Jestr

Journal of Engineering Science and Technology Review 16 (5) (2023) 91 - 101

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Modeling and Uncertainty Analysis of Wind Turbine Power Curve Based on Improved Sparrow Search Algorithm and Least Square Support Vector Machine

Hao Hu[1, 2], Junpeng Chen[1], Huazheng Wang[3], Zichao Zhang[4] and Bo Gu[1,*]

[1]*North China University of Water Resources and Electric Power, Zhengzhou 450045, China*
[2]*Yellow River Conservancy Technical Institute, Kaifeng 475004, China*
[3]*Sinohydro Bureau 11 Co., Ltd., Zhengzhou 450001, China*
[4]*Texas Tech University, Texas 79409, USA*

Received 30 June 2023; Accepted 5 September 2023

Abstract

Accurately calculating the distribution characteristics of the real operating power curve of wind turbines is a prerequisite for evaluating the operating performance of these turbines. To construct a high-precision calculation model for this real power curve, a method for modeling and uncertainty analysis was developed based on the improved sparrow search algorithm (ISSA) and least square support vector machine (LSSVM). ISSA was used to optimize the penalty factor and kernel function width of the LSSVM, the modeling accuracy of LSSVM was improved, and the real power curve model based on ISSA-LSSVM was constructed. The error distribution characteristics between the calculated value of the ISSA–LSSVM model and the real value of power curve were then analyzed. The probability density function of error distribution characteristics was calculated using the non-parametric kernel density estimation (NPKDE) method. Based on the calculated function, the confidence intervals at different confidence levels were calculated, and the uncertainty distribution range of the true power curve of the wind turbine was quantitatively analyzed. Results show that the calculation accuracy of the ISSA-LSSVM model in the real power curve modeling of wind turbines is higher than that of the PSO-LSSVM, PSO-BP, LSSVM, and LSTM models. The NPKDE method can accurately calculate the uncertainty distribution characteristics of the calculation errors of the proposed power curve model. This study provides a certain reference for the high-precision modeling of the real power curve of wind turbines.

Keywords: Wind turbine; Power curve modeling; Sparrow search algorithm; Least squares support vector machine; Uncertainty analysis

1. Introduction

The operating environment of in-service wind turbines often differs significantly from their designed environment, leading to discrepancies between their real and designed power curves. These discrepancies introduce challenges in analyzing the operating state, predicting the power output, and optimizing the control of these wind turbines [1-3]. Therefore, accurately constructing real power curve models for in-service wind turbines is a key issue in optimizing their operation. Scholars all over the world have conducted an extensive modeling and uncertainty analysis of wind turbine real power curves and achieved notable results that provide a foundation for further study in this area [4-6].

The modeling and uncertainty analysis of wind turbine power curves are often based on various machine learning methods to achieve an accurate mapping between the operating environment and output power of wind turbines. For example, local-regression-based power curve modeling takes wind speed, wind direction, air density, and pitch angle as input data to construct the power curve, and several examples have demonstrated the effectiveness of this approach [7]. Monotonic-regression-based power curve modeling can effectively remove the influence of data noise on modeling accuracy, and previous studies show that this approach outperforms other models in terms of accuracy [8]. In analyzing the uncertainty of wind turbine power curve models, normal statistical methods are typically used to examine the distribution characteristics of model calculation errors [9]. However, the model calculation errors are influenced by meteorological factors, modeling methods, and turbine operation characteristics, resulting in deviations from normal distribution. Thus, developing accurate uncertainty analysis methods remains a focus of power curve modeling research [10].

The existing power curve modeling methods for wind turbines mainly consider the mapping relationship between wind speed and power curve and often use normal distribution, quantile regression, and Gaussian regression to calculate the uncertainty of power curve models. However, due to the coupling of multiple factors, wind turbine power curve models exhibit high nonlinearity between dependent and independent variables, and the model calculation errors do not follow certain distribution characteristics, hence posing significant challenges for accurately modeling wind turbine power curves.

Scholars have extensively investigated nonlinear modeling methods for wind turbine power curves and the uncertainty distribution characteristics of model calculation errors [12-13]. However, several issues remain in the calculation accuracy of wind turbine power curve models and the quantitative analysis of model uncertainty, which lead to deviations from real power curves. Therefore, constructing high-precision nonlinear models for wind turbine power curves and quantitatively analyzing the distribution characteristics of model calculation errors remain urgent problems that need to be solved.

*E-mail address: gb19820915@163.com
ISSN: 1791-2377
doi:10.25103/jestr.165.12

To address these gaps, this study proposes ISSA-LSSVM, a real power curve modeling method for wind turbines based on the improved sparrow search algorithm (ISSA) and the least squares support vector machine (LSSVM) model. This method optimizes the penalty factor and kernel function width of LSSVM by using ISSA to improve its modeling accuracy. A power curve model based on ISSA–LSSVM is also constructed, based on which the non-parametric kernel density estimation (NPKDE) method is used to calculate the probability density function of the error distribution. The uncertainty range of the real power curve of wind turbines is also quantitatively analyzed to formulate a new strategy for the high-precision modeling of wind turbine power curves.

2. State of the art

The nonlinear modeling and uncertainty analysis of wind turbine power curves have received much scholarly interest. For instance, Han et al. [14] proposed a wind–power curve fitting method for wind turbines based on generalized least squares and analyzed the uncertainty distribution characteristics of the power curve model by using the normal distribution model. However, this method only considers the influence of wind speed on the power curve and ignores the influence of other factors. Li et al. [15] used full longitudinal filtering and segmented filtering to remove the abnormal points from the operating data and enhance the accuracy of the constructed power curve. However, they did not analyze any methods for the modeling and uncertainty analysis of power curves. Cao et al. [16] used higher-order polynomials and logistic functions to model wind–power curves and analyzed the distribution characteristics of model calculation errors by using the normal distribution method, but their modeling accuracy needs to be further improved. Alessandro et al. [17] constructed a wind turbine power curve using a multivariate polynomial regression method that accounts for wake effects, thus further improving its modeling accuracy. However, this method does not consider model uncertainty or the influence of wind direction, pitch angle, and other factors on the power curve. Yves et al. [18] built a wind turbine power curve model that fully considers environmental factors and the operation characteristics of wind turbines and then analyzed the distribution characteristics of the model calculation errors via quantile regression. However, the quantile regression method could not accurately describe the distribution characteristics of errors.

The statistical method is commonly used in the nonlinear modeling of wind turbine power curves. Yan et al. [19] built a power curve model for wind turbines using statistical methods and analyzed model uncertainty by using a quantile regression method. However, they only considered the influence of wind speed, wind direction, and pitch angle on power curve modeling and ignored the other potential influencing factors. Yun et al. [20] used Weibull distribution parameters and Monte Carlo simulation methods to construct a wind turbine power curve model. They analyzed the uncertainty distribution characteristics of the model using normal distribution methods and verified the model accuracy through numerical examples. However, the normal distribution method cannot easily achieve an accurate description of the uncertainty distribution characteristics of model calculation errors. Rogers et al. [21] used the heteroscedasticity Gaussian process (GP) to build a wind turbine power curve model and analyzed the uncertainty of this model, but the calculation accuracy of the model needs to be further improved. Bull et al. [22] used a probabilistic regression model to construct a wind turbine power curve that effectively eliminates the influence of noise on modeling accuracy. However, their approach lacks an analysis of model uncertainty. Fu et al. [23] proposed a wind turbine power curve modeling method based on correlation vector information entropy and analyzed the uncertainty distribution characteristics of the model using the quantile regression method. They verified the feasibility of this method through examples, but they did not comprehensively analyze the influencing factors of the power curve. Ravi et al. [24] proposed a power curve modeling method and uncertainty analysis method based on the GP and analyzed the wind turbine state that affects the accuracy of the power curve model. However, the GP cannot accurately describe the distribution characteristics of model calculation errors.

