

Article

A Comprehensive Characteristic Modeling Method for Francis Turbine Based on Image Digitization and RBF Neural Network

Youhan Deng ^{1,2}, Youping Li ^{2,3}, Xiaojun Hua ^{2,3}, Rui Lyu ^{2,3}, Yushu Li ^{1,2}, Lei Wang ^{2,3}, Weiwei Yao ^{1,2}, Yifeng Gu ⁴ , Fangqing Zhang ⁴  and Jiang Guo ^{4,*} 

¹ Institute of Science and Technology, China Three Gorges Corporation, Beijing 100038, China; deng_youhan@ctg.com.cn (Y.D.); li_yushu@ctg.com.cn (Y.L.); yao_weiwei@ctg.com.cn (W.Y.)

² Laboratory of Hydro-Wind-Solar Multi-Energy Control Coordination, Wuhan 430014, China; li_youping@ctg.com.cn (Y.L.); hua_xiaojun@ctg.com.cn (X.H.); lv_rui@ctg.com.cn (R.L.); wang_lei@ctg.com.cn (L.W.)

³ China Yangtze Power Co., Ltd., Wuhan 430010, China

⁴ School of Power and Mechanical Engineering, Wuhan University, Wuhan 430072, China; guyifeng@whu.edu.cn (Y.G.); zfq@whu.edu.cn (F.Z.)

* Correspondence: guo.river@whu.edu.cn

Abstract

Establishing a mathematical model of a Francis turbine is the foundation for the simulation of hydropower station operation and is of great significance for the analysis of the hydropower station's transient process. Currently, in engineering practice, the model is often established based on the comprehensive characteristic curves of the Francis turbine provided by the manufacturer, using the external characteristic method. Traditional modeling methods mostly adopt manual reading of points or the use of dedicated numerical software for curve tracing to discretely sample the comprehensive characteristic curves of the turbine. This method is labor-intensive, inefficient, and relies on manual experience, with a small sample size, which, to some extent, affects the accuracy and reliability of the numerical processing results and cannot meet the needs of transient process simulation analysis. To address these shortcomings, this paper proposes a refined modeling method based on image numerical processing and an RBF neural network. Taking the HLA685 Francis turbine as an example, the method first uses image processing to achieve large-scale automated discrete sampling of the turbine's high-efficiency zone characteristic data, then reasonably extends the small-opening and low-speed regions, and finally uses the RBF neural network method for interpolation and extrapolation to obtain the full characteristic data. This method can effectively improve the efficiency and accuracy of comprehensive characteristic modeling of the turbine and has good reference significance for the comprehensive characteristic modeling of blade-type machinery.

Keywords: comprehensive characteristic curve of Francis turbine; discrete sampling; image processing; RBF neural network



Academic Editor: Gustavo Fimbres Weihs

Received: 15 October 2025

Revised: 24 November 2025

Accepted: 25 November 2025

Published: 5 December 2025

Citation: Deng, Y.; Li, Y.; Hua, X.; Lyu, R.; Li, Y.; Wang, L.; Yao, W.; Gu, Y.; Zhang, F.; Guo, J. A Comprehensive Characteristic Modeling Method for Francis Turbine Based on Image Digitization and RBF Neural Network. *Energies* **2025**, *18*, 6380. <https://doi.org/10.3390/en18246380>

Copyright: © 2025 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As renewable energy-dominated power systems rapidly evolve, hydropower and pumped storage plants serve as crucial regulating sources to ensure grid stability, security, and operational efficiency. This has raised higher requirements for high-precision modeling and dynamic simulation of Francis turbines. As the most widely used modern turbine type, Francis turbines exhibit highly complex internal water flow dynamics. The relationships between various characteristic parameters cannot be precisely expressed, making it

difficult to establish an accurate full-characteristic mathematical model. In practice, internal characteristic methods and external characteristic methods are commonly employed for modeling.

The internal characteristic method derives mathematical relationships between a turbine's internal characteristics, flow patterns, and structural parameters through hydraulic machinery theory. However, due to computationally complex analytical equations prone to ill-conditioning, reliance on empirical parameter selection, and limited modeling accuracy [1,2], its application in practical engineering is relatively limited. The external characteristic method, based on comprehensive characteristic curves from model tests provided by manufacturers, uses mathematical approaches to reasonably infer the variation patterns of turbine characteristic parameters. This method has been widely adopted in research and engineering practice. Nevertheless, the characteristic curves provided by manufacturers are mostly in paper or image formats and primarily cover the high-efficiency operating region. In engineering applications, it is often necessary to first discretely sample the curve data and then employ appropriate methods to interpolate missing characteristic data to obtain full characteristic information.

Currently, commonly used external characteristic curve modeling methods primarily rely on mathematical interpolation, numerical fitting, and neural network algorithms to mathematically approximate known characteristic point patterns. Ma et al. [3] systematically analyzed partitioning rules and fitting methods for external characteristic curves, achieving comprehensive turbine modeling under multiple boundary conditions. Some scholars have implemented interpolation and fitting of turbine characteristic curves using methods such as Delaunay triangulation [4], principal curve theory [5], and numerical correction [6]. While these methods effectively establish full-characteristic models from limited sampling points, the discrete sampling of characteristic curves still relies on manual point reading or specialized numerical software digitization. This approach suffers from small sample sizes, high workloads, low efficiency, and susceptibility to human errors, which, to some extent, affect modeling accuracy and reliability. Simultaneously, due to the strong nonlinearity of turbine characteristics, the application of common mathematical fitting and interpolation methods still depends on practical engineering calibration, resulting in modeling effects lacking universality [7].

In recent years, neural network-based modeling methods, particularly Back Propagation (BP) neural networks and Radial Basis Function (RBF) neural networks, have attracted significant attention due to their excellent nonlinear approximation capabilities and ease of application [8–10]. However, their practical modeling effectiveness is generally constrained due to widespread limitations, including insufficient sample sizes from traditional sampling methods, a lack of reference points for low-speed and low-flow conditions, and ambiguous boundary conditions. To fundamentally address these challenges, this paper proposes an integrated methodology that combines automated image digitization with an enhanced RBF neural network framework, enabling large-scale, high-precision extraction of characteristic data while incorporating physical constraints for boundary condition consistency [11].

Using the HLA685 Francis turbine unit manufactured by Harbin Electric Machinery Company (Harbin, China) as an example, the method first preprocesses the comprehensive characteristic curve images provided by the manufacturer. It then performs connected component recognition, morphological thinning, and coordinate point conversion on the binarized, pixel-level curve data, enabling high-efficiency, high-precision, and large-scale rapid extraction of turbine characteristic data. Concurrently, boundary conditions such as runaway characteristics, zero-speed conditions, and zero-flow conditions are incorporated. Combined with engineering simulation experience, reasonable interpolation and extrapolation are applied

to the small guide vane opening and low rotational speed regions. Finally, based on these large-capacity sample points, a turbine model using the RBF neural network is constructed to obtain full characteristic data. This method can effectively enhance the efficiency and sample scale of comprehensive characteristic modeling for turbines. While reducing manual errors, it addresses the issue of insufficient sample points in traditional neural network modeling methods and further improves modeling accuracy. The approach holds significant reference value for the comprehensive characteristic modeling of turbomachinery. The parameters of the HLA685 Francis turbine are shown in Table 1.

Table 1. Parameters of HLA685 Francis turbine.

Parameter Type	Parameter Value
Rated Flow (m^3/s)	27.54
Rated Output (MW)	27.0
Rated Speed (r/min)	375
Rated Head (m)	107

2. Discrete Sampling of Turbine Characteristics Based on Image Processing

The comprehensive characteristic curve of the turbine model provided by the manufacturer is generally graphic information, which usually includes a series of constant opening lines and constant efficiency lines, reflecting the mutual relationship between the unit discharge Q_{11} , unit speed N_{11} , operating efficiency η , and guide vane opening a . Extracting the curves in the figure into a series of discrete points is the basic work for modeling analysis. The conventional operation method is to carry out discrete sampling by relying on manual direct reading points or using special numerical software to trace points. As highlighted in the Introduction, this approach presents several significant limitations. Therefore, this paper proposes an automatic discrete extraction method for the comprehensive characteristic curve of the turbine model based on image processing, mainly including steps such as image preprocessing, image discretization, automatic curve recognition, curve morphology thinning, and characteristic shape value conversion. The fundamental processing workflow is illustrated in Figure 1.

