the superiority of the AE-PLS-ELM model in modeling the complete characteristic curves of the pump turbine.



**Figure 8.** Comparison of the Scatter diagram for the test residuals between ELM and AE-PLS-ELM.



## 6. Additional Test Problem

To further demonstrate the effectiveness of the proposed AE-PLS-ELM model, a widely used nonlinear differential equation is studied as an additional test problem [38]. The nonlinear differential equation can be expressed by the equation below [9].

$$y(k+1) = \frac{y(k)y(k-1)y(k-2)u(k-1)(y(k-2)-1) + u(k)}{1 + y^2(k-2) + y^2(k-1)} \quad (36)$$

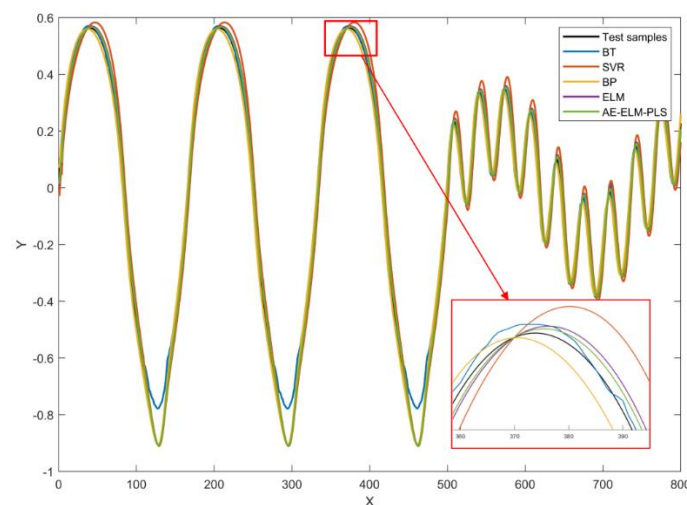
where  $u(k)$  is the control variable and is randomly generated in  $[-2,2]$  in the training period. Furthermore, 800 data pairs are generated for training.  $y(k+1)$  is taken as the output variable.  $y(k), y(k-1), y(k-2), u(k), u(k-1)$  are taken as the input variables. In the test period,  $u(k)$  is generated using the following equation.

$$u(k) = \begin{cases} \sin(3\pi k/250), & k \leq 500 \\ 0.25 \sin(2\pi k/250) + 0.2 \sin(3\pi k/50), & k > 500 \end{cases} \quad (37)$$

where  $k = 1, 2, \dots, 800$ , which means 800 data pairs are generated for testing. The RMSE, MAE, MAPE, and NSE for the test problem of the Bagtree, SVR, BP, ELM, and AE-PLS-ELM models are given Table 3. The test results of the five different models are shown in Figure 9. As can be seen in Table 3, the AE-PLS-ELM outperforms the other four models in terms of the four indices in the test period. The performance of the SVR model is the worst. The Bagtree model performed the second worst and the BP and ELM models perform better than the Bagtree model. It can also be observed from Table 3 that the SVR and ELM models encounter the overfitting problem during the test period. Their training performances are much better than the other models while the test performances are much worse.

**Table 3.** Statistics error indices for the test problem of different models.

Model	Training				Test			
	RMSE	MAE	MAPE	NSE	RMSE	MAE	MAPE	NSE
Bagtree	0.06928	0.00911	0.02077	0.98817	0.03897	0.02709	0.20570	0.99174
SVR	0.00002	0.00001	0.00004	1.00000	0.05423	0.04805	0.41004	0.98400
BP	0.00010	0.00007	0.00027	1.00000	0.02675	0.02348	0.20595	0.99611
ELM	0.00004	0.00003	0.00007	1.00000	0.02506	0.02220	0.19029	0.99658
AE-PLS-ELM	0.00010	0.00007	0.00021	1.00000	0.02038	0.01793	0.14956	0.99774



