



## Research Article

# Inverse Method of Centrifugal Pump Blade Based on Gaussian Process Regression

Renhui Zhang <sup>1,2</sup> and Xutao Zhao <sup>1</sup>

<sup>1</sup>School of Energy and Power Engineering, Lanzhou University of Technology, Lanzhou 730050, China

<sup>2</sup>Key Laboratory of Fluid Machinery and Systems, Lanzhou University of Technology, Lanzhou 730050, China

Correspondence should be addressed to Renhui Zhang; zhangrh lut@163.com

Received 13 August 2019; Revised 18 December 2019; Accepted 1 February 2020; Published 28 February 2020

Academic Editor: Nicolae Herisanu

Copyright © 2020 Renhui Zhang and Xutao Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The inverse problem is always one of the important issues in the field of fluid machinery for the complex relationship among the blade shape, the hydraulic performance, and the inner flow structure. Based on Bayesian theory of posterior probability obtained from known prior probability, the inverse methods for the centrifugal pump blade based on the single-output Gaussian process regression (SOGPR) and the multioutput Gaussian process regression (MOGPR) were proposed, respectively. The training sample set consists of the blade shape parameters and the distribution of flow parameters. The hyperparameters in the inverse problem models were trained by using the maximum likelihood estimation and the gradient descent algorithm. The blade shape corresponding to the objective blade load can be achieved by the trained inverse problem models. The MH48-12.5 low specific speed centrifugal pump was selected to verify the proposed inverse methods. The reliability and accuracy of both inverse problem models were confirmed and compared by implementing leave-one-out (LOO) cross-validation and extrapolation characteristic analysis. The results show that the blade shapes within the sample space can be reconstructed exactly by both models. The root mean square errors of the MOGPR inverse problem model for the pump blade are generally lower than those of the SOGPR inverse problem model in the LOO cross-validation. The extrapolation characteristic of the MOGPR inverse problem model is better than that of the SOGPR inverse problem model for the correlation between the blade shape parameters can be fully considered by the correlation matrix of the MOGPR model. The proposed inverse methods can efficiently solve the inverse problem of centrifugal pump blade with sufficient accuracy.

## 1. Introduction

Due to the complex internal flow constraints of fluid machinery, there is a complex implicit relationship among its geometric parameters, internal flow, and hydraulic performance. The problems related to the internal flow of fluid machinery can be summarized as the direct problem and the inverse problem. The direct problem focuses on the flow structure by experimental and numerical approaches [1–3], and the inverse problem mainly studies how to acquire the blade geometry shape according to the objective flow field distribution [4–7]. Usually, the inverse problem of fluid machinery can be considered as the design issues. In recent decades, with the rapid development of computational fluid

dynamics (CFD) and modern flow testing techniques, the researches on the direct problem of centrifugal pump have made much progress. Compared with the direct problem of flow field analysis for pumps, the inverse problem is much more difficult. Hawthorne et al. [8] and Tan et al. [9] first proposed the inverse design method of fluid machinery. Borges [4] developed the theory of inverse method to three-dimensional under incompressible conditions, and then Zangeneh and Goto [5] extended the three-dimensional inverse method to compressible conditions. Furthermore, Zangeneh et al. [6, 7] used the three-dimensional inverse method to suppress secondary flow in the pump impeller. Bonaiuti et al. [10, 11] combined the inverse design method with optimization techniques to realize the optimization

design of the pump blade. At present, the inverse methods for the centrifugal pump impeller can in general be classified into two categories: one is the inverse method based on the general theory of relative stream surfaces proposed by Wu [12] and the other is the iterative method based on the iteration between flow simulation and modification design of impeller geometry [13–15]. In the first method, the blade shape is reconstructed with the assumption that the flow must be aligned to the blade surfaces [5–7, 16, 17]. The CFD model is greatly simplified, which leads to the flow simulation without sufficient accuracy, so the accuracy of this inverse method is hard to ensure. In the second method, the more accurate turbulence model is employed to simulate the flow field and the accuracy of numerical simulation is improved. However, it is hard to get a feasible blade geometry which agrees well with the specified flow field distribution, for the modification of blade geometry is highly dependent on the experiences of designers. Recently, the proper orthogonal decomposition (POD) reduced-order model was proposed for the inverse method of centrifugal pump blade. Zhang et al. [18] proposed the inverse method of centrifugal pump blade based on the POD model and then performed the inverse design of the two-dimensional and three-dimensional centrifugal pump blades by the Gappy POD model. The inverse method based on the POD model has a feature of quicker convergence, but the accuracy of it needs to be further improved.

Gaussian process regression (GPR) is a machine learning approach based on the Bayesian theorem, which provides a flexible framework for probabilistic regression and has good adaptability to deal with the high-dimensional and small-sample problems, etc. [19]. Gaussian process regression is mainly divided into the SOGPR model and the MOGPR model. The SOGPR model has less unknown parameters and can be easily explained, which has been widely used in dimensionality reduction [20], time-series analysis [21, 22], nonlinear regression [19], etc. Based on the SOGPR model, the MOGPR model is improved to learn the correlation information between the outputs of models, which can provide more accurate predictions in comparison with modeling outputs individually by the SOGPR model [23–26]. Liu et al. [27] constructed a multiresponse surface model for airfoil design based on the MOGPR model. Chai et al. [28] took the multitask Gaussian process to compute the inverse dynamics problem for a robotic manipulator. Wu et al. [29] combined the MOGPR model with the optimization algorithm to optimize the supercritical airfoils.

The GPR method has high prediction accuracy in the case of a small number of samples. To improve the accuracy of the centrifugal pump blade inverse problem, the GPR was introduced to the inverse method of centrifugal pump blade. The blade load distributions of the pump were considered as the model input, the blade shape parameters were taken as the output, and then the inverse problem calculation of blade shapes was implemented by both of the models. The reliability of both models was verified, respectively, and the accuracy was compared and analyzed.