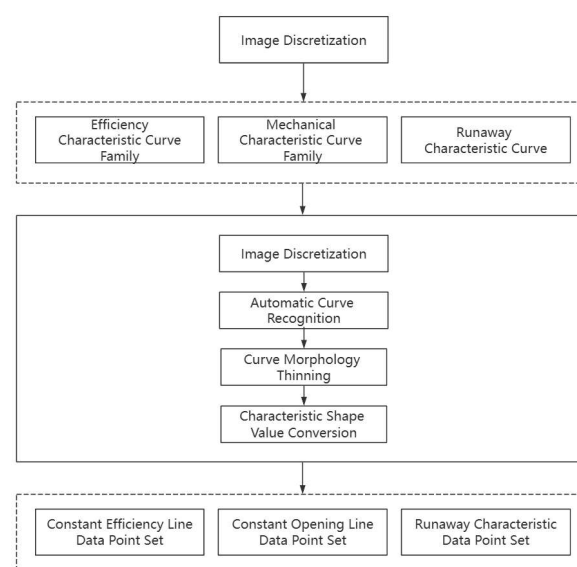


Figure 1. Flow chart of the discrete sampling method for the comprehensive characteristic curve of the Francis turbine.

2.1. Image Preprocessing

Image preprocessing involves highlighting the main features of the characteristic curve through simplification tools or methods before discrete sampling so as to facilitate better extraction of characteristic data, mainly including steps such as image rotation, filter denoising, cropping, and rough curve processing. The process is shown in Figure 2.

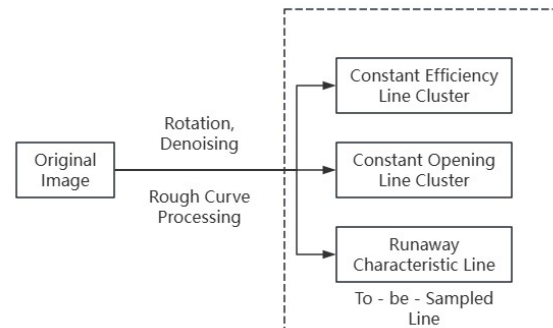


Figure 2. Image preprocessing flow chart.

(1) Image rotation. The comprehensive characteristic graphic information of the turbine often suffers from tilting, which is mainly manifested in that there is a certain angle between the characteristic data coordinate system and the graphic window coordinate system, and the shape values of the curve pixel points are inconsistent with the true values. Therefore, a mathematical method is needed to perform coordinate transformation on the foreground pixel points to achieve the correction of the curve shape value points. As shown in Figure 3, there is a point (x_0, y_0) on the curve in the original graph, and the coordinate after rotation is (x_1, y_1) . Let the distance from the coordinate point to the origin be r . According to the principle that the distance from the coordinate point to the origin remains unchanged before and after rotation, the coordinates of the shape value point after rotation can be expressed as follows:

$$\begin{cases} x_1 = r \cos(\alpha + \beta) = x_0 \cos \alpha - y_0 \sin \alpha \\ y_1 = r \sin(\alpha + \beta) = x_0 \sin \alpha + y_0 \cos \alpha \end{cases} \quad (1)$$

where α is the inclination angle of the graph, and β is the angle between the line connecting the original coordinate point and the origin and the horizontal axis of the graph window. In practice, α can be obtained by measuring the angle between the characteristic data coordinate system and the graphic window coordinate system.

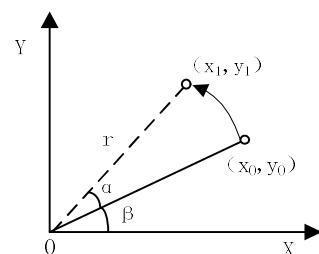


Figure 3. Schematic diagram of the image rotation principle.

(2) Image denoising. The graphic information collected often has a certain degree of blurring, and it is necessary to use filter denoising and other methods to suppress or reduce the noise in the image, which provides convenience for subsequent image processing. At present, image denoising methods are mainly divided into traditional methods based on filters and intelligent denoising methods based on models or deep learning [12]. The

traditional denoising method based on filters has been widely used in engineering due to its clear principle, simple algorithm, and easy realization, mainly including the mean filter, median filter, Gaussian filter, bilateral filter, etc. Among them, methods such as the mean filter, the median filter, and the Gaussian filter will cause blurring or loss of curve edge details while denoising, which is not conducive to subsequent graphic processing. The bilateral filter method takes into account the spatial characteristics and gray similarity, and maintains the image edge characteristics while smoothing the noise. Therefore, this paper selects the bilateral filter method for filter denoising. Its implementation idea is to use the principle of taking the spatial proximity of each point and the central point as the weighting coefficient in the Gaussian filter algorithm, and at the same time consider the similarity factor of the pixel value of each point and the central point, and use the product of the two as the filter template for the convolution operation. Its core equation is as follows:

$$I'(x) = \frac{1}{W} \sum_{y \in \Omega} I(y) \cdot f(\|x - y\|) \cdot g(|I(x) - I(y)|) \quad (2)$$

where $I'(x)$ represents the intensity value of pixel point x after processing, and $I(x)$ represents the intensity value of pixel point x before processing; $f(\|x - y\|)$ is the spatial weight reflecting the positional relationship of pixel points in the neighborhood, $g(|I(x) - I(y)|)$ is the gray value weight reflecting the color value similarity of pixel points in the neighborhood, and both are Gaussian functions; Ω represents the specified neighborhood of pixel point x ; W is the normalization coefficient to ensure that the sum of weights is 1.

(3) Rough Curve Processing. The collected graphic materials typically contain coordinate axes, grid lines, and operational guidance markers such as output limit lines and head annotation lines. The distribution of these elements increases the difficulty of automatically extracting constant opening lines and constant efficiency lines. Prior to formal image digitization, these interfering elements must be deleted, retaining only the target curves requiring discrete sampling. Some scholars have employed methods like the Hough transform for straight-line detection and image orthogonal calculations to identify and remove grid lines and coordinate axes. However, these approaches fail to effectively process irregular curves such as operational guidance markers. Additionally, straight-line detection based on binarized images is significantly affected by image quality. In practice, manual-assisted tracing recognition remains unavoidable, and critical intersection information between target curves and grid lines is easily lost. Therefore, this study utilizes general graphic processing tools for rough curve processing to erase interfering lines. First, images processed through preceding steps are cropped using horizontal and vertical coordinate axes as baselines, retaining only effective regions. Pixel values and true values of each coordinate axis origin are calibrated and recorded for subsequent coordinate point conversion. Second, considering the sparse distribution and regular patterns of grid lines, manual erasure is applied for removal [13–16]. Finally, following the grid line processing approach, constant opening line clusters and constant efficiency line clusters are obtained under identical pixel scales, with the runaway characteristic curve processed identically. This method features straightforward operation, efficiently yielding classified high-quality curve images with consistent scaling to the original material, thereby facilitating subsequent automated large-scale discrete sampling.

As shown in Figure 4, key curve clusters are obtained after rough processing: Figure 4a,b represent collected comprehensive characteristic curves and runaway characteristic curves of the HLA685 Francis turbine; Figure 4c–e display constant opening line clusters, constant efficiency line clusters, and runaway characteristic curves obtained through image preprocessing. Note: In this paper, n_{11} is denoted as N_{11} (unit speed, unit: r/min); a represents guide vane opening (unit: mm); Q_{11} represents unit discharge (unit: L/s).