The machine learning method is mainly used in the nonlinear modeling of wind turbine power curves. Mehrdad et al. [25] proposed a power curve modeling method for wind turbines based on the weighted equilibrium loss function and analyzed the uncertainty of the model using the normal distribution method but did not comprehensively consider the influencing factors of the power curve. Moreover, their normal distribution analysis did not accurately describe the uncertainty distribution characteristics of the model. Zha et al. [26] used the improved Transformer network to build a wind turbine power curve and verified the progressiveness of their power curve modeling method for wind turbines from the perspective of modeling error and its distribution characteristics. However, their modeling method did not consider the impact of the wind turbine parameters on power curve modeling. Gustavo et al. [27] combined GP regression, probability-kernel-based machine learning models, and standard logic functions to construct a power curve model for wind turbines, but they did not analyze the uncertainty of the model. Li et al. [28] integrated isolation forest, asymmetric fuzzy mean, radial basis function neural network (RBFNN), and meta-heuristic algorithm to construct a power curve model of wind turbine and analyzed the uncertainty distribution characteristics of the model by quantile. However, they ignored the influence of meteorological and geographical factors on the modeling accuracy of the power curve. Despina et al. [29] proposed a power curve modeling method for wind turbines based on RBFNNs and used statistical analysis to analyze the uncertainty distribution characteristics of the model. However, they did not analyze those factors that may influence the power curve model. Bartolome et al. [30] proposed a power curve modeling method for wind turbines and an uncertainty analysis method based on the GP and artificial neural network. However, they did not provide a detailed analysis of those factors that may influence the power curve modeling process.

Previous studies have focused on modeling and uncertainty analysis methods for wind turbine power curves, but high-precision models for wind turbine power curves, especially precise uncertainty analysis methods for power curve models, remain lacking. This study then proposes a power curve modeling method for wind turbines based on the ISSA–LSSVM model. This method uses correlation analysis to identify the key factors affecting the power curve modeling of wind turbines and constructs a power curve model for wind turbines based on ISSA–LSSVM using these key factors as model inputs. A non-parametric kernel density

estimation method (NPKDE) is also used to quantitatively analyze the uncertainty distribution characteristics of the constructed power curve model.

The rest of this study is organized as follows. Section 3 provides a detailed introduction to the principles of sparrow search algorithm (SSA), LSSVM, and NPKDE method. Based on these principles, a power curve model for wind turbines is constructed based on the ISSA–LSSVM model. Section 4 verifies the accuracy and superiority of the ISSA-LSSVM model in modeling wind turbine power curves by analyzing the error distribution characteristics between the calculated values of the ISSA–LSSVM model and the true values of wind turbine power within different wind speed zones. Section 5 summarizes the paper and presents relevant conclusions.

3. Methodology

3.1 Principle of SSA

SSA is a new swarm intelligence optimization algorithm proposed in 2020 that was inspired by sparrows' foraging and anti-predation behaviors [31-32]. The SSA has the advantages of high optimization ability, convergence speed, stability, and robustness and has been widely studied and applied in solving high-dimensional and complex optimization control problems. The mathematical model of SSA is expressed as follows:

(1) Suppose that the population with N sparrows is $X(x_1,x_2,...x_i,...x_N)$, where i=1, 2, ..., N. The position and fitness value of the sparrow in the d-dimensional space can be described by Eqs. (1) and (2).

$$X=\begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \cdots & \vdots \\ x_N^1 & x_N^2 & \cdots & x_N^d \end{bmatrix} \tag{1}$$

$$F_x=\begin{bmatrix} f_1(x_1^1,\ x_1^2,\ \cdots, x_1^d) \\ f_2(x_2^1,\ x_2^2,\ \cdots, x_2^d) \\ \vdots \\ f_N(x_N^1,\ x_N^2,\ \cdots, x_N^d) \end{bmatrix} \tag{2}$$

where d represents the dimension of the optimization variable, N is the number of sparrow population, and f_i is the fitness value of the sparrow.

(2) In SSA, the producers with good fitness values will prioritize obtaining food during the search process. In addition, given that the producers are responsible for finding food for the entire sparrow population and providing foraging directions for all scroungers, they have a larger range of foraging searches than scroungers. During each iteration, the location update of the producers can be calculated using Eq. (3).

$$x_{i,j}^{t+1}=\begin{cases} x_{i,j}\cdot\exp\left(-\dfrac{i}{\alpha\cdot \mathrm{Iter}_{\max}}\right), & \text{if } R_2<ST \\ x_{i,j}+Q\cdot L, & \text{if } R_2\geq ST \end{cases},\ \alpha\in(0,\ 1] \tag{3}$$

(3) During foraging, scroungers continuously monitor producers. Upon detecting that the producers have located superior food sources, the scroungers will immediately leave their current position to compete with the producers for food. The successful scroungers immediately acquire the food of producers as described by Eq. (4).

$$x_{i,j}^{t+1}=\begin{cases} Q\cdot\exp\left(\dfrac{x_{\mathrm{worst}}-x_{i,j}^t}{i^2}\right), & \text{if } i>N/2 \\ x_{\mathrm{p}}^{t+1}+\left|x_{i,j}-x_{\mathrm{p}}^{t+1}\right|\cdot A^+\cdot L, & \text{otherwise} \end{cases} \tag{4}$$

where x_{p} is the optimal position occupied by the current producers, x_{worst} represents the worst position of the current sparrow population, and A is the $1\times d$ dimension vector, where each element in A is randomly assigned a value of 1 or -1, and $A^+=A^T(AA^T)^{-1}$. When $i>N/2$, the i-th scrounger with a lower fitness value has not obtained any food and is in a very hungry state. At this time, the i-th scrounger needs to fly to other places to forage for food in order to obtain energy.

(4) In SSA, the number of sparrows where predators are found is generally assumed to account for 10% to 20% of the total population. A sparrow that finds a predator can update its position according to Eq. (5).

$$x_{i,j}^{t+1}=\begin{cases} x_{\mathrm{best}}^t+\beta\cdot\left|x_{i,j}^t-x_{\mathrm{best}}^t\right|, & \text{if } f_i>f_g \\ x_{i,j}^t+K\cdot\left(\dfrac{x_{i,j}^t-x_{\mathrm{worst}}^t}{(f_i-f_w)+\varepsilon}\right), & \text{if } f_i=f_g \end{cases} \tag{5}$$

where x_{best}^t is the global optimal position of the sparrow population, β is a step size control parameter, and β follows a random number of normal distribution with a mean of 0 and variance of 1. K is the direction and step size control parameter of sparrow movement, and K is a random number between [-1, 1]. f_i is the fitness value of a sparrow individual, and f_{g} and f_{w} are the global optimal and worst fitness values of the sparrow population, respectively. When $f_i>f_g$, the sparrow is at the edge of population and is vulnerable to predators' attacks. When $f_i=f_g$, the sparrow in the middle of the population, discovers predators, and needs to approach other sparrows to reduce risk. ε is the smaller constant values to avoid zero in the denominator.

3.2 ISSA based on multi-strategy fusion

Although the basic SSA has strong global search ability and fast convergence speed in high-dimensional optimization problems, this method can easily fall into the local optimal solution in the late iteration. Therefore, the Sin chaos model, Cauchy variation, and hyperparameter adaptive adjustment are introduced to enhance the global optimization ability of SSA.