## 2. Gaussian Process Regression Models

*2.1. Single-Output Gaussian Process Regression Model.* A brief introduction to the SOGPR model is provided here, and a more detailed description can be found in [19]. A training sample set is defined as  $D = \{(\mathbf{x}_i, y_i) | i = 1, 2, \dots, n\}$ , where  $\mathbf{x}_i$  is the  $d$ -dimensional input and  $y_i$  is the one-dimensional output. The input  $\mathbf{x}_i$  corresponds to the random variable function  $f(\mathbf{x}_i)$ , and the collection of  $f(\mathbf{x}_i)$  satisfies the joint Gaussian distribution, which can be interpreted as

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (1)$$

where  $m(\mathbf{x})$  is the mean function, which is commonly assumed to be zero, and  $k(\mathbf{x}, \mathbf{x}')$  is the covariance function.  $f(\mathbf{x})$  is completely specified by the mean function and the covariance function. In many realistic scenarios, the values of outputs are replaced by observations, which can be expressed as  $y_i = f(\mathbf{x}_i) + \varepsilon$ , where  $\varepsilon \sim N(0, \sigma_n^2)$  is independent and identically distributed, which accounts for the measurement errors and is known as noise in the Gaussian process models. Therefore, the joint Gaussian distribution of the vector  $\mathbf{y}$  is expressed as

$$\mathbf{y} \sim GP(\mathbf{0}, k(\mathbf{x}, \mathbf{x}') + \sigma_n^2 \delta_{ij}), \quad (2)$$

where  $\delta_{ij}$  is the Kronecker delta function. The joint prior distribution of observed values  $\mathbf{y}$  and the function value  $f(\mathbf{x}_*)$  at a test point  $\mathbf{x}_*$  is

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & K(\mathbf{X}, \mathbf{x}_*) \\ K(\mathbf{x}_*, \mathbf{X}) & K(\mathbf{x}_*, \mathbf{x}_*) + \sigma_n^2 \end{bmatrix}\right), \quad (3)$$

where  $K(\mathbf{X}, \mathbf{X}) \in R^{n \times n}$  is the covariance matrix of inputs with the element  $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ , which can describe the information between the inputs in the training sample set. According to the inference of conditional distribution based on multivariate Gaussian distribution, the posterior distribution of  $f_*$  is analytically derived as

$$f_* | \mathbf{x}_*, \mathbf{X}, \mathbf{y} \sim N(\bar{f}_*, \text{cov}(f_*)), \quad (4)$$

where the prediction mean  $\bar{f}_*$  can approximate the value of the unknown output of the test sample and the prediction variance  $\text{cov}(f_*)$  can provide the uncertainty in the solution of the unknown output; they are, respectively, given as

$$\bar{f}_* = K(\mathbf{x}_*, \mathbf{X}) [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y}, \quad (5)$$

$$\text{cov}(f_*) = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{X}) \cdot [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} K(\mathbf{X}, \mathbf{x}_*). \quad (6)$$

During the calculation of the covariance matrix  $K$  in the previously mentioned SOGPR model, according to the Mercer theorem, the well-known squared exponential (SE) kernel function was used to calculate the covariance matrix  $K$ . The SE kernel function has good performance within the kernel machine field, and it is expressed as

$$k_{SE}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2l^2}\right), \quad (7)$$

where the signal variance  $\sigma_f^2$  controls the output scale of the kernel function and the characteristic length scale  $l$  represents the level that the output result is impacted by different dimensionalities of the input  $\mathbf{x}$ . The vector of hyperparameters  $\boldsymbol{\theta}$  is defined, which contains the characteristic length scale  $l$ , the signal variance  $\sigma_f^2$ , and the noise variance  $\sigma_n^2$ . The values of the hyperparameters were optimized by the maximum likelihood estimation and the gradient descent algorithm.

It is known that the vector of the outputs  $\mathbf{y}$  obeys the multivariate Gaussian distribution, the negative log marginal likelihood (NLML) of which is defined as

$$\begin{aligned} \text{NLML} = -\log P(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) &= \frac{1}{2} \mathbf{y}^T (K + \sigma_n^2 I)^{-1} \mathbf{y} \\ &+ \frac{1}{2} \log |K + \sigma_n^2 I| + \frac{n}{2} \log 2\pi. \end{aligned} \quad (8)$$

The optimal values of  $\boldsymbol{\theta}$  can be obtained by using the gradient descent algorithm to minimize the NLML. In order to avoid optimization falling into local minima in the NLML, the hyperparameters usually need to be initialized randomly for multiple times and the hyperparameters with the lowest NLML will be selected [30].

**2.2. Multioutput Gaussian Process Regression Model.** Based on the SOGPR model, the MOGPR is improved to model  $t$  outputs simultaneously to learn their correlation, which can outperform individual modeling. The flow charts of the SOGPR model and the MOGPR model are shown in Figure 1. The main difference between MOGPR and SOGPR is the construction of the kernel function matrix  $K$ . A correlation matrix based on the Kronecker product is employed to describe the correlation between  $t$  outputs.

Consider a set of training samples  $D = \{(\mathbf{x}_i, \mathbf{y}_i) | i = 1, 2, \dots, n\}$ , where  $\mathbf{x}_i$  is the  $d$ -dimensional input and  $\mathbf{y}_i$  contains  $t$  one-dimensional outputs. For a model with  $n$  samples, each sample has  $t$  outputs, and the kernel function matrix can be expressed as

$$\mathbf{K}(\mathbf{X}, \mathbf{t}, \boldsymbol{\theta}_c, \boldsymbol{\theta}_x) = \mathbf{K}_c(\mathbf{t}, \boldsymbol{\theta}_c) \otimes \mathbf{K}_x(\mathbf{X}, \boldsymbol{\theta}_x), \quad (9)$$

where  $\otimes$  represents the Kronecker product and  $\mathbf{K}_x$  is the same as the kernel function matrix in the SOGPR model and has a size of  $n \times n$ . It describes the relationship between the inputs of all samples.  $\mathbf{K}_c$  expresses the correlation between outputs, which is referred to as the correlation matrix and has a size of  $t \times t$ . The diagonal elements of  $\mathbf{K}_c$  represent the correlation between outputs and themselves, and the non-diagonal elements describe the correlation between the different outputs. If  $\mathbf{K}_c$  is the identity matrix, all outputs would be considered to be independent and have no correlation. Generally,  $\mathbf{K}_c$  is initialized as the identity matrix.  $\mathbf{K}_c$  and  $\mathbf{K}_x$  lead to a matrix of size  $nt \times nt$  for  $\mathbf{K}$ .  $\boldsymbol{\theta}_c$  and  $\boldsymbol{\theta}_x$  are vectors including hyperparameters for  $\mathbf{K}_c$  and  $\mathbf{K}_x$ , respectively. The number of elements in  $\boldsymbol{\theta}_c$  increases rapidly with

the increase of outputs.  $\boldsymbol{\theta}_c$  and  $\boldsymbol{\theta}_x$  can be learned together by optimizing the NLML.