**Figure 9.** Comparison of the test results of the five different models for the nonlinear differential equation.

## 7. Conclusions

PTGS plays an extremely important role in maintaining the safe and stable operation of the power system. However, it is a closed-loop control system with a complex structure, variable parameters, and strong nonlinearity. As a crucial part of PSU, an accurate pump turbine model is the key to the accurate modeling and simulation of PTGS. This study first introduced an improved Suter transformation to process the complete characteristic curves of the pump turbine. The crossing, aggregating phenomena, and multi-value problems in the “S” characteristic region of the pump turbine were reduced through the improved Suter transformation. Furthermore, an AE-PLS-ELM model was proposed to model the pump turbine characteristics precisely. The AE technique was introduced to the single ELM model for feature extraction of input data to improve its stability. In addition, the PLSR algorithm was employed to replace the Moore-Penrose generalized inverse in ELM to reduce the multicollinearity of the output weight. Results have shown that the proposed AE-PLS-ELM model has better fitting precision and generalization performance than traditional models such as Bagtree, SVM, BP, and ELM. Essentially, the proposed modeling framework is an effective technique in modeling the pump turbine characteristics and the proposed AE-PLS-ELM can be used in other regression problems in a future study. However, the performances of some other data-driven methods with different structures, such as the multivariate adaptive regression spline (MARS), the gene expression programming (GEP) [39], the general regression neural network (GRNN), the genetic programming (GP), and the cascaded neural network (CCNN) have not been studied. More attention will be paid to the performances of different data-driven methods for nonlinear modeling of pump turbine characteristics in the future study.

**Author Contributions:** C.Z. designed and performed the experiments and drew the figures. T.P. analyzed the results, and wrote the original draft. J.Z. provided the data, reviewed the paper, and gave constructive advice. J.J. checked the whole paper and improved the writing of this paper. X.W. collected relevant material and gave some advice for advisement.

**Funding:** This research received no external funding.

**Acknowledgments:** The Natural Science Foundation of Jiangsu Province (No. BK20191052), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 19KJB470012), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 19KJB480007), the National Natural Science Foundation of China (No. 51741907), the National Natural Science Foundation of China (NSFC) (No. 51709121), and the National Natural Science Foundation of China (No. 51709122) supported this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

<i>Pump turbine model</i>		<i>Improved Suter transformation</i>	
$M_{11}$	unit moment	$a$	relative speed
$Q_{11}$	unit flow	$q$	relative flow
$N_{11}$	unit speed	$h$	relative water head
$a$	guide vane opening	$m$	relative moment
		$x$	relative flow angle
		$y$	relative opening
<i>ELM</i>		<i>AE-PLS-ELM</i>	
$(x_k, y_k)$	Training samples	$u_i, v_i$	Score vectors
$\hat{y}_k$	Simulated output	$\omega_i, c_i$	Load factors
$N$	Number of training samples	$\lambda_1, \lambda_2$	Lagrange multipliers
$L$	Number of hidden neurons	$E_r, F_r$	Residual matrix
$a_i, \beta_i$	Connection weights	$\alpha_i, \gamma_i$	Regression coefficients
$b_i$	Bias	$\theta_i$	Objective function
$g(\cdot)$	Activation function	$W$	Component matrix
$H$	Hidden layer output matrix	$B$	Diagonal matrix
$H^\dagger$	Generalized inverse matrix of $H$		

## Abbreviations

AE	Autoencoder	GS	Grid search
AE-ELM	Autoencoder based extreme learning machine	MAE	Mean absolute error
	Autoencoder and partial least squares regression based extreme learning machine	MAPE	Mean absolute percent error
AE-PLS-ELM		NSE	Nash-Sutcliffe Efficiency
		PLSR	Partial least squares regression
BP	Backpropagation	PSU	Pumped storage unit
RBF	Radial basis function	PTGS	Pump turbine governing system
ELM	Extreme learning machine	RMSE	Root mean square error
		SVR	Support vector regression