(1) Initializing sparrow population using the Sin chaos model

The initial distribution state of the sparrow population has a great influence on the optimal result of the population. The more dispersed the sparrow population is in the solution space, the more likely it is to reach the global optimal solution. In this study, the Sin chaotic model is used to initialize the sparrow population. The Sin chaotic model has an unlimited number of map folds with good chaotic characteristics and can improve the dispersion degree of the

initial population of sparrows. The Sin chaos model is shown in Eq. (6).

$$\begin{cases} x_{n+1} = \sin\left(2 / x_n\right) & n = 0,1,\ldots,N \\ -1 \le x_n \le 1 & x_n \ne 0 \end{cases} \tag{6}$$

where x_n is a d dimension vector.

(2) Cauchy mutation in the process of scrounger position updating

During the search process, scroungers often forage around the best positioned producers, which may lead to the entire sparrow population falling into a local optimal solution. Therefore, Cauchy mutation is used to perturb the scroungers' position update process, thereby expanding the scroungers' search range and improving the ability of SSA to jump out of the local optimal solution. The Cauchy mutation strategy for the scrounger position updating process is shown in Eq. (7).

$$x_{i,j}^{t+1} = x_{\mathrm{p}}^{t+1} + \mathrm{C}\,\mathrm{auchy}(0,1) \oplus x_{\mathrm{p}}^{t+1} \tag{7}$$

where Cauchy(0, 1) is the standard Cauchy distribution function, and $\oplus$ represents multiplication.

(3) Hyperparameter adaptive adjustment of the producer's position updating process

The sparrow population is divided into producers and scroungers, with the producers obtaining food by searching the space and the scroungers searching for food based on the location of the producers. Therefore, the optimization capability of SSA is closely related to the producers' search scope. To enhance the optimization ability of SSA, the search space of the producers should be increased. The hyperparameter adaptive adjustment strategy is introduced into the producer position updating Eq. (3) to increase the producer search space. The producer position updating formula with hyperparameter adaptive adjustment is shown in Eq. (8).

$$x_{i,j}^{t+1} = \begin{cases} w = w_0 \cdot c^t \\ x_{i,j} \cdot \exp\left(-\dfrac{i}{w \cdot \alpha \cdot Iter_{\max}}\right), & \text{if } R_2 < ST \quad , \alpha \in (0,\,1] \\ x_{i,j} + Q \cdot L, & \text{if } R_2 \ge ST \end{cases} \tag{8}$$

where w_0 is the initial weight value, and c is the adaptive factor, whose size can be set according to the actual problem. By selecting appropriate initial weights w_0 and adaptive factor c, the search range of producers can be changed, and the search space of the producers can be increased.

3.3 LSSVM algorithm principle

To address the difficulty in solving inequality-constrained optimization in SVM, Suykens et al. [33] proposed the LSSVM optimization algorithm that converts the inequality constraints in the SVM algorithm into equality constraints and achieves the final objective function by solving a linear equation system, thereby reducing the difficulty of problem solving to a certain extent and increasing the solving speed.

The goal of the SVM optimization algorithm is to find an optimal hyperplane $Wx+b=0$ so that the input samples can be correctly separated by the hyperplane. When there are M n-dimensional training samples $\{x_k, y_k\}, y_k \in \{-1,1\}$, $x_k \in R^n$, k=1, 2, 3, ..., M, the objective function and inequality constraints of the SVM algorithm are shown in Eqs. (9) and (10).

$$\min_{W,b,\zeta} J_P(W,\zeta) = \frac{1}{2} W^T W + c\sum_{k=1}^{M} \zeta_k \tag{9}$$

$$\text{s.t.} \quad y_k = W^T \varphi(x_k) + b + \zeta_k \tag{10}$$

Where $\xi_k \ge 0$ is the relaxation variable that represents the loss function in the support vector, c is the penalty factor, W is the normal vector of the hyperplane, $\varphi(x_k)$ is the kernel function, and b is a constant term.

The LSSVM optimization algorithm uses equality constraints instead of inequality constraints, and its objective functions and constraints are shown in Eqs. (11) and (12).

$$\min_{W,b,e} J_P(W,e) = \frac{1}{2} W^T W + \frac{\gamma}{2}\sum_{k=1}^{M} e_k \tag{11}$$

$$\text{s.t.} \quad y_k = W^T \varphi(x_k) + b + e_k \tag{12}$$

where γ is the punishment factor, and e is the square of the SVM loss function ζ, i.e., $e = \xi^2$. Eqs. (11) and (12) are transformed by the Lagrange multiplier method.

$$L(W,\, b,\, e,\, \alpha) = J_P(W,e) - \sum_{k=1}^{M} \alpha_k \left\{ W^T \varphi(x_k) + b + e_k - y_k \right\} \tag{13}$$

where α_k, k=1, 2, ..., M, is a Lagrange multiplier. A set of linear equations is obtained by taking the partial derivatives of W, b, e_k, and α_k and setting the derivatives equal to 0. The calculation formula of LSSVM optimization algorithm can be obtained by solving the linear equation (14).

$$y(x) = \sum_{k=1}^{M} \alpha_k K(x, x_k) + b \tag{14}$$

where $K(x, x_k)$ is the kernel function, whose type determines the performance of the LSSVM optimization algorithm. The kernel functions commonly used in LSSVM optimization algorithms include the radial basis kernel function (RBF), exponential RBF, polynomial kernel function, and Sigmoid kernel function. Given that RBF has good performance in LSSVM regression fitting and prediction applications, this study chooses RBF as the kernel function of the LSSVM optimization algorithm. The RBF function is shown in Eq. (15).

$$K(x, x_k) = \exp\left(-\frac{\left\| x - x_k \right\|}{2\sigma^2}\right) \tag{15}$$

3.4 Calculation model of ISSA–LSSVM

When the penalty factor γ and kernel function width σ of the LSSVM algorithm are suitable, this algorithm shows good modeling accuracy. Otherwise, this algorithm falls into the local optimization solution, thus reducing the calculation accuracy of the power curve model. This study then uses the ISSA proposed in Sections 3.2 and 3.3 to optimize the penalty factor γ and kernel function width σ of the LSSVM algorithm. The construction process of the ISSA LSSVM model is shown in Fig. 1.

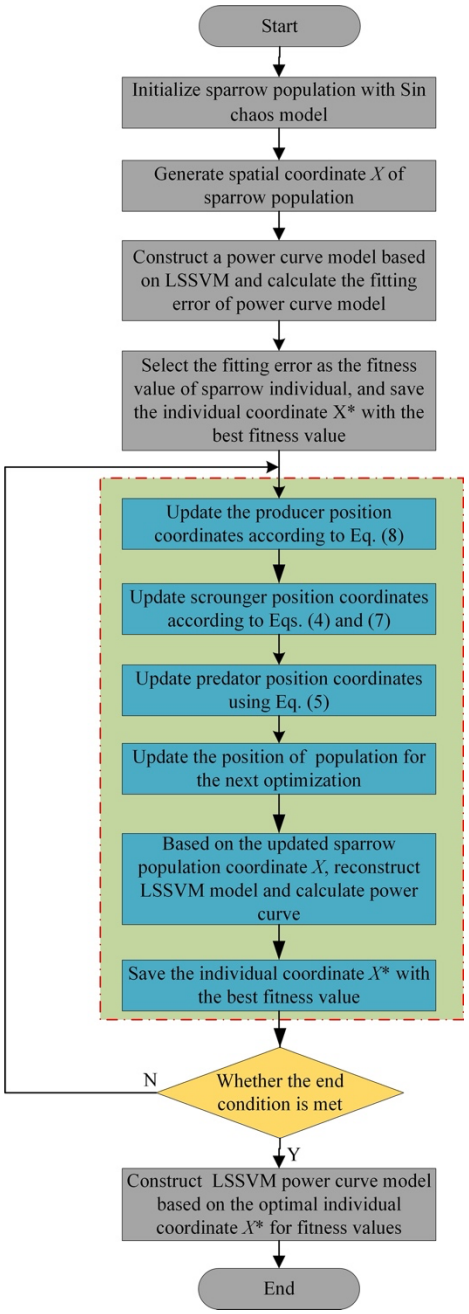

Fig. 1. ISSA–LSSVM model construction process

The specific steps are as follows:

Step 1: The parameter values of SSA are initialized, including the number of sparrow population, the number of producers, the number of scroungers, the maximum number of iterations, the dimension of optimization parameters, and the value range of optimization parameters.