### 3. Inverse Method of Pump Blade Based on GPR

The inverse problem of centrifugal pump blade also belongs to the pump design issues, the target of which is to get the desired blade shape which can produce the prescribed flow field distribution. In this research, according to the models mentioned above, the blade load distributions  $\mathbf{x}_i$  and the blade shape parameters  $\mathbf{y}_i$  were defined as the input and output of both models, respectively. The objective blade shapes can be obtained by both of the trained inverse problem models when the objective blade load distributions are given. The algorithm of the proposed centrifugal pump blade inverse methods can be summarized as follows:

*Step 1.* Parameterize the prototype blade shape and obtain the initial samples of the blade shape by experimental design

*Step 2.* Simulate the inner flow by the CFD method and calculate the blade load distribution of all samples

*Step 3.* Combine the blade load distribution data and blade shape parameters of all initial samples into a training sample set  $D = \{(\mathbf{x}_i, \mathbf{y}_i) | i = 1, 2, \dots, n\}$

*Step 4.* Give the initial values of the hyperparameters and get the optimal hyperparameters  $\boldsymbol{\theta}[l, \sigma_f, \sigma_n]$  of SOGPR or  $\boldsymbol{\theta}[l, \sigma_f, \sigma_n, K_c]$  of MOGPR by using the gradient descent algorithm to minimize the NLML

*Step 5.* Get the mean and variance of blade shape parameters corresponding to the given objective blade load by equations (5) and (6)

*Step 6.* Plot the blade shape and the corresponding 95% confidence interval according to the mean and the variance of the blade shape parameters, respectively

In step 1, as we can see in Figure 2, the prototype blade shape was parameterized by the cubic Bezier curve. The blade inlet and outlet angles were, respectively, controlled by the slopes of edges  $AB$  and  $CD$ . The blade wrap angle was controlled by the movement of point  $D$  in the circumferential direction of the outlet diameter of the impeller.

### 4. Results and Discussion

**4.1. Sample Set and Training Data Generation.** The MH48-12.5 low specific speed centrifugal pump ( $Q = 12.5 \text{ m}^3/\text{h}$ ,  $H = 30.7 \text{ m}$ ,  $n = 2900 \text{ r/min}$ ,  $\eta = 53\%$ ) was applied to verify the inverse method of centrifugal impeller based on Gaussian process regression. The meridional plane and end view of the prototype centrifugal impeller are shown in Figure 3. The wrap angle  $\Phi$  of the impeller is  $143^\circ$ , the inlet blade angle  $\beta_1$  is  $30^\circ$ , and the outlet blade angle  $\beta_2$  is  $17^\circ$ . The prototype blade shape was parameterized by the cubic Bezier curve. The uniform design of experiments was employed to generate the sample set. In the condition that the meridional plane of the impeller was fixed, there were eight initial blade samples generated by the perturbation of  $\beta_1$ ,  $\beta_2$ , and  $\Phi$  for  $5^\circ$ ,  $5^\circ$ , and  $10^\circ$  on the basis of the prototype blade shape,

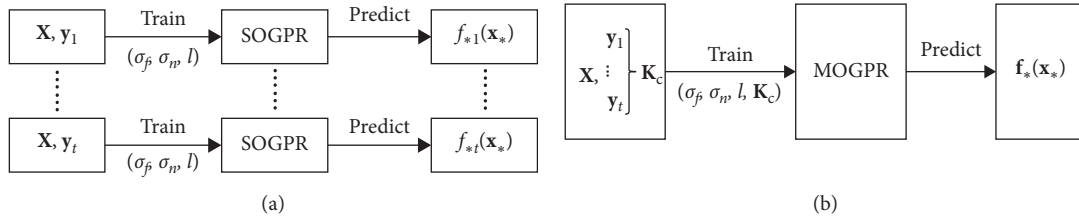


FIGURE 1: Illustration of the (a) SOGPR model and (b) MOGPR model.

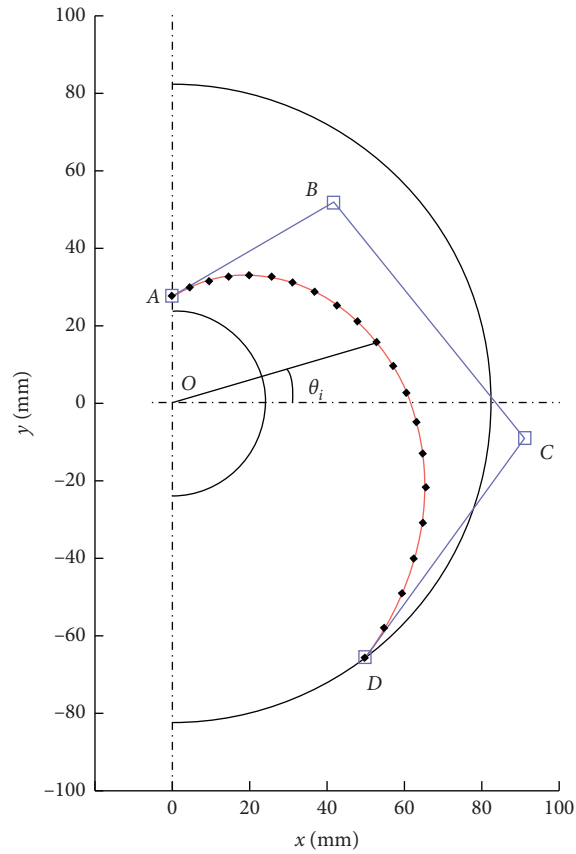


FIGURE 2: Parametric control of the prototype blade shape.

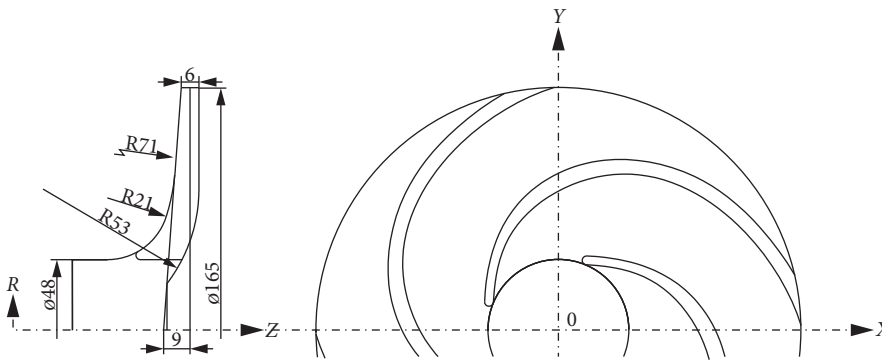


FIGURE 3: Meridional plane and end view of the prototype impeller.