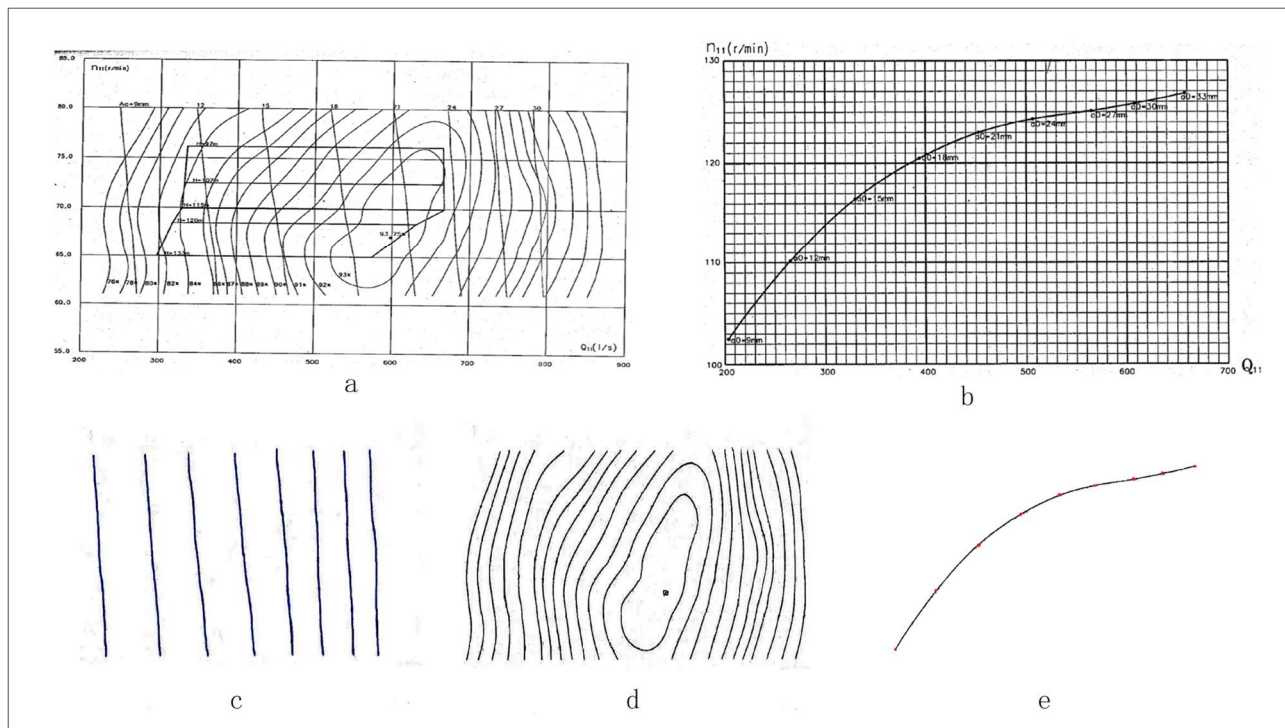


Figure 4. Schematic diagram of the image preprocessing effect: (a) collected comprehensive characteristic curves of the HLA685 Francis turbine; (b) collected runaway characteristic curves; (c) extracted constant guide-vane opening line clusters; (d) extracted constant efficiency line clusters; (e) extracted runaway characteristic curve after image preprocessing.

2.2. Image Discretization

Image discretization is the basis for numerical extraction of characteristic curves. Generally, the target image is grayscaled and binarized, and the characteristic curve is discretized into a series of pixel point sets for subsequent recognition, extraction, analysis, and processing. Image grayscaling is the process of converting the RGB three-channel color image into a single-channel grayscale image. The weighted average method is generally used to calculate the grayscale value, and the grayscale value range of each pixel point is 0–255. Binarization is based on image grayscaling. By setting a threshold, the pixel value greater than or equal to the threshold is forcibly converted to 0, which is displayed as black, and the pixel value lower than the threshold is forcibly converted to 255, which is displayed as white. It can not only effectively highlight the outline of the target curve but also facilitate subsequent processing. The coordinate information of pixel-level data points on the characteristic curve can be obtained through grayscaling and binarization processing, that is, the discretization of the characteristic curve is realized.

2.3. Automatic Curve Recognition

The automatic recognition and tracking of characteristic curves are the key steps in the discrete sampling of turbine characteristic curves. The characteristic curve image after image discretization is actually stored as a pixel matrix of the same size. Each pixel point has a unique grayscale value (0 or 255), and each characteristic curve is represented as a pixel point set with morphological correlation in image storage, which needs to be classified and recognized by mathematical methods to achieve large-scale and automatic collection. Due to the certain regularity in the distribution of constant opening lines and constant efficiency lines, the connected component recognition method can be used to realize the automatic extraction of curve pixel point coordinates. The connected component refers

to the pixel point set with the same pixel value and mutual connection in the binarized image. Connected component recognition has been widely used in image feature extraction, pattern recognition, and other fields [17]. The basic idea of realizing automatic tracking and recognition of characteristic curves is to traverse all pixel points of the binarized image, detect the connection of foreground pixels one by one, and make marks. The basic steps are as follows:

- (1) Load the characteristic curve image and initialize the connected domain set and the mark matrix with the same size as the image to be detected.
- (2) Scan the pixel points one by one in the direction from left to right and top to bottom, and perform the following operations:
 - (a) If the current pixel point is a background pixel (pixel value is 255), skip it directly, and the corresponding mark matrix is marked as 0.
 - (b) If the current pixel point is a foreground pixel (pixel value is 0), detect whether there is a neighborhood mark in the left, upper left, upper, and upper right adjacent points of the point according to the 8-neighborhood connectivity principle. If there is no neighborhood mark in the above adjacent points, the corresponding mark of the point is a new neighborhood mark (the value is increased by 1); if there is a neighborhood mark and the value is the same in the above adjacent points, the corresponding mark of the point is assigned as the adjacent mark value. If there is a conflict in the neighborhood mark values existing in the adjacent points, the corresponding mark of the point is assigned as the minimum value among them, and at the same time, it is recorded that all the neighborhood mark values in the adjacent points are equivalent.
- (3) Merge the mark values with equivalent relations into the same set and clarify that the minimum value in the same set is the root mark.
- (4) Traverse all pixel points again, and replace the temporary mark values in the mark matrix with root marks.
- (5) According to the final mark matrix, classify and extract the pixel point coordinates with the same mark, that is, obtain the connected component set in the image.
- (6) According to the set threshold of the area block size, screen the identified connected components to exclude the blurred noise point blocks existing in the image.
- (7) Finally, the identified connected components are classified and stored as characteristic curve pixel coordinate point sets according to the order of characteristic parameters (such as opening value, efficiency value).

The algorithmic flow is illustrated in Figure 5. The foundation of connected component recognition lies in the pixel matrix obtained after image binarization. Its core concept involves traversing the pixel matrix to identify neighboring points with identical pixel values, then merging and integrating them through adjacent labeling to acquire connected region coordinate sets. In the diagram, the next Position function sequentially returns the coordinates of the next pixel following a left-to-right, top-to-bottom scanning pattern. The purpose of maintaining a connected component label collection is to record all distinct label values, with insertion operations requiring that labels with equivalent relationships be stored within the same array.