Step 2: The Sin chaotic model in Eq. (6) is used to initialize the spatial location of the sparrow population.

Step 3: According to the position of individual sparrows, the power curve model based on LSSVM is constructed.

Step 4: According to their power curve results, the calculation error of the power curve of each LSSVM model is calculated, and the calculation error is taken as the individual fitness value of the sparrow. The optimal fitness value and the corresponding individual position coordinate of sparrow are then saved.

Step 5: Eq. (8) is used to update the producer position coordinates.

Step 6: Eqs. (4) and (7) are used to the update scrounger position coordinates.

Step 7: Eq. (5) is used to update the position coordinates of the predators.

Step 8: According to the updated population location, the LSSVM model is reconstructed, and the power curves are calculated. According to the calculation results of the LSSVM model, the calculation errors of the power curves of these models are recalculated, and the position coordinate of the individual sparrow with the best fitness value is determined.

Step 9: If the end condition is met, then the optimal sparrow position coordinate and fitness value are outputted. Otherwise, Step 4 is repeated.

3.5 Non-parametric kernel density estimation

Accurately calculating the distribution characteristics of the calculation error of the power curve model is a prerequisite for the uncertainty analysis of this model. Accordingly, this study uses the NPKDE method to calculate these distribution characteristics. NPKDE does not require prior assumptions about the distribution characteristics of the sample data and can directly calculate its probability density function from the sample data. In this method, the choice of kernel function will directly affect the quality of its probability density function estimation results. The Gaussian kernel function is the most commonly used kernel function in NPKDE. Therefore, this study chooses the Gaussian kernel function as the kernel function for NPKDE.

The expression of the Gaussian kernel function is shown in Eq. (16).

$$g(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{16}$$

where $g(x)$ is the Gaussian kernel function, μ is the mean, and σ is the standard deviation.

The probability density distribution of NPKDE is shown in Eq. (17).

$$f(x)=\frac{1}{N}\sum_{i=1}^{N}g\left(\frac{x-x_i}{h}\right) \tag{17}$$

where N is the number of interval samples, h is the bandwidth coefficient, and x_i is the i-th sample.

3.6 Evaluation index of model error

The root mean square error (RMSE) and mean absolute error (MAE) are commonly used to evaluate the calculation accuracy of power curve models. The formulas for calculating RMSE are shown in Eqs. (18) and (19).

$$f(x)=\frac{1}{N}\sum_{i=1}^{N}g\left(\frac{x-x_i}{h}\right) \tag{18}$$

$$P_{\mathrm{RMSE}}=\frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(P_{\mathrm{true}}-P_{\mathrm{fore}}\right)^2}}{P_{\mathrm{cap}}} \tag{19}$$

where RMSE is the root mean square error, N is the number of samples, P_{true} is the true value of the output power of wind turbines, P_{fore} is the calculated value of the power curve model of wind turbines, P_{cap} is the rated output power of wind turbines, and P_{RMSE} is the ratio of the RMSE to the

rated output power. These values are usually expressed in percentages in practice.

MAE truly reflects the error value between the output power of wind turbine sand the calculated value of the power curve model, and its calculation formula is shown in Eqs. (20) and (21).

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|P_{\mathrm{true}} - P_{\mathrm{fore}}\right| \tag{20}$$

$$P_{\mathrm{MAE}} = \frac{\frac{1}{N}\sum_{t=1}^{N}\left|P_{\mathrm{true}} - P_{\mathrm{fore}}\right|}{P_{\mathrm{cap}}} \tag{21}$$

where MAE is the mean absolute error, and P_{MAE} is the ratio of the MAE to the rated power of wind turbines. These values are usually expressed in percentages in practical applications.

3.7 Confidence interval calculation method

After calculating the probability density distribution of the calculation error of the power curve model using NPKDE, the confidence interval can be used to quantitatively calculate such distribution. The calculation error of the power curve model is computed as the difference between the model calculation value P_{fore} and the true power value P_{true} at a certain moment as shown in Eq. (22).

$$e = P_{\mathrm{fore}} - P_{\mathrm{true}} \tag{22}$$

For the calculation error e of the power curve model, its confidence level is calculated using Eq. (23).

$$P\left(e_{\mathrm{low}} < e < e_{\mathrm{up}}\right) = 1 - \theta \tag{23}$$

where the interval $[e_{low}, e_{up}]$ is the confidence interval with a confidence level of 1-θ, e_{low} is the lower limit of the confidence interval, and e_{up} is the upper limit of the confidence interval. $P\left(e_{\mathrm{low}} < e < e_{\mathrm{up}}\right)$ represents the probability that the calculated error value e of the power curve model falls into the interval $[e_{low}, e_{up}]$, from which the confidence interval of the power curve model can be obtained as $[P_{fore} - |e_{up}|, P_{fore} + |e_{low}|]$.

4. Results analysis and discussion

4.1 Data sources and correlation analysis of operating parameters

The data used in this study are collected from a wind farm in a province in western China. The wind farm has 24 wind turbines, with each turbine having a capacity of 2 MW, and the time resolution of SCADA system data is 10 minutes. The data collection period is from 0:00 on May 5, 2020 to 24:00 on July 31, 2021. The SCADA system data includes the state parameters of 24 wind turbines, such as power, wind speed, wind direction, generator speed, rotor speed, nacelle temperature, tower temperature, ambient temperature, gearbox oil temperature, and pitch angle. To investigate the impact of these parameters on the accuracy of wind turbine power curve modeling, the correlation between these parameters and the output power of the wind turbines is analyzed. Table 1 shows the correlation coefficients for the 106th wind turbine.

Table 1. Correlation coefficients between the state parameters and output power of the 106th wind turbine

State parameter	Output power	State parameter	Output power
Output power	1	Nacelle temperature	-0.17
Wind speed	0.91	Tower temperature	-0.10
Wind direction	-0.23	Environmental temperature	-0.14
Generator speed	0.95	Gearbox oil temperature	0.71
Tower base cabinet temperature	-0.06	Gearbox oil inlet temperature	-0.07
Nacelle cabinet temperature	-0.18	Gearbox oil inlet pressure	0.94
State parameter	**Output power**	**State parameter**	**Output power**
Gearbox oil pump pressure	0.70	Generator phase A stator winding temperature	0.76
Gearbox primary bearing temperature	0.58	Generator phase B stator winding temperature	0.74
Gearbox secondary bearing temperature	0.76	Generator phase C stator winding temperature	0.74
Gearbox third stage bearing temperature	0.83	Generator water inlet temperature	-0.37
Generator bearing drive end temperature	0.10	Angle of twisted cable	-0.09
Generator bearing non-driven end temperature	0.30	Pitch angle	0.87

Only five state parameters, namely, wind speed, generator speed, gearbox oil inlet pressure, gearbox third stage bearing temperature, and pitch angle, are strongly correlated with the output power of the wind turbine (correlation coefficient greater than 0.8). Therefore, when modeling the power curve of wind turbines, these state parameters are selected as input data of the power curve model to ensure that the wind turbine state parameters that affect the accuracy of power curve model are fully considered.