TABLE 1: Perturbation parameter values of all blade samples.

	Prototype blade	Extrapolation blade	Sample blades							
			Blade 1	Blade 2	Blade 3	Blade 4	Blade 5	Blade 6	Blade 7	Blade 8
$\Phi$ ( $^\circ$ )	143	113	153	153	153	153	133	133	133	133
$\beta_1$ ( $^\circ$ )	30	35	25	35	35	25	25	35	35	25
$\beta_2$ ( $^\circ$ )	17	20	22	22	12	12	22	22	12	12

respectively. The wrap angle, inlet blade angle, and outlet blade angle of all blade samples are listed in Table 1, and all blade shapes are shown in Figure 4. The angular coordinates of 20 points uniformly distributed along the radius of the impeller from the blade inlet to the outlet, as shown in Figure 2, were defined to control the blade shape. Hence, the blade shape can be expressed as  $\mathbf{y}_i = [\theta_i^1, \theta_i^2, \dots, \theta_i^{20}]$ .

In the field of hydraulic machinery, the change of the energy gradient of fluid from the impeller inlet to the outlet determines the internal flow characteristics and hydraulic performances of pump impellers. Therefore, the gradient of the flow head  $\partial H_i / \partial m$  was directly defined as the blade load, which can be obtained by CFD simulation [31, 32]. In order to get blade load data conveniently, the hexahedral structured grids were adopted to discretize the computational domains of the impeller, volute, inlet pipe, and outlet pipe. The details of the grid independence test at the design flow rate are shown in Figure 5, and the final grid number for the computation was determined as 1.1 million. The commercial code ANSYS FLUENT 16.0 with the RNG  $k - \varepsilon$  turbulence model and SIMPLEC algorithm was employed for CFD calculation. A uniform velocity was set to the inlet boundary condition, and the free outflow was set to the outlet boundary condition. The near wall flow was treated by standard wall function. The convergence criterion of all residuals was set as  $10^{-5}$ . The calculated blade load distributions of all samples are shown in Figure 6. As we can see, the blade load distributions vary with the blade shapes, the blade load distributions on the front section of the blades are obviously affected by the inlet blade angles, and the blade load distributions on the trailing section of the blades are significantly influenced by the outlet blade angles. The blade load values of 63 grid nodes from the inlet to the outlet of the pump blade were considered as the model input, and they were expressed as  $\mathbf{x}_i = [\Delta H_i^1, \Delta H_i^2, \dots, \Delta H_i^{63}]$ . The whole sample set was described as  $D = [\mathbf{X} \ \mathbf{Y}]$ , where  $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_8^T]^T$  and  $\mathbf{Y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_8^T]^T$ . The rows of  $\mathbf{X}$  and  $\mathbf{Y}$  both represent the number of samples. The columns of  $\mathbf{X}$  and  $\mathbf{Y}$  represent the dimension of the model input and output, respectively.

**4.2. Results of Inverse Methods.** The SOGPR model and the MOGPR model were, respectively, used to construct the inverse methods for the centrifugal pump blade. Both of the inverse problem models were programed by using the MATLAB code. In the SOGPR model,  $\mathbf{X}$  was used as the model input, and each column of  $\mathbf{Y}$  was taken as the one-dimensional output of the model, so that 20 SOGPR models were trained independently. The angular coordinates of 20 points on blade shape can be obtained by the 20 trained

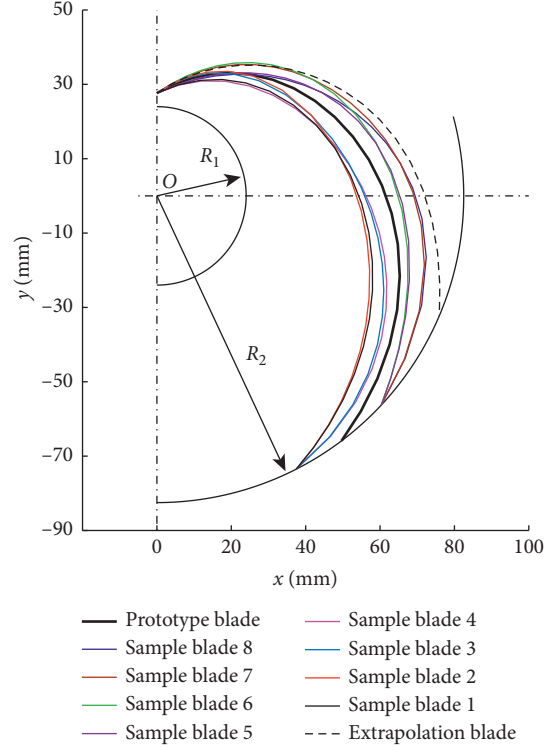


FIGURE 4: Sample set of all blade shapes.

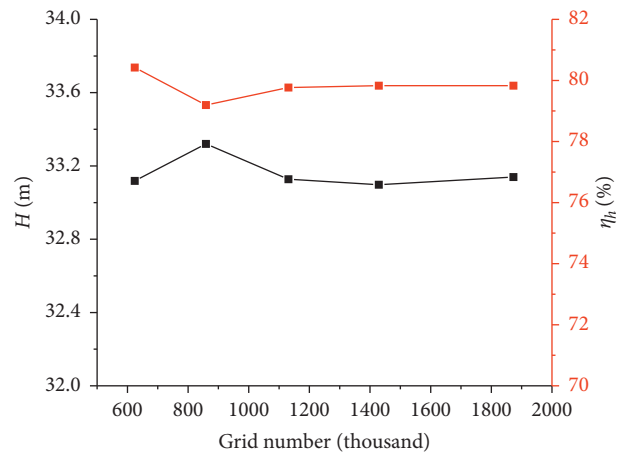


FIGURE 5: Grid independence test.

models according to the given blade load distribution. The impeller blade geometry can be reconstructed by the prescribed meridional plane and these 20 points on blade shape. In the MOGPR model, however, the entire  $\mathbf{Y}$  was taken as the output, which has 20 dimensionalities. During the