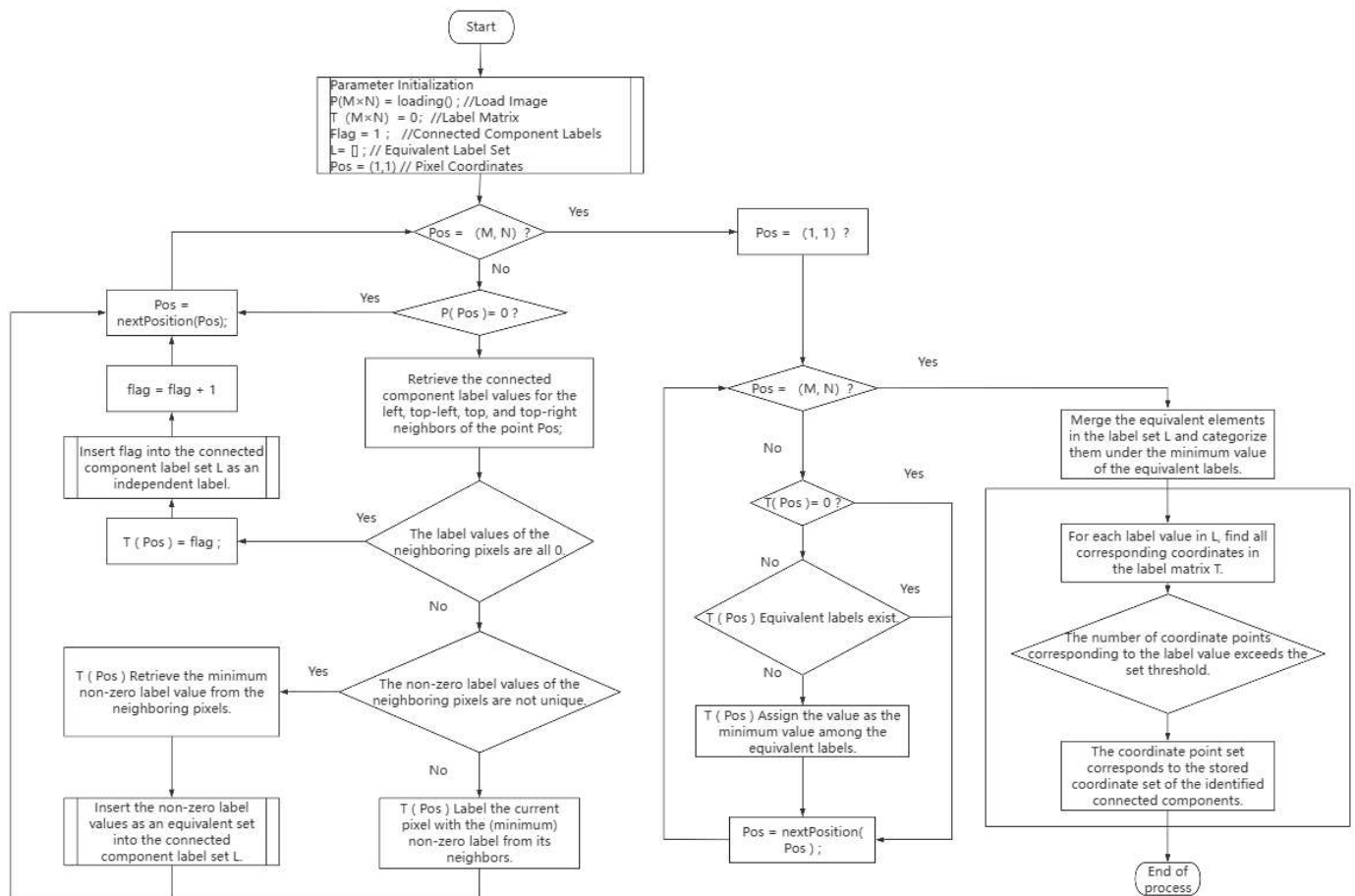


Figure 5. Flowchart of the connected component recognition algorithm.

2.4. Curve Morphology Thinning

Because each curve in the image has a certain width, the curve after discretization in the above steps is represented as a set of pixel points with a line-width area along the curve direction in the image. Considering the uniqueness of the curve shape value point, it is necessary to thin the curve, that is, to obtain the central pixel point along the curve direction as the discrete sampling point of the curve. There are many image thinning algorithms, among which the mathematical morphology thinning algorithm is widely used. The extracted skeleton of the curve is continuous, effectively removing burrs and providing good edge extraction. This paper draws on the principle of the traditional mathematical morphology thinning algorithm. Considering that the extracted constant-opening lines and constant-efficiency lines have good continuity and exhibit a certain distribution law, it is directly simplified to traverse and obtain the central pixel coordinates of each curve along the X and Y axis directions in turn, so as to achieve the purpose of curve thinning extraction.

As shown in Figure 6, the constant opening line extracted by discretization is thinned morphologically. Each constant-opening line is actually composed of discrete pixel points with a certain thickness, and the red line in the middle is its single-pixel morphological center-line, which is used as the sampling pixel point set for subsequent calculations. The XY axis coordinates in the figure are all image pixels.

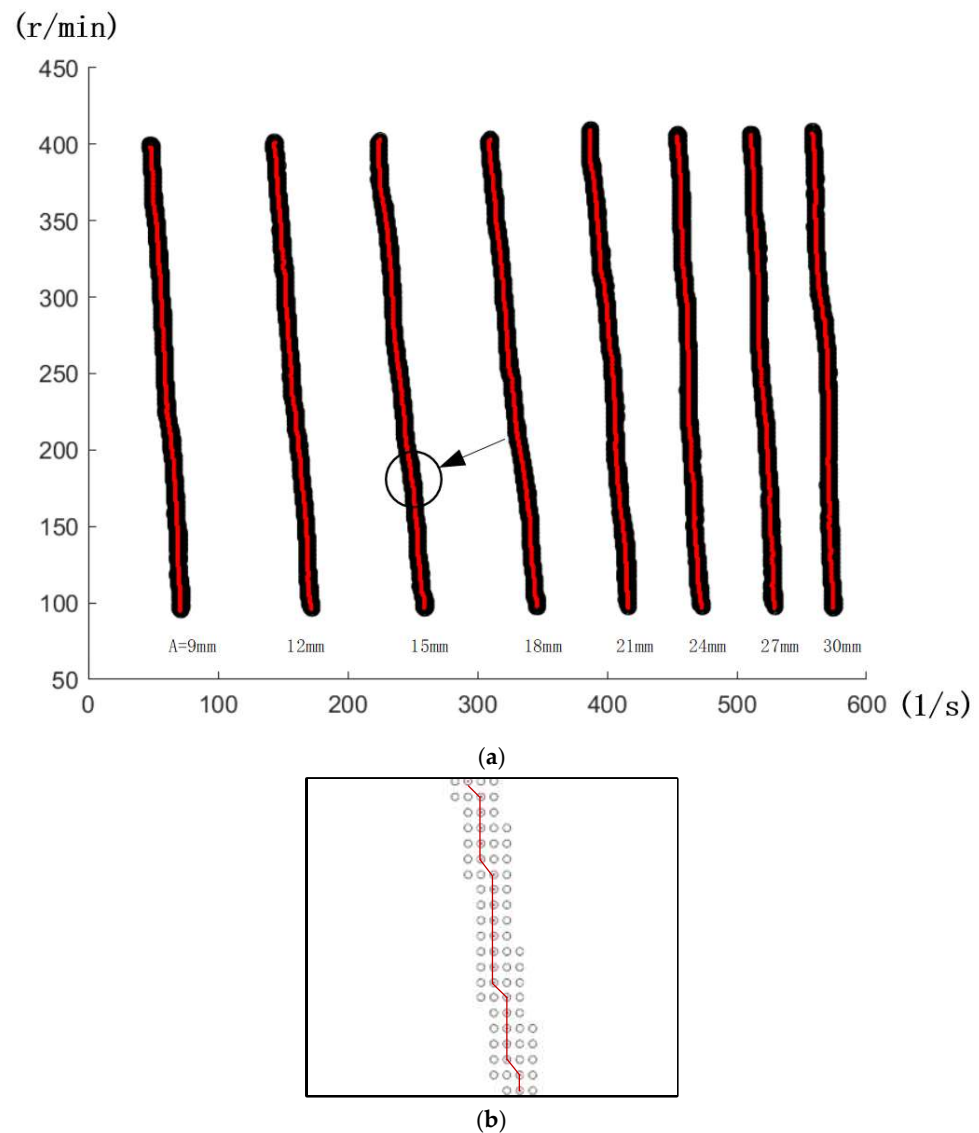


Figure 6. (a) Schematic diagram of the morphological thinning effect of a constant opening line. (b) Magnified view of the morphological thinning effect (constant opening line).