4.2 Calculation and analysis of the power curve model

According to Sections 3.1 to 3.4, the power curve model based on ISA-LSSVM is constructed, and the power curve of the 106th wind turbine is calculated and analyzed using this model. The power output of wind turbines varies greatly under different wind speeds. The 2 MW wind turbine selected in this study has output powers of 60 kW and 2000 kW at wind speeds of 3 m/s and 12 m/s, respectively. Assume that the error between the actual output power of the wind turbine and the calculated power of the power curve model is 40 kW. If the wind speed is 12 m/s, then the calculation error is within an acceptable range. However, if the wind speed is 3 m/s, then the calculation error of the power curve model is relatively large.

Therefore, this study divides the wind speed range of 3 m/s–12 m/s into 9 wind speed zones at a 1 m/s interval. When the wind speed is between 12 m/s and 20 m/s, the wind turbine is at rated power output and is thus divided into 1 zone. To verify the high computational accuracy of the ISSA-LSSVM power curve model constructed in different wind speed zones, this study selects three wind speed zones,

namely, 4 m/s-5 m/s, 8 m/s-9 m/s, and 12 m/s-20 m/s, for calculation and analysis. Figures 2 to 4 compare the calculated power of the ISSA-LSSVM power curve model (blue solid line) and the actual power of the wind turbine (red solid line) when the wind speed is 4 m/s-5 m/s, 8 m/s-9 m/s, and 12 m/s-20 m/s, respectively.

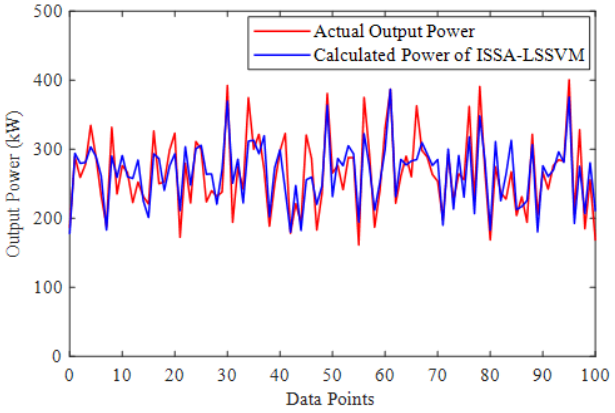

Fig. 2. Calculated power of the power curve model within the wind speed zone of 4 m/s-5 m/s

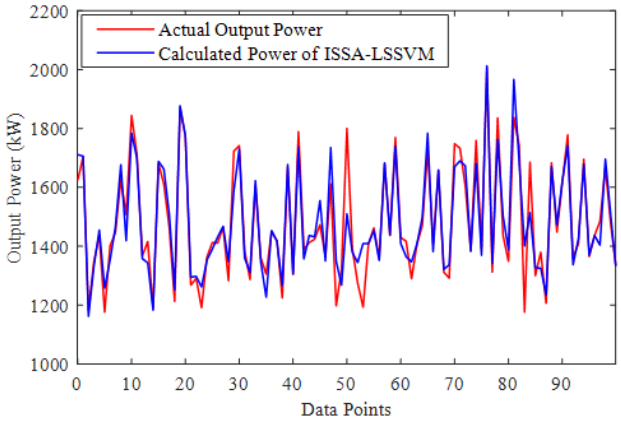

Fig. 3. Calculated power of the power curve model within the wind speed zone of 8 m/s-9 m/s

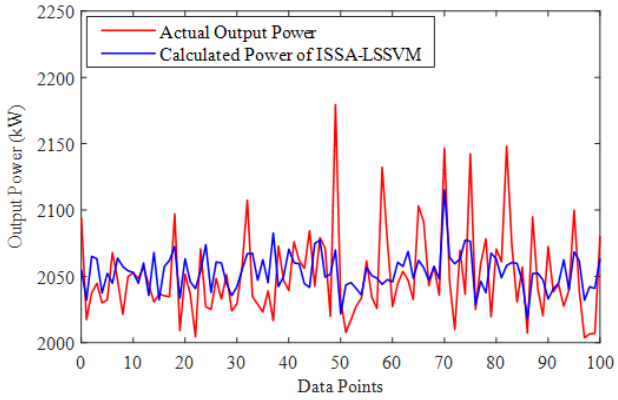

Fig. 4. Calculated power of the power curve model within the wind speed zone of 12 m/s-20 m/s

Figures 2 to 4 show that the ISSA–LSSVM model can accurately calculate the power distribution characteristics of wind turbines in different wind speed zones, thereby proving that the constructed ISSA-LSSVM model can accurately describe the power curve distribution characteristics of wind turbines. Table 2 shows the calculated error values of the ISSA-LSSVM power curve model according to Section 3.6 under the conditions of dividing and not dividing the wind speed range. In Table 2, the average value of the model calculation error is shown in the last row. Results in Table 2 show that dividing the wind speed range can further reduce the calculation error of the ISSA-LSSVM model, thereby indicating that the method of dividing the wind speed zones first and then constructing the curve model of wind turbine power is reasonable and effective.

Table 2. Calculation error of the ISSA-LSSVM model

Wind speed zone (m/s)	Type of error	Wind speed range divided	Wind speed range not divided
3-4	P_{MAE}	1.17%	1.45%
	P_{MAE}	1.42%	2.75%
8-9	P_{MAE}	2.12%	3.06%
	P_{MAE}	3.27%	4.02%
12-20	P_{MAE}	1.16%	1.32%
	P_{MAE}	1.51%	1.71%
Mean error	P_{MAE}	1.48%	1.94%
	P_{MAE}	2.07%	2.83%

4.3 Comparative analysis of power curve models

According to Section 4.2, dividing the wind speed range into wind speed zones can significantly improve the calculation accuracy of the ISSA-LSSVM power curve model. To further demonstrate the superiority of the proposed model, its calculation results are compared with those of the LSTM, LSSVM, PSO-BP, and PSO-LSSVM models. The training and testing sample sets used in the comparative analysis are the same as those used in Section 4.2.

Figures 5 to 7 compare the calculated power of these power curve models in the 4 m/s-5 m/s, 8 m/s-9 m/s, and 12 m/s–20 m/s wind speed zones, respectively. The red solid line denotes the real output power of the wind turbine, the black, blue, and yellow solid lines are the calculated power of the ISSA-LSSVM, PSO-LSSVM, and PSO-BP models, respectively, and the black and blue dashed lines are the calculated power of the LSSVM and LSTM models. These machine learning models can calculate the power curve distribution characteristics of wind turbines accurately, thereby indicating that using machine learning algorithms to build the power curve model of wind turbines is reasonable and feasible.