## References

- Li, C.; Zou, W.; Zhang, N.; Lai, X. An evolving T-S fuzzy model identification approach based on a special membership function and its application on pump-turbine governing system. *Eng. Appl. Artif. Intell.* **2018**, *69*, 93–103. [\[CrossRef\]](#)
- Zhou, J.; Zhang, C.; Peng, T.; Xu, Y. Parameter Identification of Pump Turbine Governing System Using an Improved Backtracking Search Algorithm. *Energies* **2018**, *11*, 1668. [\[CrossRef\]](#)
- Fu, W.; Wang, K.; Li, C.; Li, X.; Li, Y.; Zhong, H. Vibration trend measurement for a hydropower generator based on optimal variational mode decomposition and an LSSVM improved with chaotic sine cosine algorithm optimization. *Measurement Science and Technology* **2018**, *30*, 015012. [\[CrossRef\]](#)
- Xu, Y.; Zheng, Y.; Du, Y.; Yang, W.; Peng, X.; Li, C. Adaptive condition predictive-fuzzy PID optimal control of start-up process for pumped storage unit at low head area. *Energy Convers. Manag.* **2018**, *177*, 592–604. [\[CrossRef\]](#)
- Lai, X.; Li, C.; Zhou, J.; Zhang, N. Multi-objective optimization for guide vane shutting based on MOASA. *Renew. Energy* **2019**, *139*, 302–312. [\[CrossRef\]](#)
- Wang, L.; Han, Q.; Chen, D.; Wu, C.; Wang, X. Non-linear modelling and stability analysis of the PTGS at pump mode. *IET Renew. Power Gener.* **2017**, *11*, 827–836. [\[CrossRef\]](#)
- Zhang, H.; Chen, D.; Xu, B.; Patelli, E.; Tolo, S. Dynamic analysis of a pumped-storage hydropower plant with random power load. *Mech. Syst. Signal Process.* **2018**, *100*, 524–533. [\[CrossRef\]](#)
- Zhang, C.; Peng, T.; Li, C.; Fu, W.; Xia, X.; Xue, X. Multiobjective Optimization of a Fractional-Order PID Controller for Pumped Turbine Governing System Using an Improved NSGA-III Algorithm under Multiworking Conditions. *Complexity* **2019**, 2019. [\[CrossRef\]](#)
- Zhang, C.; Li, C.; Peng, T.; Xia, X.; Xue, X.; Fu, W.; Zhou, J. Modeling and Synchronous Optimization of Pump Turbine Governing System Using Sparse Robust Least Squares Support Vector Machine and Hybrid Backtracking Search Algorithm. *Energies* **2018**, *11*, 3108. [\[CrossRef\]](#)
- Palikhe, S.; Zhou, J.; Bhattarai, K.P. Hydraulic Oscillation and Instability of a Hydraulic System with Two Different Pump-Turbines in Turbine Operation. *Water* **2019**, *11*, 692. [\[CrossRef\]](#)
- Zuo, Z.; Fan, H.; Liu, S.; Wu, Y. S-shaped characteristics on the performance curves of pump-turbines in turbine mode—A review. *Renew. Sustain. Energy Rev.* **2016**, *60*, 836–851. [\[CrossRef\]](#)
- Lima, G.M.; Júnior, E.L. Method to Estimate Complete Curves of Hydraulic Pumps through the Polymorphism of Existing Curves. *J. Hydraul. Eng.* **2017**, *143*, 04017017. [\[CrossRef\]](#)
- Huang, W.; Yang, K.; Guo, X.; Ma, J.; Wang, J.; Li, J. Prediction Method for the Complete Characteristic Curves of a Francis Pump-Turbine. *Water* **2018**, *10*, 205. [\[CrossRef\]](#)
- Li, J.; Han, C.; Yu, F. A New Processing Method Combined with BP Neural Network for Francis Turbine Synthetic Characteristic Curve Research. *Int. J. Rotating Mach.* **2017**, 2017. [\[CrossRef\]](#)
- Qian, J.; Zeng, Y.; Guo, Y.; Zhang, L. Reconstruction of the complete characteristics of the hydro turbine based on inner energy loss. *Nonlinear Dyn.* **2016**, *86*, 963–974. [\[CrossRef\]](#)
- Tanaka, H.; Tsunoda, S. The development of high head single stage pump-turbines. In Proceedings of the 10th IAHR Symposium, Tokyo, Japan, 1–4 January 1980; pp. 429–440.