2.5. Coordinate Shape-Value Transformation

In the discretization processing shown in the above image, the shape value point coordinates are all pixel coordinates, and the pixel point coordinates of the extracted characteristic curve need to be converted to the corresponding shape value coordinates of the comprehensive characteristic curve of the turbine model according to the corresponding relationship. It is assumed that the pixel size of the image is $PX \times PY$, and the corresponding X-axis value range is $[Q_{11min}, Q_{11max}]$, and the Y-axis value range is $[N_{11min}, N_{11max}]$. Because the pixel size of the image has a linear corresponding relationship with the value range of the shape value. For any known pixel point (p_x, p_y) on the characteristic curve, its corresponding shape value point coordinates (Q, N) , the conversion relationship between the two can be expressed as follows:

$$\begin{cases} Q = (Q_{11max} - Q_{11min}) \cdot p_x / PX \\ N = (N_{11max} - N_{11min}) \cdot p_y / PY \end{cases} \quad (3)$$

Following data standardization and coordinate transformation to obtain normalized sample points, this study implements a refined color-mapping scheme to ensure visual

consistency across all three-dimensional characteristic surfaces. A segmented color mapping scheme (based on the MATLAB parula colormap, MATLAB R2022b) is employed, transitioning from dark blue to bright yellow to represent the normalized values (0 to 1) of the parameters.

2.6. Comparison with Existing Research Findings

The comprehensive characteristic modeling method for Francis's turbines, integrating image digitization and Radial Basis Function (RBF) neural networks, demonstrates transformative advantages through its application to the HLA685 turbine unit. In terms of processing efficiency, the automated sampling via the connected-component recognition algorithm achieves a paradigm shift, extracting 12,205 flow characteristic points and 8692 torque characteristic points, 20 times the speed of manual methods (this acceleration ratio is derived from measured comparisons of processing time between manual annotation and the automated algorithm for the same dataset). This breakthrough marks a significant departure from contemporary research. In contrast to earlier studies constrained by traditional sampling methods [18] or physical sensor dependencies [19], our image digitization pipeline achieves fully automated, industrial-scale data acquisition through connected-component recognition and morphological thinning.

In terms of modeling precision, our physics-informed methodology eliminates pixel-level errors via morphological refinement and enforces boundary integrity through run-away characteristic extrapolation and engineering-guided interpolation in low-opening regions. The resulting RBF model achieves high training accuracy (e.g., a Mean Squared Error of 0.00001) and generates seamless full-characteristic surfaces. This approach outperforms generic algorithms. For instance, in contrast to the generalized sparse RBF method developed by Dai et al. (2024), which attained 92.7% classification accuracy but ignored mechanical continuity requirements [20], our method explicitly incorporates these constraints. Furthermore, it directly resolves the issue of boundary sample scarcity, which has been shown to cause significant performance degradation (e.g., the 12% drop observed by Sun et al. (2022) in wind turbine control [21]), demonstrating its robustness across different turbomachinery domains. The topological consistency of our surfaces validates the synergy between image-derived data density and domain expertise.

The core innovation of our work is its integrated workflow. This process begins with raw data purification through image rotation correction and noise-adaptive bilateral filtering, progresses to high-fidelity sampling via connected-domain recognition and coordinate transformation, and culminates in nonlinear mapping through an enhanced RBF network. In contrast to the cyber-physical preprocessing framework developed by Sai et al. (2021), which prioritized real-time sensor data streams [22], our method introduces gridline-erasure techniques specifically optimized for digitizing historical technical diagrams. This hybrid RBF design—which integrates Gaussian activation functions for feature abstraction with a linear output layer for boundary continuity—effectively eliminates the oscillatory artifacts prevalent in traditional piecewise fitting methods. This capability ensures the numerical stability required for realistic transient simulations in hydropower plants.

Validated in engineering practice, this methodology significantly enhances modeling efficiency and reliability while offering direct transferability to pumps, fans, and other blade machinery. Future work will integrate Shao's sparsification strategies for real-time deployment and expand to transient process characterization, advancing precision modeling capabilities for renewable-integrated power grids.

3. Full Characteristic Modeling Based on RBF Neural Network

3.1. Sample Point Collection

The constant opening line cluster point set, constant efficiency curve cluster point set, and runaway characteristic curve point set are discretely extracted by the image digitization method, and each sample point can be expressed as follows:

$$a_i, N_{11i}, Q_{11i} \quad (i = 1, 2, 3, \dots, U) \quad (4)$$

$$\eta_j, N_{11j}, Q_{11j} \quad (j = 1, 2, 3, \dots, V) \quad (5)$$

$$a_k, N_{11k}, Q_{11k} \quad (k = 1, 2, 3, \dots, W) \quad (6)$$

In Equations (4)–(6), U , V , and W represent the number of sample points for the constant-opening lines, constant-efficiency lines, and runaway characteristic curve, respectively. The collected sample points are shown in Figure 7. In the figure, the green curve represents the sample points of the runaway characteristic curve, indicating the operating parameters of the unit under runaway conditions. The blue curves represent the energy characteristic sample points (i.e., efficiency characteristic sample points), where points on each curve correspond to the same unit efficiency, also known as the constant-efficiency lines. The red curves represent the opening characteristic sample points, where points on each curve correspond to the same opening, also known as constant-opening lines.

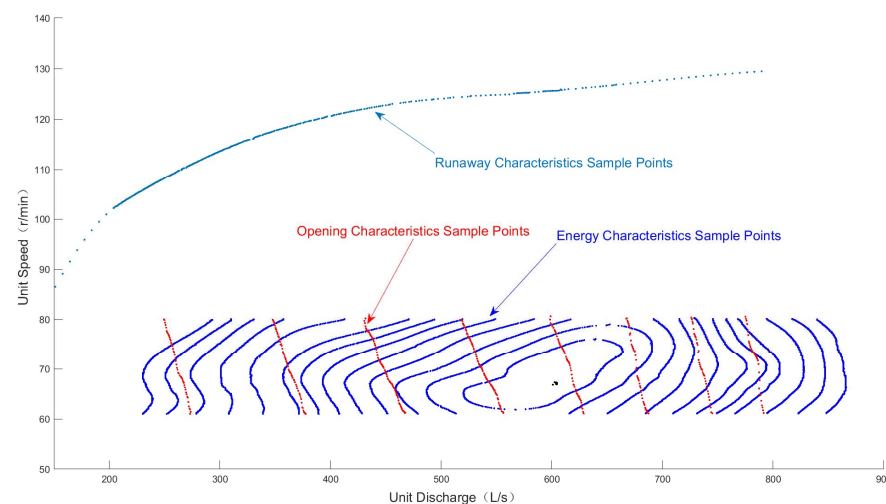


Figure 7. Schematic diagram of collected sample points.

Unit efficiency is a crucial parameter for calculating output torque, which is a nonlinear function of unit speed and discharge, as expressed in Equation (7). The unit efficiency characteristic curve consists of a series of iso-efficiency lines, as shown in Figure 7, exhibiting a contour-like distribution. By treating the unit efficiency value as the “elevation” and unit discharge and unit speed as the base coordinates, surface-fitting interpolation can be performed to obtain the efficiency characteristic surface. The data provided by the manufacturer only includes iso-efficiency lines in the high-efficiency region. It is generally assumed that the unit efficiency is zero during runaway conditions; therefore, the runaway characteristic curve can be used as a boundary condition. The interpolated efficiency characteristic surface is shown in Figure 8, while Figure 9 displays the vertical projection of the efficiency interpolation surface. Combined with Figure 7, the distribution pattern of the iso-efficiency curves can be observed. The red scatter points in the figures represent the collected sample points and the extended-boundary-condition sample points. The

following colormap is qualitative, indicating relative values. Quantitative values should be read from the vertical axis.

$$\eta = f_{\eta}(Q_{11}, N_{11}) \quad (7)$$

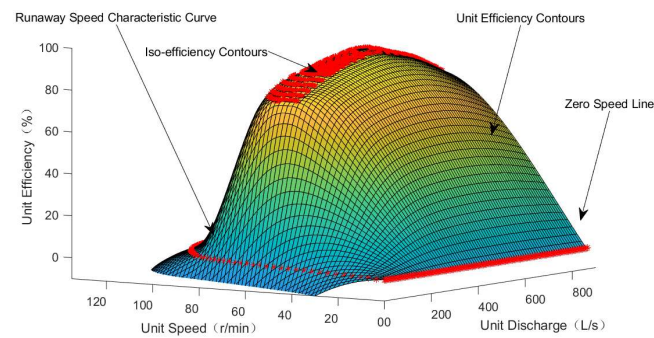


Figure 8. Efficiency characteristic interpolation surface diagram (color indicates normalized unit efficiency from low (blue) to high (yellow)).