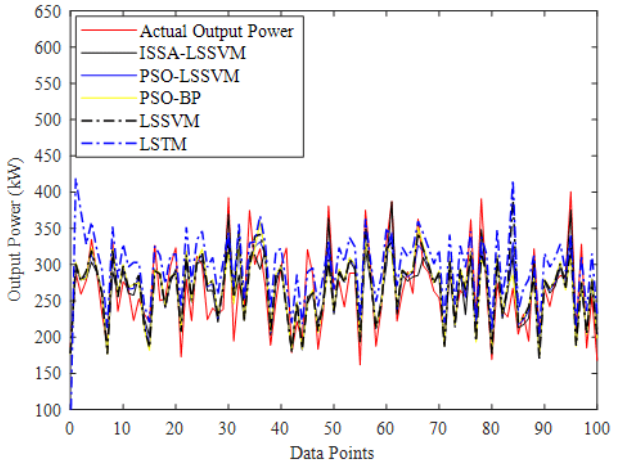

Fig. 5. Comparative analysis of various power curve models in the 4 m/s-5 m/s wind speed zone

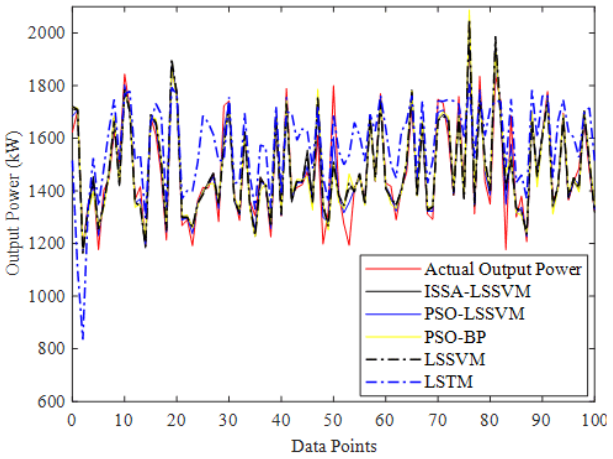

Fig. 6. Comparative analysis of various power curve models in the 8 m/s-9 m/s wind speed zone

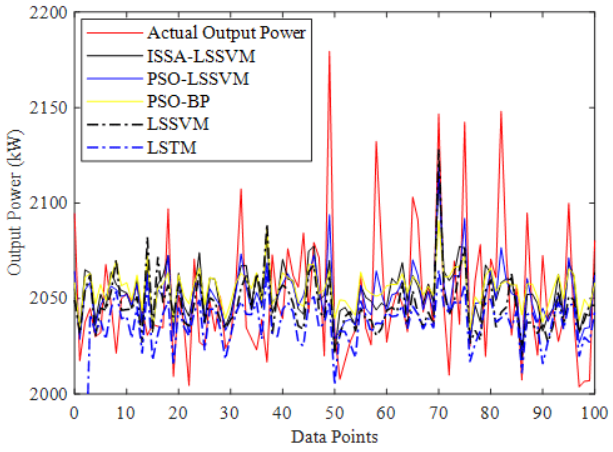

Fig. 7. Comparative analysis of various power curve models in the 12 m/s-20 m/s wind speed zone

Table 3 shows the calculation error values of different power curve models, with the last row showing the average value calculation error. The calculation errors of the ISSA-LSSVM model are smaller than those of the other models across different wind speed zones, thereby confirming that the proposed model has good calculation accuracy.

Table 3. Comparative analysis of power curve models

Wind speed range (m/s)	Type of error	ISSA-LSSVM	PSO-LSSVM	PSO-BP	LSSVM	LSTM
3-4	P_{MAE}	1.17%	1.29%	1.36%	1.98%	2.22%
	P_{RMAE}	1.42%	1.58%	1.69%	6.79%	2.62%
8-9	P_{MAE}	2.12%	2.34%	2.45%	2.24%	7.03%
	P_{RMAE}	3.27%	3.53%	3.64%	3.41%	8.60%
12-20	P_{MAE}	1.16%	1.19%	1.22%	1.21%	2.29%
	P_{RMAE}	1.51%	1.59%	1.66%	1.63%	2.88%
Mean value of error	P_{MAE}	1.48%	1.61%	1.68%	1.81%	3.85%
	P_{RMAE}	2.07%	2.23%	2.33%	3.94%	4.7%

4.4 Uncertainty analysis of the power curve model

There is always a difference between the calculated value of the power curve model and the true power value. Accurately calculating the error distribution characteristics of these values is a prerequisite for quantitatively analyzing the uncertainty of the power curve model. The NPKDE method introduced in Section 3.5 is used to calculate these distribution characteristics, and the results are shown in Figures 8 to 10. The black solid line denotes the probability density function of the calculation error of the power curve model obtained by the NPKDE method, the red dashed line is the probability density function of the calculation error of the power curve model obtained by the normal distribution method, and the histogram is the statistical histogram of the calculation error of the power curve model.

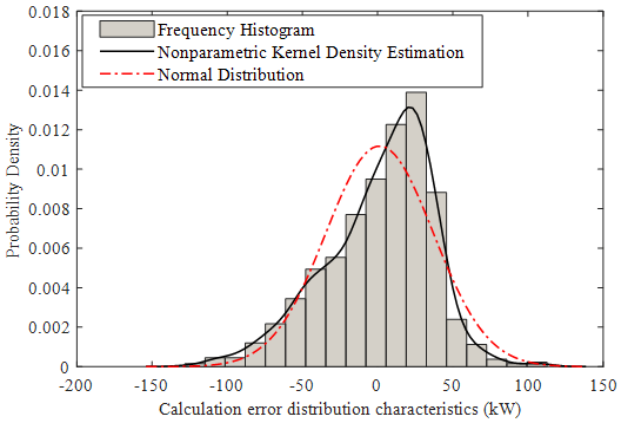

Fig. 8. Calculation error distribution of the power curve model in the 4 m/s-5 m/s wind speed zone

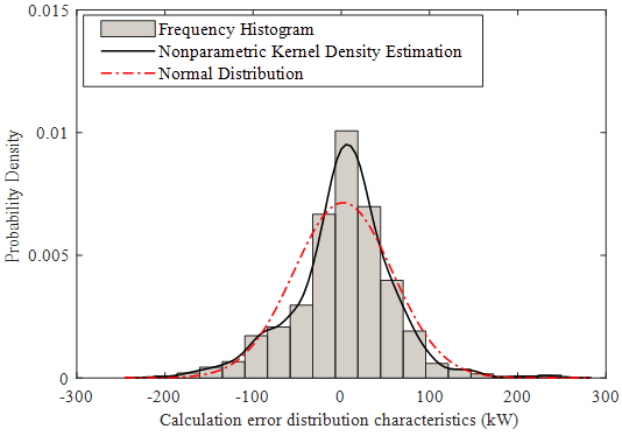

Fig. 9. Calculation error distribution of the power curve model in the 8 m/s-9 m/s wind speed zone

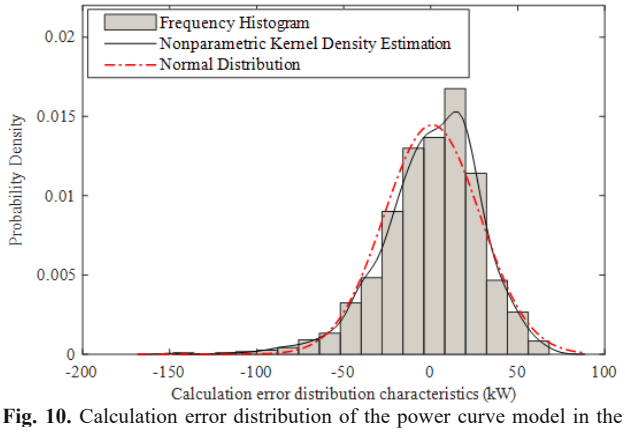

Fig. 10. Calculation error distribution of the power curve model in the 12 m/s-20 m/s wind speed zone

Figures 8 to 10 show that the probability density distribution characteristics calculated by the NPKDE method can accurately describe the distribution characteristics of the calculation error of the power curve model.

The uncertainty distribution characteristics of the power curve model can be calculated following the procedures described in Section 3.7. Figures 11 to 13 show the uncertainty distribution characteristics of the power curve model under different wind speed zones and confidence levels. Most of the calculated values of the power curve models fall within the confidence interval corresponding to the confidence level, thereby proving that the NPKDE method introduced in Section 3.5 can accurately calculate the distribution characteristics of the calculation errors of these models.