17. Yang, J.; Pavesi, G.; Yuan, S.; Cavazzini, G.; Ardizzon, G. Experimental Characterization of a Pump–Turbine in Pump Mode at Hump Instability Region. *J. Fluids Eng.* **2015**, *137*, 051109. [[CrossRef](#)]
18. Boldy, A.; Walmsley, N. Representation of the characteristics of reversible pump turbines for use in waterhammer simulations. In Proceedings of the 4th International Conference on Pressure Surges, Bath, UK, 21–23 September 1983; pp. 287–296.
19. Suter, P. Representation of Pump Characteristics for Calculation of Water Hammer. *Sulzer Tech. Rev.* **1966**, *4*, 45–48.
20. Liu, Z.M.; Zhang, D.H.; Liu, Y.Y.; Zhang, X. New Suter-transformation Method of Complete Characteristic Curves of Pump-turbines Based on the 3-D Surface. *China Rural Water Hydropower* **2015**, *1*, 143–145.
21. Zhiyong, Z.; Bing, S.; Xu, L.; Xuerui, G. Nonlinear simulation of Kaplan turbine regulating system based on RBF networks. In Proceedings of the 2011 International Conference on Electric Information and Control Engineering, Wuhan, China, 15–17 April 2011; pp. 4302–4306.
22. Peng, T.; Zhou, J.; Zhang, C.; Zheng, Y. Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine. *Energy Convers. Manag.* **2017**, *153*, 589–602. [[CrossRef](#)]
23. Zhang, C.; Zhou, J.; Li, C.; Fu, W.; Peng, T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *143*, 360–376. [[CrossRef](#)]
24. Fu, W.; Wang, K.; Zhou, J.; Xu, Y.; Tan, J.; Chen, T. A Hybrid Approach for Multi-Step Wind Speed Forecasting Based on Multi-Scale Dominant Ingredient Chaotic Analysis, KELM and Synchronous Optimization Strategy. *Sustainability* **2019**, *11*, 1804. [[CrossRef](#)]
25. Chang, C.-W.; Lee, H.-W.; Liu, C.-H. A Review of Artificial Intelligence Algorithms Used for Smart Machine Tools. *Inventions* **2018**, *3*, 41. [[CrossRef](#)]
26. Han, M.; Zhang, R.; Xu, M. Multivariate Chaotic Time Series Prediction Based on ELM–PLSR and Hybrid Variable Selection Algorithm. *Neural Process. Lett.* **2017**, *46*, 1–13. [[CrossRef](#)]
27. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; pp. 985–990.
28. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B* **2012**, *42*, 513–529. [[CrossRef](#)] [[PubMed](#)]
29. Li, K.; Xiong, M.; Li, F.; Su, L.; Wu, J. A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine. *Neurocomputing* **2019**, *350*, 261–270. [[CrossRef](#)]
30. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning internal representation by back-propagation of errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
31. Lekamalage, L.; Kasun, C.; Zhou, H.; Huang, G.-B.; Cambria, E. Representational learning with ELMs for big data. *IEEE Intell. Syst.* **2013**, *11*, 31–34.
32. Sun, K.; Zhang, J.; Zhang, C.; Hu, J. Generalized extreme learning machine autoencoder and a new deep neural network. *Neurocomputing* **2017**, *230*, 374–381. [[CrossRef](#)]
33. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
34. Duzan, H.; Shariff, N.S.B.M. Ridge regression for solving the multicollinearity problem: Review of methods and models. *J. Appl. Sci.* **2015**, *15*, 392. [[CrossRef](#)]
35. Guo, Y.; Zeng, Y.; Fu, L.; Chen, X. Modeling and Experimental Study for Online Measurement of Hydraulic Cylinder Micro Leakage Based on Convolutional Neural Network. *Sensors* **2019**, *19*, 2159. [[CrossRef](#)] [[PubMed](#)]
36. Fu, W.; Wang, K.; Li, C.; Tan, J. Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM. *Energy Convers. Manag.* **2019**, *187*, 356–377. [[CrossRef](#)]

37. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
38. Vuković, N.; Miljković, Z. Robust sequential learning of feedforward neural networks in the presence of heavy-tailed noise. *Neural Netw.* **2015**, *63*, 31–47. [[CrossRef](#)] [[PubMed](#)]
39. Mirabbasi, R.; Kisi, O.; Sanikhani, H.; Meshram, S.G. Monthly long-term rainfall estimation in Central India using M5Tree, MARS, LSSVR, ANN and GEP models. *Neural Comput. Appl.* **2018**, 1–20. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).