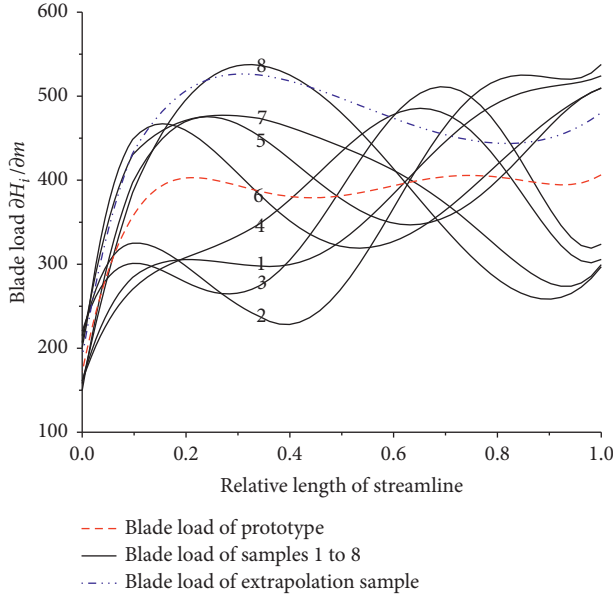


FIGURE 6: Blade load distributions of all sample blades.

training of this model, all dimensionalities of the output were trained simultaneously to consider their correlation so that there was only one model that needs to be trained.

Firstly, the prototype blade was selected as the objective blade to achieve the inverse problem calculation for the centrifugal pump blade. The hyperparameters were initialized according to the training samples [33]. The characteristic length scale  $l$ , the signal variance  $\sigma_f^2$ , and the noise variance  $\sigma_n^2$  were finally initialized to 2.6, 0.9, and 0.05 for both models, respectively. The correlation matrix of the MOGPR model was initialized as an identity matrix. Regarding hyperparameter learning, the gradient descent algorithm was employed to minimize the NLML and the maximum number of function evaluations was set as 800 for both models. The blade shapes, shown in Figures 7 and 8, were, respectively, calculated by the trained MOGPR and SOGPR models according to the objective blade load distribution. It can be seen that the blade shapes generated by the SOGPR model and the MOGPR model both are almost coincident with the objective blade shape, and the variances of blade shape parameters for each model are low enough, which are illustrated by the 95% confidence interval. The average variances of 20 blade shape parameters calculated by SOGPR and MOGPR models are 1.20 and 0.94, respectively. It can be concluded that the blade shapes obtained by both inverse problem models have sufficient accuracy and low uncertainty.

Secondly, the LOO cross-validation was employed to fully confirm the reliability of the Gaussian process regression models to achieve the inverse problem calculation for the centrifugal pump blade. In the LOO cross-validation, the sample set was reconstructed including eight training samples and the prototype sample. Each blade shape was predicted by both of the constructed inverse problem models, which were trained by the other eight samples, and the prediction errors were analyzed. These nine samples

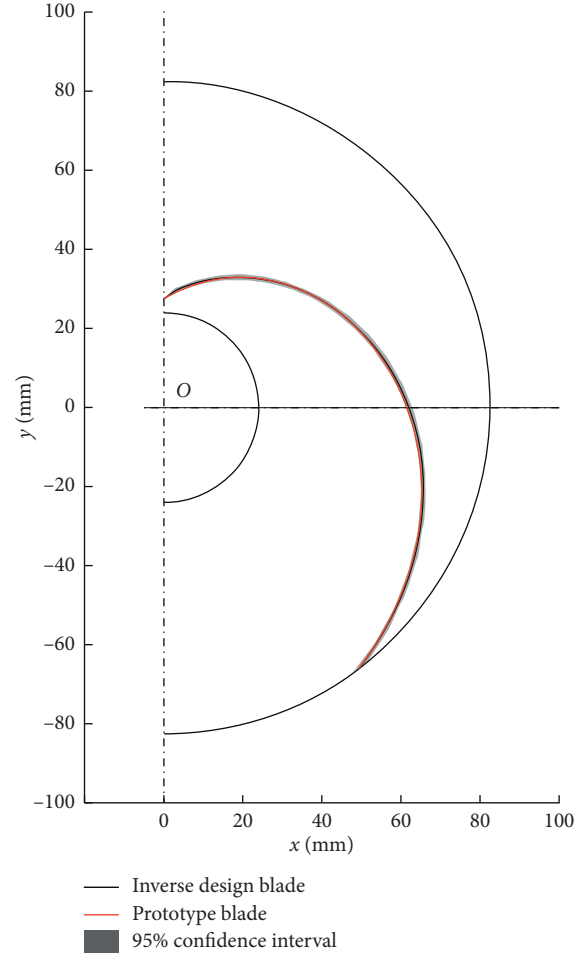


FIGURE 7: Prototype blade calculated by MOGPR.

were cross-validated, respectively. The blade shape parameters were calculated by both models, respectively, according to the blade load distribution of the test sample. Table 2 shows the values of the root mean square error (RMSE) between the inverse design blade shapes calculated by both inverse problem models and their objective blade shapes during the LOO cross-validation. The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{d=1}^N \left( \frac{\theta_d - \bar{\theta}_d}{\Phi} \right)^2} \times 100\%, \quad (10)$$

where  $\theta_d$  and  $\bar{\theta}_d$  denote the angular coordinates of the objective blade shape and the inverse design blade shape, respectively;  $\Phi$  is the wrap angle corresponding to the blade shape; and  $N$  stands for the total number of angular coordinates. We can see in Table 2 that the RMSE values of SOGPR and MOGPR models are almost within 1%, and this would be a reasonable range for pump design. It indicates that both models are robust to calculate the inverse problem of pump blade. Moreover, the RMSE values of the MOGPR model are generally lower than those of the SOGPR model. It can be concluded that the MOGPR model provides a significant improvement based on the SOGPR model. In the process of calculating blade 7 and blade 8, however, we can