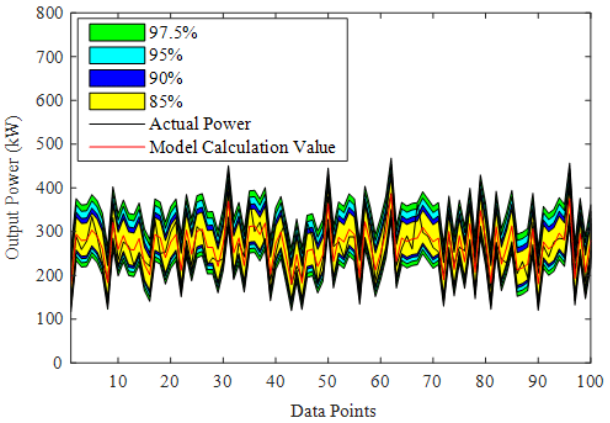

Fig. 11. Uncertainty distribution characteristics of the power curve model in the 4 m/s-5 m/s wind speed zone

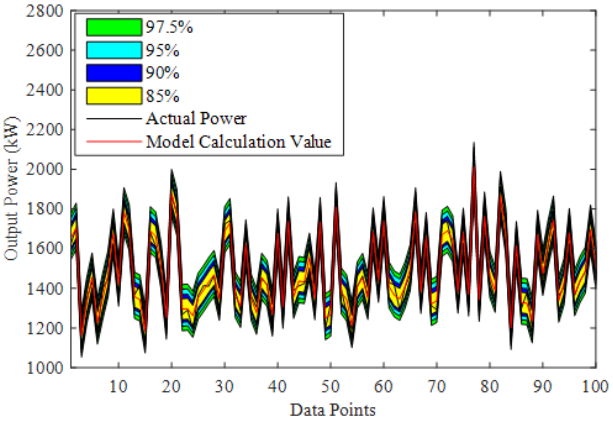

Fig. 12. Uncertainty distribution characteristics of the power curve model in the 8 m/s-9 m/s wind speed zone

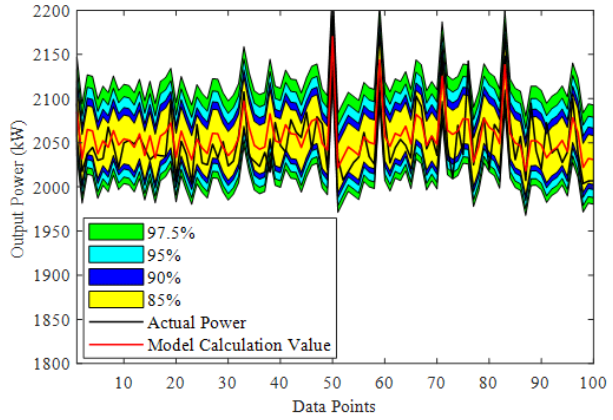

Fig. 13. Uncertainty distribution characteristics of the power curve model in the 12 m/s-20 m/s wind speed zone

Table 4 shows the coverage of confidence intervals in different wind speed zones and at different confidence levels. Across these wind speed zones and confidence levels, the coverage of confidence intervals is greater than the corresponding confidence levels, which further proves that the NPKDE method introduced in Section 3.5 can accurately calculate the distribution characteristics of the calculation errors of the power curve model.

Table 4. Confidence interval coverage

Confidence level / Wind speed zone	97.5%	95%	90%	85%
4 m/s-5 m/s	99%	97%	92%	84%
8 m/s-9 m/s	99%	99%	92%	82%
12 m/s-20 m/s	98%	96%	93%	90%

Figure 14 shows the power distribution characteristics of the wind turbine calculated by the ISA–LSSVM power curve model at a 97.5% confidence level. In Figure 14, the blue dots represent the power distribution calculated by the ISA–LSSVM power curve model, and the red dots represent the real power distribution of the wind turbine. The ISA-LSSVM power curve model can accurately calculate the power distribution characteristics of the wind turbine, thus providing strong support for our analysis of wind turbine operating status.

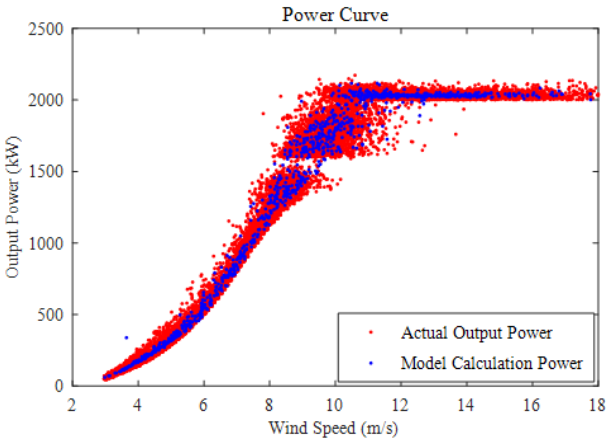

Fig. 14. Power distribution characteristics of the ISA-LSSVM model at a 97.5% confidence level

5. Conclusion

To accurately calculate the distribution characteristics of the real power curve of wind turbines and construct a high-precision power curve calculation model, this study analyzes those factors that influence wind turbine power curve modeling, constructs a power curve model of wind turbines based on ISSA-LSSVM, and compares the calculation accuracy and uncertainty distribution characteristics of this power curve model with that of other models. The following conclusions are drawn from the results:

(1) The state parameters that affect the output power of wind turbines mainly include wind speed, generator speed, gearbox oil inlet pressure, gearbox third stage bearing temperature, and pitch angle.

(2) The wind speed is divided into different wind speed zones, and then the power curve modeling method can improve the precision of power curve modeling for wind turbines.

(3) The power curve modeling precision based on the ISSA-LSSVM model is higher than that of the LSTM, LSSVM, PSO-BP, and PSO-LSSVM models.

(4) The NPKDE method can accurately calculate the distribution characteristics of the calculation error of the power curve model for wind turbines.

In this study, a novel power curve modeling method for wind turbines was proposed, and its uncertainty was calculated. The results have a certain role in promoting the health management and analysis of wind turbine operating conditions. However, the precise modeling of the power curve of wind turbines is influenced by external environment and spatiotemporal factors. In the modeling process, this study only considers the impact of the external environment on the modeling accuracy of the power curve and ignores the impact of spatiotemporal factors. Future research should comprehensively consider the impact of the external environment and spatiotemporal factors on the modeling accuracy of the power curve to further improve its modeling accuracy.

Acknowledgements

This work was supported by the Funding Plan for Young Backbone Teachers in Colleges and Universities of Henan Province (2019GGJS105), the Key Research Projects of Henan Higher Education Institutions (22A570006), and the Henan Natural Science Foundation (232300420152). The authors would like to thank the editors and reviewers for their helpful comments on this manuscript. The authors also thank those people who helped in this study but whose names are not listed.

References

[1] J. Liu, B. R. An, W. B. Zhang, and Q. Y. Gan. "Review of Health Status Evaluation of Large Wind Turbines," *Pow. Syst. Pr. and Control*, vol. 51, no. 1, pp. 176-187, Jan. 2023.

[2] Y. B. Wu, J. Z. Zhang, Z. X. Yuan, and F. J. Deng. "Review on Identification and Cleaning of Abnormal Wind Power Data for Wind Farms," *Pow. Syst. Tech.*, vol. 47, no. 6, pp. 2367-2379, Jun. 2023.

[3] G.W. Qian and T. Ishihara. "A Novel Probabilistic Power Curve Model to Predict the Power Production and Its Uncertainty for a Wind Farm over Complex Terrain," *Energy*, vol. 261, 2022, Art. no. 125171.