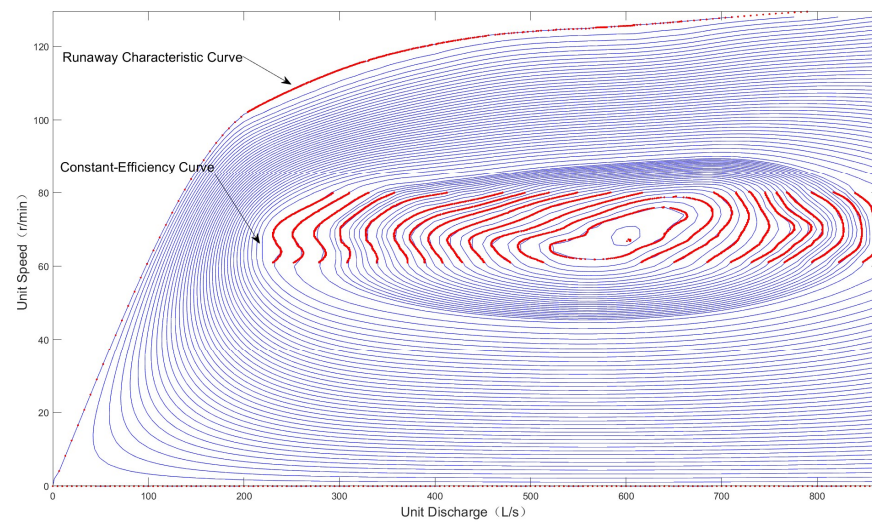


Figure 9. Effect diagram of iso-efficiency curve interpolation.

Bring each point in Equation (4) into Equation (7) to obtain the corresponding efficiency value, and then use Equation (8) to calculate the unit torque value. Note that when the opening is 0, it is generally considered that the unit torque is proportional to the square of the unit speed.

$$M_{11} = \begin{cases} 0.93726 \frac{Q_{11}}{N_{11}} \eta & (a > 0) \\ KN_{11}^2 & (a = 0) \end{cases} \quad (8)$$

In the equation, K is a pre-given negative constant. The torque characteristic sample point set can be obtained through calculation, which is expressed as Equation (9).

$$M_{11i}, a_i, N_{11i} \quad (i = 1, 2, 3, \dots, U) \quad (9)$$

3.2. Sample Point Expansion

The sample points discretely extracted through image digitization methods are primarily concentrated in the high-efficiency zone. Therefore, it is necessary to supplement operating conditions in regions with low gate openings and low rotational speeds.

(1) Expansion of flow characteristic sample points. The runaway characteristic curve marks the runaway points under some openings, and the point set in Equation (6) can be added to the sample point set in Equation (5). At the same time, when the opening is 0,

regardless of the unit speed, the unit discharge is 0. Therefore, the speed sequence N_{11z} ($z = 1, 2, 3, \dots, R$) can be taken to form a zero-opening sample point set, which is merged into Equation (5), and the corresponding representation is as follows:

$$a_z = 0, N_{11z}, Q_{11z} = 0 \quad (z = 1, 2, 3, \dots, R) \quad (10)$$

$$a_i, N_{11i}, Q_{11i} \quad (i = 1, 2, 3, \dots, U + Wa + R) \quad (11)$$

Formula 10 is the zero-opening sample point set, and N_{11z} generally takes equally spaced data points between the minimum and maximum speeds of the known runaway line. Equation (11) is the expanded flow characteristic sample point set, and Wa is the number of intersection points between the constant opening line and the runaway characteristic curve.

(2) Expansion of the torque characteristic sample points. In practice, Equation (5) does not include the zero-opening point. Drawing on engineering simulation experience for hydraulic units, Equation (8) is used to calculate the torque characteristics at zero opening, which can be expressed in Equation (12).

$$M_{11l} = KN_{11l}, a_l = 0, N_{11l} \quad (l = 1, 2, 3, \dots, L) \quad (12)$$

In addition, when the unit runs away, the efficiency is 0, and the output torque to the outside is 0. The expanded runaway characteristic curve sample points can be used to construct zero-torque characteristics, which can be expressed by Equation (13).

$$M_{11o} = 0, a_o, N_{11o} \quad (o = 1, 2, 3, \dots, W) \quad (13)$$

The expanded torque characteristic points can be expressed as follows:

$$M_{11m}, a_m, N_{11m} \quad (m = 1, 2, 3, \dots, U + W + L) \quad (14)$$

After expansion, 12,205 sets of flow characteristic sample points and 8692 sets of torque characteristic sample points were obtained. Table 2 compares the processing efficiency, sample size, and precision among conventional manual graph reading, software tracing, and the method proposed in this study, demonstrating that the proposed approach achieves a larger sample size, higher precision (due to automated pixel-level data extraction), and superior processing efficiency compared to manual or software-based methods, as summarized in Table 2 (comparison of sample point acquisition methods).

Table 2. Comparison of sample point acquisition methods.

Method	Manual Graph Reading	Software Tracing	Proposed Method
Processing Efficiency	Relatively low	Faster than manual extraction	High
Sample Size	234 sets total	576 sets total	20,897 sets total
Sample Precision	Low (typically to two decimal places)	Relatively high (human-induced errors during tracing)	High

3.3. Construction of RBF Neural Network Model

The complete characteristic curves of hydraulic turbines exhibit highly nonlinear behavior. Artificial neural networks (ANNs) are well-suited for modeling such nonlinearities, as they can, in theory, approximate any continuous function with arbitrary accuracy without requiring explicit mathematical relationships between the parameters. Furthermore,

by leveraging specific activation functions in the output layer, ANNs can effectively overcome the boundary discontinuity issues commonly associated with traditional piecewise fitting and interpolation methods, thereby ensuring convergence in simulations of large-fluctuation transient processes. Among the commonly used ANNs, the Backpropagation (BP) and Radial Basis Function (RBF) networks are both powerful nonlinear approximators. However, they possess distinct advantages and disadvantages, as summarized in Table 3 for a direct comparison. The BP network is often criticized for its slow training convergence and tendency to become trapped in local minima. In contrast, the RBF network generally features a simpler architecture, faster training speed, and superior generalization capability. Based on this comparative analysis, the RBF neural network algorithm is selected for this study to acquire the full characteristic data.

Table 3. Performance comparison between BP and RBF neural networks.

Feature	BP Neural Network	RBF Neural Network
Training Speed	Slow	Fast
Convergence Reliability	Prone to local minima	More reliable
Architectural Complexity	Complex	Simpler
Generalization Ability	Standard	Stronger

The structure of an RBF network generally consists of an input layer, a hidden layer, and an output layer. The input layer, composed of input variable nodes, serves only to transmit data. The hidden layer contains a series of radial basis function neurons that perform a spatial mapping transformation on the input information. The output layer typically consists of linear kernel function neurons, which linearly weight the outputs of the hidden layer neurons to produce the final result. In the simulation application of Francis turbines, the guide vane opening (a) and unit speed (N_{11}) are used as inputs, while the unit flow (Q_{11}) and unit torque (M_{11}) are used as outputs to establish separate network models. The network structure is shown in Figure 10.

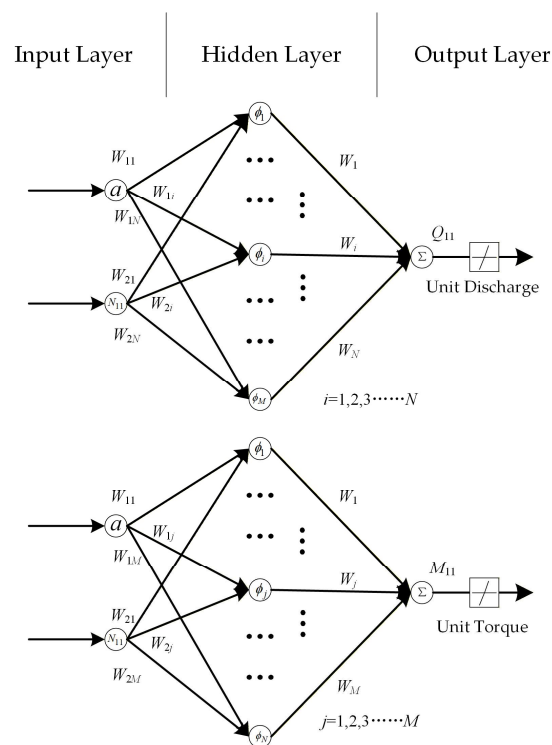


Figure 10. Topology of the radial basis function neural network.