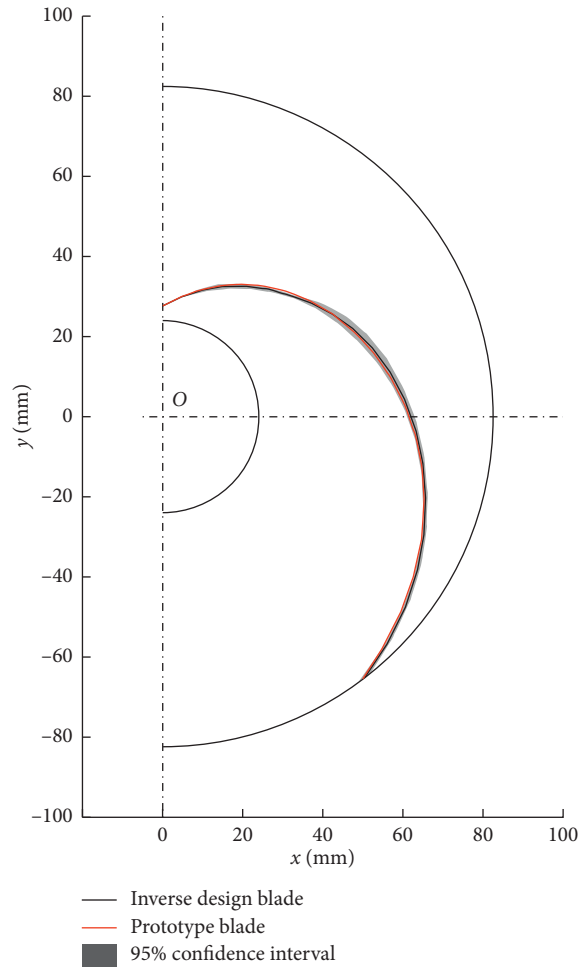


FIGURE 8: Prototype blade calculated by SOGPR.

TABLE 2: Root mean square errors in LOO cross-validation (%).

	Prototype blade	Blade 1	Blade 2	Blade 3	Blade 4	Blade 5	Blade 6	Blade 7	Blade 8
SOGPR	0.83	0.89	1.08	1.30	0.95	0.92	1.35	0.22	0.55
MOGPR	0.60	0.85	1.04	1.12	0.75	0.87	1.08	0.55	0.86

see that the accuracy of the SOGPR model is higher than that of the MOGPR model, which is not as expected. This phenomenon may be caused by the following reason: We mentioned above that the number of hyperparameters in the MOGPR model is much larger than that in the SOGPR model. During the optimization of hyperparameters in the MOGPR model, the results of optimization may only succumb to local minima in the NLML, not the global minima. As a consequence, the training of the MOGPR model is not enough, and finally, the accuracy of it is lower than that of the SOGPR model. Therefore, in order to ensure the accuracy, efficiency, and reliability of inverse problem calculation for pump blade design, reasonable selection of the initial values of hyperparameters during the model training is necessary.

Finally, an objective blade shape out of the sample space was generated to analyze the extrapolation characteristics of both inverse problem models. The blade shape out of the

sample space is shown in Figure 4, and we refer to this blade as an extrapolation blade. The sample set reconstructed in the LOO cross-validation was used to train both models. Regarding model parameter settings of both models, the characteristic length scale  $l$  was initialized to 2.68, the signal variance  $\sigma_f^2$  was initialized to 1.08, and the noise variance  $\sigma_n^2$  was initialized to 0.57. The correlation matrix of the MOGPR model was also initialized as an identity matrix. The maximum number of function evaluations was set as 800 for both models. In the gradient descent algorithm of the MOGPR model, the NLML value shows a tendency toward stabilization with the increase of iteration numbers. As shown in Figure9, we can see that the NLML value remains almost constant after 800 steps of iteration, so the optimization results of hyperparameters at 800 steps were used to calculate the inverse problem of extrapolation blade.

The blade shapes corresponding to the objective extrapolation blade load were obtained by the trained SOGPR

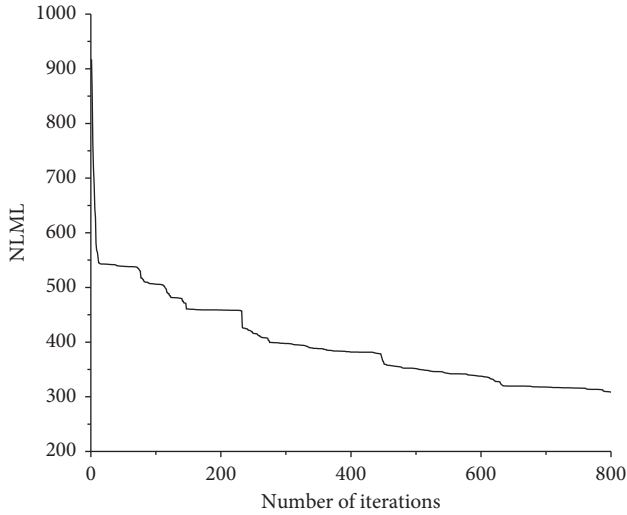


FIGURE 9: NLML value varying with the number of iterations in the MOGPR model.

and MOGPR models, respectively. We can see in Figure 10 that the extrapolation blade shape calculated by the MOGPR model almost approaches its objective blade shape, and the blade shape is smoother and continuous for the correlation between the blade shape parameters that is taken into account by the correlation matrix. Some elements of the learned correlation matrix  $\mathbf{K}_c$  are shown in equation (11). It can be seen that the correlation matrix is a symmetric matrix, and the farther the elements are from the principal diagonal, the smaller the element values are, which reveals that the correlation between any output parameters of blade shape decreases gradually with the increase of their distance. As shown in Figure 11, however, the extrapolation blade shape acquired by the SOGPR model is messy in the middle section of the blade because the 20 parameters for blade shape were calculated by the 20 SOGPR models individually and the correlation information between the blade shape parameters is ignored. Consequently, the MOGPR inverse problem model has better extrapolation characteristic than the SOGPR inverse problem model. The correlation matrix can restrict the relationship between blade shape parameters so that the blade shape characteristic is formulated more exactly. In addition, comparing Figures 7 and 10, we can find that the inverse problem of extrapolation blade has higher uncertainty than that of interpolation blade.

$$\mathbf{K}_c = \begin{bmatrix} 1 & 0.960 & 0.854 & 0.729 & 0.614 & \cdots & -0.502 \\ 0.960 & 1 & 0.964 & 0.881 & 0.768 & & -0.561 \\ 0.854 & 0.964 & 1 & 0.972 & 0.884 & & -0.584 \\ 0.729 & 0.881 & 0.972 & 1 & 0.962 & & -0.579 \\ 0.614 & 0.768 & 0.884 & 0.962 & 1 & & -0.550 \\ \vdots & & & & & \ddots & \vdots \\ -0.502 & -0.561 & -0.584 & -0.579 & -0.550 & \cdots & 1 \end{bmatrix}_{20 \times 20} \quad (11)$$