[4] T. Liang, J. Cui, H. Shi, and Z. Q. Li. "Comparative Study of Wind Turbine Power Curve Modeling Method," *Comp. Sim.*, vol. 38, no. 2, pp. 62-66, Feb. 2021,

[5] Q. Xia, A. Lei, Z. Yan, and F. Li. "Comparative Study of Multiple Power Curve Modelling Methods," *Renew. Energ. Res.*, vol. 36, no. 4, pp. 580-585, Apr. 2018.

[6] Y. Wang, Q. Hu, L. Li, A. M. Foley, and D. Srinivasan. "Approaches to Wind Power Curve Modeling: A Review and Discussion," *Renew. Sust. Energ. Rev.*, vol. 116, Dec. 2019, Art. no. 109422.

[7] P. Nursery and A. A. Ezzat. "Yaw-adjusted Wind Power Curve Modeling: A Local Regression Approach," *Renew. Energ.*, vol. 202, pp. 1368-1376, Dec. 2022.

[8] M. Mehrjoo, M. J. Jozani, and M. Pawlak. "Wind Turbine Power Curve Modeling for Reliable Power Prediction Using Monotonic Regression," *Renew. Energ.*, vol. 147, pp. 214-222, Aug. 2019.

[9] J. C. Chu, P. Guo, and J. Y. Xie. "Wind Turbine Power Curve Modeling and Application Based on AAKR Method," *ACTA Energiae Solaris SINICA*, vol. 47, no. 7, pp. 372-377, Jul. 2021.

[10] Y. Wang, X. C. Duan, R. M. Zou, F. Zhang, Y. F. Li, and Q. H. Hua. "A Novel Data-driven Deep Learning Approach for Wind Turbine Power Curve Modeling," *Energy*, vol. 270, May 2023, Art. no. 126908.

[11] K. Y. Xu, J. Yan, H. Zhang, H. R. Zhang, S. Han, and Y. Q. Liu. "Quantile Based Probabilistic Wind Turbine Power Curve Model," *Appl. Energ.*, vol. 296, Aug. 2021, Art. no. 116913.

[12] R. M. Zou, J. X. Yang, Y. Wang, F. Liu, M. Essaaidi, and D. Srinivasan. "Wind Turbine Power Curve Modeling Using an Asymmetric Error Characteristic-based Loss Function and a Hybrid Intelligent Optimizer," *Appl. Energ.*, vol. 304, Dec. 2021, Art. no. 117707.

[13] P. Wang, Y. T. Li, and G. Y. Zhang. "Probabilistic Power Curve Estimation Based on Meteorological Factors and Density LSTM," *Energy*, vol. 269, Apr. 2023, Art. no. 126768.

[14] W. Z. Han, F. Y. Wang, and Z. H. Ge. "Study on Wind Speed-power Curve Fitting Model of Wind Turbine Based on Generalized Least Squares Method," *Advan. Tech. Elec. Eng. and Energ.*, vol. 37, no. 9, pp. 67-73, Sep. 2018.

[15] H. T. Li, P. Guo, and X. Y. Yang. "Study on Wind Turbine Power Curve Drawing Method Based on Discrete Degree Analysis," *ACTA Energiae Solaris SINICA*, vol. 40, no. 1, pp. 237-241, Jan. 2019.

[16] L. X. Cao, W. M. Liu, and H. Q. Guo. "Cleaning and Modeling of Abnormal Data of Wind Farm Power Curve," *Journ. Lanz. Univ. Tech.*, vol. 48, no. 4, pp. 64-69, Apr. 2022.

[17] S. Alessandro, P. Alfredo, and T. Niels. "Numerical Evaluation of Multivariate Power Curves for Wind Turbines in Wakes Using Nacelle Lidars," *Renew. Energ.*, vol. 202, pp. 419-431, Jan. 2023.

[18] M. S. D. Yves, et al. "A Parametric Model for Wind Turbine Power Curves Incorporating Environmental Conditions," *Renew. Energ.*, vol. 157, pp. 754-768, Sep. 2020.

[19] J. Yan, H. Zhang, Y. Liu, S. Han, and L. Li. "Uncertainty Estimation for Wind Energy Conversion by Probabilistic Wind Turbine Power Curve Modelling," *Appl. Energ.*, vol. 239, pp. 1356-1370, Apr. 2019.

[20] E. J. Yun and J. Hur. "Probabilistic Estimation Model of Power Curve to Enhance Power Output Forecasting of Wind Generating Resources," *Energy*, vol. 223, May. 2021, Art. no. 120000.

[21] T. J. Rogers, P. Gardner, N. Dervilis, and E. J. Cross. "Probabilistic Modelling of Wind Turbine Power Curves with Application of Heteroscedastic Gaussian Process Regression," *Renew. Energ.*, vol. 148, pp. 1124-1136, Apr. 2020.

[22] L. A. Bull, et al. "Bayesian Modelling of Multivalued Power Curves from an Operational Wind Farm," *Mech. Syst. Signal Pr.*, vol. 169, Mar. 2022, Art. no. 108530.

[23] D. Y. Fu, S. Q. Gao, L. X. Kong, and H. K. Jia. "Wind Turbine Power Curve Construction Based on Correlation Vector Information Entropy," *ACTA Energiae Solaris SINICA*, vol. 43, no. 5, pp. 252-259, May 2022.

[24] K. P. Ravi, I. David, and K. Athanasios. "Gaussian Process Power Curve Models Incorporating Wind Turbine Operational Variables," *Energy Rep.*, vol. 6, pp. 1658-1669, Nov. 2020.

[25] M. Mehrdad, J. J. Mohammad, and P Miroslaw. "Toward Hybrid Approaches for Wind Turbine Power Curve Modeling with Balanced Loss Functions and Local Weighting Schemes," *Energy*, vol. 218, Mar. 2021, Art. no. 119478.

[26] W. Zha, Y. Jin, Y. Li, and Y. Sun. "A Wind Speed Vector-wind Power Curve Modeling Method Based on Data Denoising Algorithm and the Improved Transformer," *Electr. Pow. Syst. Res.*, vol. 214, Jan. 2023, Art. no. 108838.

[27] G. C.M. Virgolino, C. L. C. Mattos, J. A. F. Magalhaes, and G. A. Barreto. "Gaussian Processes with Logistic Mean Function for Modeling Wind Turbine Power Curves," *Renew. Energ.*, vol. 162, pp. 458-465, Dec. 2020.

[28] T. Li, X. Liu, Z. Lin, and R. Morrison. "Ensemble Offshore Wind Turbine Power Curve modelling - An Integration of Isolation Forest, Fast Radial Basis Function Neural Network, and Metaheuristic Algorithm," *Energy*, vol. 239, Jan. 2022, Art. no. 122340.

[29] K. Despina, K. Vasiliki, and A. Alex. "Wind Turbine Power Curve Modeling Using Radial Basis Function Neural Networks and Tabu Search," *Renew. Energ.*, vol. 163, pp. 2137-2152, Jan. 2021.

[30] M. Bartolome, *et al*. "Wind Turbine Power Curve Modeling Based on Gaussian Processes and Artificial Neural Networks," *Renew. Energ.*, vol. 125, pp. 1015-1020, Sep. 2018.

[31] J. K. Xue and B. Shen. "A Novel Swarm Intelligence Optimization Approach: Sparrow Search Algorithm," *Syst. Sci. Control. Eng.*, vol. 8, no. 1, pp. 22-34, Jan. 2020.

[32] C. L. Zhang and S. F. Ding. "A Stochastic Configuration Network Based on Chaotic Sparrow Search Algorithm," *Knowl-based Syst.*, vol. 220, no. 10, May 2021, Art. no. 106924.

[33] J. A. K. Suykens and J. Vandewalle. "Least Square Support Vector Machine Classifiers," *Neural Process Lett.*, vol. 9, no. 3, pp. 293-300, Apr. 1999.