The Gaussian function is uniformly adopted as the activation function in the hidden layer, as shown in the following equation. In the equation, c_i and σ_i represent the center and width parameters of the kernel function for the i -th neuron, respectively, and x denotes the input value.

$$\Phi(x) = e^{-\frac{(x-c_i)^2}{2\sigma_i^2}} \quad (15)$$

The output layer function uses a simple linear function, as shown in the following equation.

$$y = \sum_{i=1}^N W_i \Phi_i + b_i \quad (16)$$

In the equation, W_i , Φ_i , and b_i represent the output weight, output value, and bias of the i -th neuron, respectively, and y represents the network output value.

The collected sample points were randomly divided into training (70%), validation (15%), and prediction (15%) sets according to a predetermined ratio. To ensure numerical stability and accelerate convergence during network training, all input and output variables were normalized to a common scale. Given the well-defined operating ranges of the turbine parameters, min-max normalization was applied to linearly transform the data into the $[0, 1]$ interval. The normalization formula is defined as follows:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (17)$$

where x represents the original data (including guide vane opening, unit speed, unit discharge, and unit torque), and x_{min} and x_{max} are the minimum and maximum values of the respective variables, determined from the expanded sample set. This preprocessing step mitigates the potential dominance of variables with larger numerical magnitudes during RBF network training. After normalization, the datasets were input into the RBF neural network for simulation. Network modeling was implemented using the Neural Network Toolbox in MATLAB, where a radial basis network was created with the `newrb` function (P , T , goal, spread). Here, P is the input matrix, consisting of sample points formed by guide vane opening and unit speed; T is the target matrix, representing the unit flow or unit torque vector; goal is the mean squared error target, where a smaller value indicates higher fitting accuracy but also results in longer computation time; and spread is the spread parameter of the radial basis function. A smaller spread yields more detailed fitting effects but may lead to overfitting if too small, while a larger spread enhances the network's generalization ability but significantly reduces fitting accuracy if excessively large.

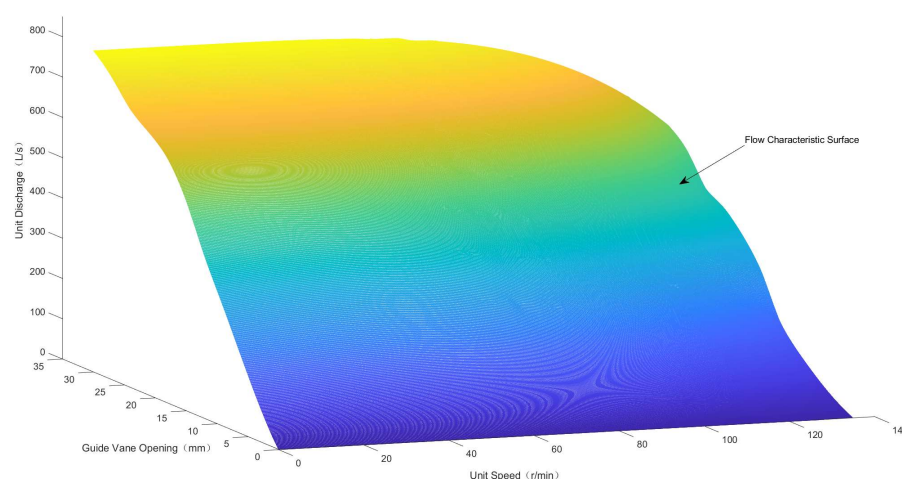
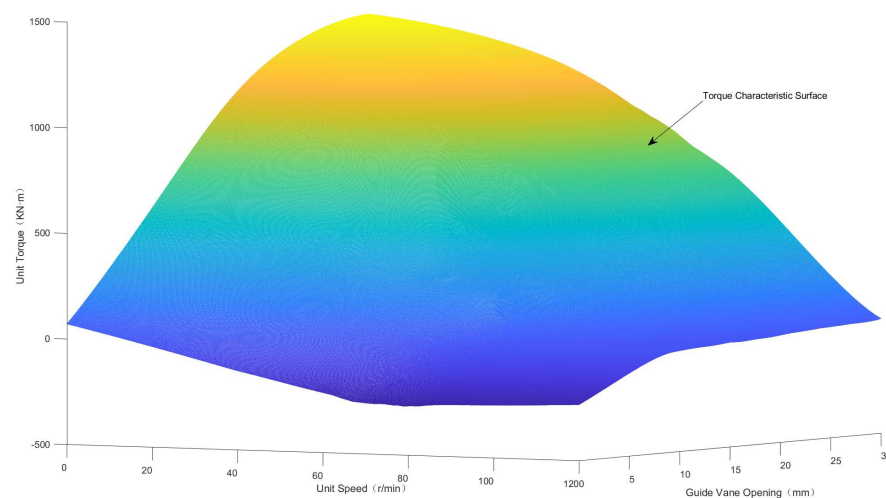
Based on the high-precision requirements of the simulation and parameter tuning tests, the parameter sets (goal = 1×10^{-6} , spread = 3) and (goal = 1×10^{-6} , spread = 1.5) were selected for training. The training results are shown in Table 4. It can be seen from the table that the prediction error levels of both the unit flow and unit torque network models meet the design requirements, and the number of training iterations is relatively low, demonstrating that the RBF neural network converges quickly.

To ensure the reproducibility of the RBF models, the key structural and performance metrics are documented. The `newrb` function automatically determines the optimal number of radial basis neurons (centers) required to achieve the prescribed mean squared error goal. For the unit flow network, the final architecture comprised 215 neurons in the hidden layer, while the unit torque network required 189 neurons. The training process was highly efficient; the unit flow network converged in approximately 4.5 s, and the unit torque network in 3.2 s, on a standard workstation equipped with an Intel Core i7 processor and 16 GB RAM manufactured by Lenovo (Beijing, China). These details, combined with the parameters in Table 3, provide a complete reference for replicating the study.

Table 4. RBF neural network training results.

	Unit Flow Network			Unit Torque Network		
	Training	Validation	Testing	Training	Validation	Testing
Number of Samples	8543	1831	1831	6084	1304	1304
MSE	1.66×10^{-6}	1.70×10^{-6}	1.95×10^{-6}	2.16×10^{-6}	3.02×10^{-6}	2.94×10^{-6}
R ²	0.9999	0.9997	0.9998	0.9999	0.9998	0.9996
Iterations	6	-	-	7	-	-

By taking equally spaced sequences of values covering the entire operating range of the guide vane opening and unit speed, and using the trained model for predictive interpolation, the unit flow and unit torque characteristic surfaces of the unit can be obtained, as shown in Figures 11 and 12, respectively. In these figures, a segmented color map is used, which transitions from dark blue (indicating low values) to bright yellow (indicating high values) to represent the normalized values of unit discharge (Q_{11}) and unit torque (M_{11}), respectively. For engineering applications, the trained model can predict unit flow and unit torque at any operating condition, thereby enabling comprehensive modeling of the unit's full-range characteristics.

**Figure 11.** Full characteristic surface diagram of flow (color indicates normalized unit efficiency from low (blue) to high (yellow)).**Figure 12.** Full characteristic surface diagram of torque (color indicates normalized unit efficiency from low (blue) to high (yellow)).

The negative torque area shown in Figure 12 corresponds to the Braking Zone and Reverse Pump Zone of the hydraulic turbine. When the unit speed is high and the guide vane opening is small, the propulsive effect of the water flow on the runner weakens, and the runner may even experience resistance. This requires the runner to consume mechanical energy to maintain rotation, resulting in a negative torque value.