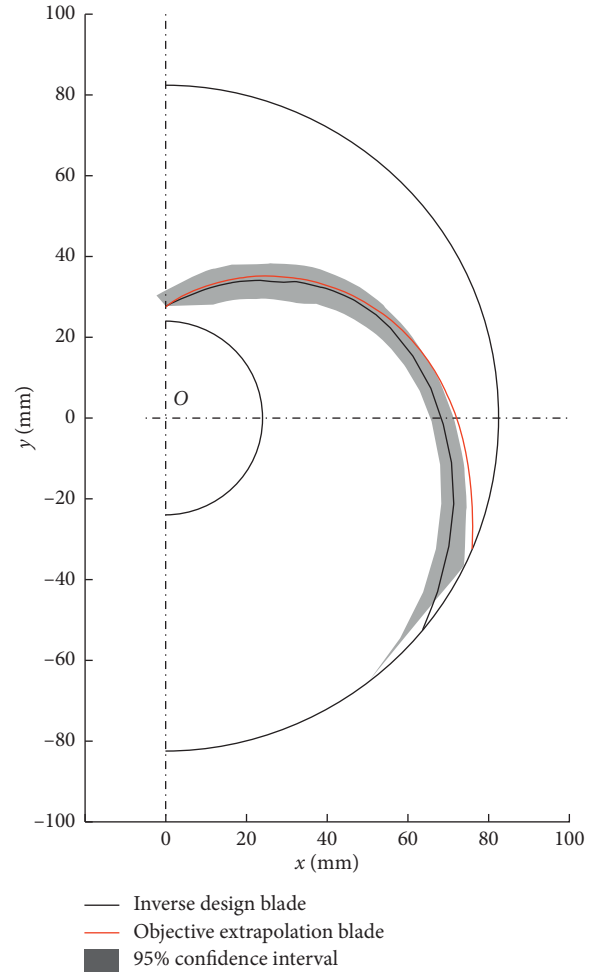


FIGURE 10: Extrapolation blade calculated by MOGPR.

## 5. Conclusions

- (1) The Gaussian process regression method was introduced to the inverse problem of centrifugal pump blade. The complicated inverse problem was converted into the problem of getting the posterior distribution from the known prior distribution based on the Bayesian theorem in the background of machine learning. Both of the inverse problem models for pump blade design have good interpolation characteristics.
- (2) The LOO cross-validation was carried out, respectively, on both models, and the results were compared and analyzed. The blade shapes within the sample space can be achieved exactly and efficiently by both of the SOGPR and MOGPR inverse problem models according to the given objective blade load distributions. Both inverse problem models are robust to calculate the inverse problem of pump blade. The RMSE values of the MOGPR inverse problem model are generally lower than those of the SOGPR inverse problem model. The research shows that the accuracy of the MOGPR inverse problem model to



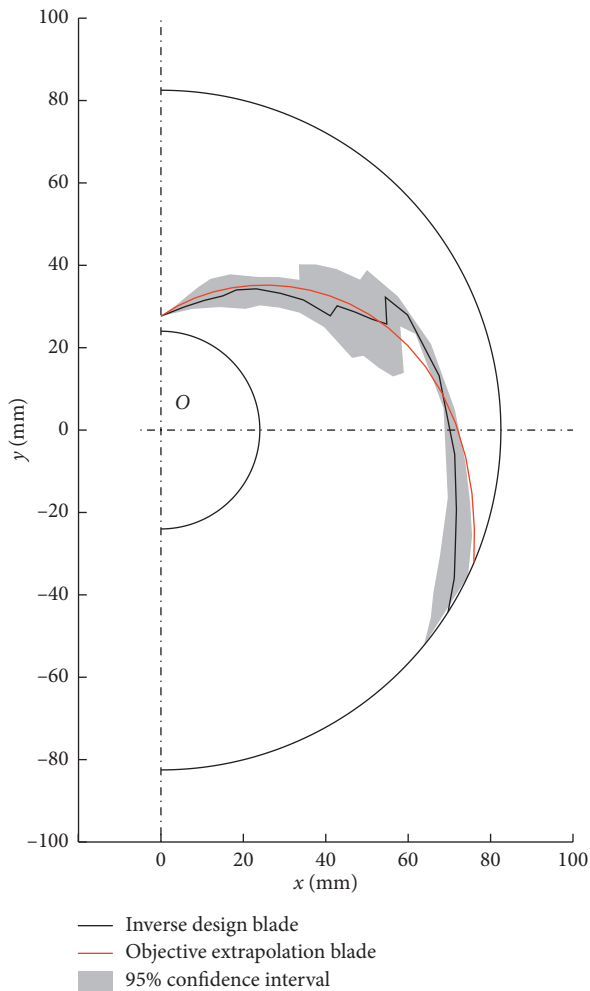


FIGURE 11: Extrapolation blade calculated by SOGPR.

calculate the inverse problem of pump blade is better than that of the SOGPR inverse problem model.

- (3) The extrapolation characteristics of both models were tested and compared. The extrapolation blade obtained by the MOGPR inverse problem model almost approaches its objective blade shape, and the blade shape is continuous and smoother. However, the extrapolation blade shape acquired by the SOGPR inverse problem model is messy, which is unable to achieve the inverse design. Since the outputs are considered dependent on each other in the MOGPR inverse problem model, the correlation between the outputs is adequately learned by the kernel function matrix. The extrapolation characteristic of the MOGPR inverse problem model is much better than that of the SOGPR inverse problem model. In addition, the accuracy of the proposed inverse methods for the interpolation blade is higher than that for the extrapolation blade.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was funded by the National Key Research and Development Program of China (grant no. 2016YFB0200901), National Natural Science Foundation of China (grant no. 51979135), and Longyuan Young Innovative Talents Program.

## References

- [1] S. R. Shah, S. V. Jain, R. N. Patel, and V. J. Lakhera, "CFD for centrifugal pumps: a review of the state-of-the-art," *Procedia Engineering*, vol. 51, pp. 715–720, 2013.
- [2] M. H. Shojaeefard, M. Tahani, M. B. Ehghaghi, M. A. Fallahian, and M. Beglari, "Numerical study of the effects of some geometric characteristics of a centrifugal pump impeller that pumps a viscous fluid," *Computers & Fluids*, vol. 60, pp. 61–70, 2012.
- [3] Y. X. Bai, F. Y. Kong, F. Zhao et al., "Experiments and numerical simulation of performances and internal flow for high-speed rescue pump with variable speeds," *Mathematical Problems in Engineering*, vol. 2018, Article ID 9168694, 18 pages, 2018.
- [4] J. E. Borges, "A three-dimensional inverse method for turbomachinery: Part I-theory," *Journal of Turbomachinery*, vol. 112, no. 3, pp. 346–354, 1990.
- [5] M. Zangeneh and A. Goto, "A compressible three-dimensional design method for radial and mixed flow turbomachinery blades," *International Journal for Numerical Methods in Fluids*, vol. 13, no. 5, pp. 599–624, 1991.
- [6] M. Zangeneh, A. Goto, and T. Takemura, "Suppression of secondary flows in a mixed-flow pump impeller by application of three-dimensional inverse design method: Part I-design and numerical validation," *Journal of Turbomachinery*, vol. 118, no. 3, pp. 536–543, 1996.
- [7] M. Zangeneh, A. Goto, and H. Harada, "On the role of three-dimensional inverse design methods in turbomachinery shape optimization," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 213, no. 1, pp. 27–42, 1999.
- [8] W. R. Hawthorne, C. Wang, C. S. Tan, and J. E. McCune, "Theory of blade design for large deflections: Part I—two-dimensional cascade," *Journal of Engineering for Gas Turbines and Power*, vol. 106, no. 2, pp. 346–353, 1984.
- [9] C. S. Tan, W. R. Hawthorne, J. E. McCune, and C. Wang, "Theory of blade design for large deflections: Part II—Annular cascades," *Journal of Engineering for Gas Turbines and Power*, vol. 106, no. 2, pp. 354–365, 1984.
- [10] D. Bonaiuti and M. Zangeneh, "On the coupling of inverse design and optimization techniques for the multiobjective, multipoint design of turbomachinery blade," *Journal of Turbomachinery*, vol. 131, no. 2, Article ID 2950065, 2009.
- [11] D. Bonaiuti, M. Zangeneh, R. Aartojarvi, and J. Eriksson, "Parametric design of a waterjet pump by means of inverse design, CFD calculations and experimental analyses," *Journal of Fluids Engineering*, vol. 132, no. 3, pp. 031104–0311018, 2010.
- [12] C. H. Wu, "A General theory of three-dimensional flow in subsonic and supersonic turbomachines of axial, radial, and mixed-flow types," *Technical Report Archive & Image Library*, vol. 1, no. 1, 1952.

- [13] J. Wu, K. Shimmei, K. Tani, K. Niikura, and J. Sato, "CFD-based design optimization for hydro turbines," *Journal of Fluids Engineering*, vol. 129, no. 2, pp. 159–168, 2007.
- [14] L. Tan, S. L. Cao, S. B. Gui, and B. S. Zhu, "Centrifugal pump impeller design by using direct inverse problem iteration," *Transactions of the Chinese Society of Agricultural Machinery*, vol. 41, no. 7, pp. 30–35, 2010.
- [15] S. Cao, G. Peng, and Z. Yu, "Hydrodynamic design of rotodynamic pump impeller for multiphase pumping by combined approach of inverse design and CFD analysis," *Journal of Fluids Engineering*, vol. 127, no. 2, pp. 330–338, 2005.
- [16] N. P. Kruyt and R. W. Westra, "On the inverse problem of blade design for centrifugal pumps and fans," *Inverse Problems*, vol. 30, no. 6, Article ID 065003, 2014.
- [17] G. Peng, S. Cao, M. Ishizuka, and S. Hayama, "Design optimization of axial flow hydraulic turbine runner: Part I: an improved Q3D inverse method," *International Journal for Numerical Methods in Fluids*, vol. 39, no. 6, pp. 517–531, 2002.
- [18] R.-H. Zhang, R. Guo, J.-H. Yang, and J.-Q. Luo, "Inverse method of centrifugal pump impeller based on proper orthogonal decomposition (POD) method," *Chinese Journal of Mechanical Engineering*, vol. 30, no. 4, pp. 1025–1031, 2017.
- [19] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, USA, 2006.
- [20] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems*, pp. 514–520, MIT Press, Cambridge, MA, USA, 1996.
- [21] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, Article ID 20110550, 2013.
- [22] R. Durichen, M. A. F. Pimentel, L. Clifton, A. Schweikard, and D. A. Clifton, "Multitask Gaussian processes for multivariate physiological time-series analysis," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 1, pp. 314–322, 2015.
- [23] E. V. Bonilla, K. M. A. Chai, and C. K. I. Williams, "Multi-task Gaussian process prediction," in *Advances in Neural Information Processing Systems*, pp. 153–160, MIT Press, Cambridge, MA, USA, 2008.
- [24] A. Melkumyan and F. Ramos, "Multi-kernel Gaussian processes," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pp. 1408–1413, Barcelona, Spain, July 2011.
- [25] M. A. Alvarez, N. D. Lawrence, and E. Rasmussen, "Computationally efficient convolved multiple output Gaussian processes," *Journal of Machine Learning Research*, vol. 12, no. 4, pp. 1459–1500, 2011.
- [26] H. Liu, J. Cai, and Y.-S. Ong, "Remarks on multi-output Gaussian process regression," *Knowledge-Based Systems*, vol. 144, pp. 102–121, 2018.
- [27] X. Liu, Q. Zhu, and H. Lu, "Modeling multiresponse surfaces for airfoil design with multiple-output-Gaussian-process regression," *Journal of Aircraft*, vol. 51, no. 3, pp. 740–747, 2014.
- [28] K. M. Chai, S. Klanke, C. Williams, and S. Vijayakumar, "Multi-task Gaussian process learning of robot inverse dynamics," in *Advances in Neural Information Processing Systems*, pp. 265–272, MIT Press, Cambridge, MA, USA, 2009.
- [29] K. Z. Wu, X. J. Liu, and H. Q. Lu, "A supercritical airfoil design based on multi-output surrogate model," *Acta Aerodynamica Sinica*, vol. 33, no. 6, pp. 728–732, 2015.
- [30] Z. Chen and B. Wang, "How priors of initial hyperparameters affect Gaussian process regression models," *Neurocomputing*, vol. 275, pp. 1702–1710, 2018.
- [31] R. H. Zhang, M. Guo, J. H. Yang, Y. Liu, and R. N. Li, "New inverse method of centrifugal pump blade based on free form deformation," *Transactions of the Chinese Society of Agricultural Machinery*, vol. 45, no. 1, pp. 84–88, 2014.
- [32] R. H. Zhang, J. C. Fan, J. H. Yang, and R. N. Li, "Optimization of centrifugal pump blade load based on free form deformation," *Transactions of the Chinese Society of Agricultural Machinery*, vol. 46, no. 10, pp. 38–43, 2015.
- [33] Z. P. Liu, *Integrated Aerodynamics-Steath Optimal Design of Airfoil Based on Machine-Learning*, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2016.