This accurately reflects the real physical states that a hydraulic turbine may experience during transient processes (such as start-up or load rejection), which is an important manifestation of the model's accuracy.

4. Conclusions

This study addresses the limitations of conventional sampling methods—low efficiency and susceptibility to human error—and the restricted modeling accuracy of traditional algorithms due to small sample sizes in the comprehensive characteristic modeling of Francis turbines. An integrated methodology combining image digitization and an RBF neural network is proposed. Using the HLA685 turbine unit as a case study, the approach automates large-scale characteristic data extraction via image processing, incorporates engineering expertise to ensure reasonable boundary extrapolation, and employs an RBF neural network to achieve high-accuracy prediction of unit discharge and unit torque across full operating characteristics. Practical applications demonstrate that this method significantly enhances sampling efficiency, scale, and precision, effectively improving the reliability of the neural network model. It thus presents an effective solution for the comprehensive characteristic modeling of Francis turbines and other blade machinery. It should be explicitly acknowledged that the sample expansion in the low-opening and low-speed regions, while based on established engineering judgment and simulation experience, represents an extrapolation beyond the manufacturer-provided data. Future work will focus on validating the model's performance in these critical regions through computational fluid dynamics (CFD) simulations or experimental measurements to further enhance the model's physical fidelity and generalizability.

Author Contributions: Writing—original draft preparation, Y.D.; resources, Y.L. (Youping Li); writing—review and editing, X.H. and J.G.; data curation, Y.L. (Yushu Li) and R.L.; software, Y.L. (Yushu Li) and Y.G.; validation, L.W., W.Y. and Y.L. (Yushu Li); methodology, Y.G.; formal analysis, F.Z.; project administration, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the China Yangtze Power Co., Ltd. project “Impact of wind and solar access on power system and power plant operation and simulation study” (No. Z242302010) and the Key R&D Program of Hubei Province “Research on Key Technologies for Integrated Hydro-Wind-Solar Multi-Energy Complementarity” (No. 2022AAA007).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Youhan Deng, Yushu Li and Weiwei Yao were employed by the company China Three Gorges Corporation. Authors Youping Li, Xiaojun Hua, Rui Lyu and Lei Wang were employed by the company China Yangtze Power Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that this study received funding from China Yangtze Power Co., Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

- Men, C.S.; Nan, H.P. Improvement of Internal Characteristic Model of Francis Turbine and Its Application in the Extension of External Characteristic Curves. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 58–66. [\[CrossRef\]](#)
- Zhu, Y.P.; Shi, X.Y.; Zhou, L.J. Complete Composite Characteristic Curves of Hydraulic Turbines Based on Internal Characteristic Method. *J. China Agric. Univ.* **2006**, *11*, 88–91.
- Ma, W.C.; Yang, J.B.; Zhao, Z.G.; Yang, W.J.; Yang, J.D. Zoning Method for Characteristic Curves of Francis Turbine under Multiple Boundary Conditions. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 31–39.
- Zhang, R.S.; Liu, Z.; Wang, L.C.; Zhang, Y.C. Data Interpolation of Turbine Composite Characteristic Curves Based on Delaunay Triangulation. *J. Hydroelectr. Eng.* **2011**, *30*, 197–201+221.
- Wang, Y.H.; Shen, Z.Y.; Sun, Y. Numerical Fitting of Turbine Characteristic Curves Based on Principal Curves Method. *J. Hydroelectr. Eng.* **2009**, *28*, 181–186+151.
- Liu, D.; Hu, X.; Zeng, Q.; Zhou, H.K.; Xiao, Z.H. A Refined Model of Turbine Characteristic Curves Based on Input-Output Correction. *J. Hydraul. Eng.* **2019**, *50*, 555–564.
- Sun, Z.; Yao, Q.; Jin, H.; Xu, Y.; Hang, W.; Chen, H.; Li, K.; Shi, L.; Gu, J.; Zhang, Q.; et al. A novel in-situ sensor calibration method for building thermal systems based on virtual samples and autoencoder. *Energy* **2024**, *297*, 131314. [\[CrossRef\]](#)
- Wu, Q.Q.; Zhang, L.K.; Ma, Z.Y. Extension and Reconstruction of Hydraulic Turbine Composite Characteristic Curves Based on Engineering Experience and RBF Neural Network. *J. Basic. Sci. Eng.* **2019**, *27*, 996–1007.
- Li, J.Y.; Chen, Q.J.; Chen, G.D. Research on Fitting Method of Turbine Composite Characteristic Curves Using BP Neural Network. *J. Hydroelectr. Eng.* **2015**, *34*, 182–188.
- Zhu, Y.; Luo, J.; Pan, K. Reconstruction of Integrated Hydraulic Turbine Characteristics Curve Based on Classification Weight Neural Network. *Energy Rep.* **2023**, *9*, 884–896. [\[CrossRef\]](#)
- Wang, H.Y.; Yang, H.T.; Wang, J.Y.; Zhou, X.X.; Zhang, H.G.; Xu, Y.F. A Survey of Remote Sensing Image Denoising Methods. *Comput. Eng. Appl.* **2024**, *60*, 55–65. [\[CrossRef\]](#)
- Li, L.; Xu, P.; Li, Q.; Zheng, R.; Xu, X.; Wu, J.; He, B.; Bao, J.; Tan, D. A coupled LBM-LES-DEM particle flow modeling for microfluidic chip and ultrasonic-based particle aggregation control method. *Appl. Math. Model.* **2025**, *143*, 116025. [\[CrossRef\]](#)
- Sehgal, R.; Kaushik, D.V. ST Bilateral-Deep Filter: Shearlet Transform Based Bilateral Filter and Deep Learning Approach for Noise Reduction in CT Images. *Int. J. Bio-Inspired Comput.* **2025**, *25*, 226–238. [\[CrossRef\]](#)
- Qi, X.; Chen, J.F.; Luo, W.L. FPGA Implementation of Image Preprocessing Algorithm. *Foreign Electron. Meas. Technol.* **2021**, *40*, 102–107.
- Li, X.; Zheng, X.; Song, T.; Fang, H.; Jiang, K. Automatic Extraction Method of Microtremor Surface-Wave Dispersion Curves Based on Connected Domain Analysis. *Oil Geophys. Prospect.* **2025**, *60*, 679–688.
- Zhang, Y.; Xia, Y.F.; Wang, Z.Q.; Liu, Y. Light Stripe Center Extraction Method Based on Skeleton Adjacent Pixel Matching. *J. Northeast. Univ. Nat. Sci.* **2023**, *44*, 1529–1536.
- Liu, R.; Liu, H.; Sun, X.; Chen, Z. Automatic Recognition Method for RQD Value of Core Images Based on Connected Components. *Min. Metall. Explor.* **2025**, *42*, 705–717. [\[CrossRef\]](#)
- Zhang, P.; Chen, G.D.; Zhang, X. Comparison of BP and RBF Neural Networks in Fitting Nonlinear Characteristics of Hydraulic Turbines. *China Rural. Water Hydropower* **2011**, *11*, 125–128+131.
- Babes, B.; Boutaghane, A.; Reddafi, A.; Boudjerda, M.; Amar, H.; Hamouda, N.; Ghoneim, S.S. An RBF neural network-based parameter tuning for an ADRC regulator of electrode wire feed mechanism: Arc welding Applications. *Weld. World* **2024**, *68*, 987–999. [\[CrossRef\]](#)
- Dai, Y.; Wu, Q.; Zhang, Y. Generalized sparse radial basis function networks for Multi-classification Problems. *Appl. Soft Comput.* **2024**, *154*, 111361. [\[CrossRef\]](#)
- Sun, C.; Wang, Z.; Wang, Y.; Sun, H.; Liu, T. Flutter Suppression of Wind Turbine Blade Based on RBF Neural Network Compensation Backstepping Control. *J. Phys. Conf. Ser.* **2022**, *2173*, 012030. [\[CrossRef\]](#)
- Sai, B.K.; Mayer, Y.T.; Bauernhansl, T. Dynamic Data Acquisition and Preprocessing for Automatic Behavioral Modeling of Cyber-physical Systems. *Procedia CIRP* **2021**, *104*, 363–369. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.