lubricants

MDPI

Article

Multi-Source Data Fusion-Driven Performance Prediction and Method Evaluation for Spiral Groove Dry Gas Seal

Jiashu Yu, Xuexing Ding and Jianping Yu *

College of Petrochemical Engineering, Lanzhou University of Technology, Lanzhou 730050, China; 231080706001@lut.edu.cn (J.Y.); dingxxseal@126.com (X.D.)
* Correspondence: yujp@lut.edu.cn

Abstract

Spiral-groove dry gas seals are widely used in various rotating machinery, and their performance prediction is of great significance for structural design and operational optimization. Existing studies still face several limitations, including the limited fidelity of numerical simulations, the insufficient number of experimental samples, and the restricted generalization capability of models based on a single data source. To address these issues, this study constructed a multi-source data system integrating numerical simulation data and experimental data, and systematically compared four representative data fusion methods, namely the uncertainty-weighted fusion algorithm, TrAdaBoost, MFDNN, and CoKriging, with analysis of their applicability and predictive performance. The results show that multi-source data fusion can effectively exploit the complementary advantages of different data sources and improve the prediction accuracy of dry gas seal performance. In terms of the comparison of data fusion methods, all four methods achieved good results for the groove-depth problem; however, for the spiral-angle and groove-number problems, which exhibit stronger nonlinear characteristics, clear differences were observed among the methods. Among them, TrAdaBoost showed the best overall performance, followed by MFDNN, then CoKriging, while the uncertainty-weighted method was relatively weaker. In terms of seal performance, the influence of groove depth on seal performance was relatively direct; the spiral angle is recommended to be controlled within 10–14°, and the groove number within 12–16, so as to balance opening force and leakage rate. This study can provide a reference for the rapid performance prediction and parameter optimization of spiral-groove dry gas seals.

Keywords: dry gas seal; data fusion; performance prediction; MFDNN; CoKriging; TrAdaBoost

Check for updates

Received: 7 April 2026
Revised: 26 April 2026
Accepted: 27 April 2026
Published: 28 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Dry gas seals (DGSs) have been widely used in rotating machinery such as compressors and turbomachinery because of their excellent sealing performance and suitability for high-speed and high-pressure operating conditions [1,2]. The working mechanism of a dry gas seal is essentially that a high-speed gas flow between the two seal faces forms a hydrodynamic gas film with load-carrying capacity within a micron-scale gap, thereby separating the faces and enabling non-contact sealing [3,4]. Therefore, the geometric parameters of the seal faces, together with the operating conditions, jointly determine key performance indicators such as the load-carrying capacity of the gas film, opening force, and leakage rate [5]. Performance prediction of spiral-groove dry gas seals not only helps reveal the influence of structural and operating parameters on sealing characteristics [6]

but also provides a basis for structural design, parameter matching, and performance optimization [7,8].

In engineering practice, dry gas seals usually involve design problems characterized by the coupling of multiple parameters. Therefore, how to achieve rapid and accurate performance prediction for spiral-groove dry gas seals has become an issue of considerable interest. Although the structural configuration of a spiral-groove dry gas seal is not particularly complicated, its performance prediction is far from straightforward. First, the thickness of the sealing gas film is usually on the micron scale, and the flow process exhibits pronounced microscale characteristics. Gas compressibility, rarefaction effects, and wall slip can all influence the pressure distribution and leakage behavior of the gas film [5,9,10]. Second, the gas-film flow field shows significant non-uniformity. Under different groove depths, spiral angles, groove numbers, and rotational speeds [4,5,11], the variations in the pressure field and velocity field often exhibit strong nonlinear characteristics [6,12]. In addition, during actual operation, structural parameters and operating parameters often vary simultaneously, resulting in a complex flow-field state. Their influences on opening force and leakage rate are not simply linearly superimposed [13,14].

A large number of studies have been carried out on the flow mechanism, pressure distribution, load-carrying characteristics, and leakage behavior of spiral-groove dry gas seals. The main approaches include theoretical analysis, numerical simulation, and experimental investigation. These studies have provided an important foundation for understanding the gas-film flow behavior between sealing faces. For example, Sneck and McGrew [15] conducted early theoretical research on spiral-groove face seals; Ruan [3], as well as Miller and Green, developed relevant numerical analysis models; and Ding [16], Shi [17], and Chen [14], among others, investigated the temperature field and dynamic characteristics under high-speed and high-pressure operating conditions. However, in general, traditional methods are more suitable for analyzing seal performance under a specific structural configuration or a specific operating condition. They are usually computationally expensive and have difficulty achieving rapid prediction and analysis when parameters vary over a wide range. Therefore, relying solely on traditional methods is no longer sufficient to meet the demands for performance prediction of spiral-groove dry gas seals under conditions involving multiple parameters, wide parameter ranges, and high efficiency requirements.

A review of the existing literature shows that research on spiral groove dry gas seals has evolved from classical mechanism analysis toward structural optimization and rapid performance prediction. Early studies mainly focused on the fundamental operating mechanism and structural characteristics of noncontact spiral groove face seals, clarifying the key role of spiral grooves in establishing hydrodynamic gas-film pressure and separating the sealing faces, thereby laying the theoretical foundation for subsequent research [18]. On this basis, Salant and Homiller analyzed the stiffness and leakage characteristics of upstream-pumping spiral groove mechanical seals, demonstrating the significant influence of groove geometry on both the static and dynamic performance of the seal [19]. Faria further proposed a finite element method suitable for the analysis of high-speed spiral groove gas face seals, which improved the efficiency of numerical analysis under complex operating conditions [20]. These studies advanced the understanding of spiral groove dry gas seal performance from the perspectives of theoretical modeling and numerical solution. However, they were still primarily concerned with mechanism elucidation and the analysis of individual structural parameters, rather than rapid prediction over a wide multi-parameter range [18–20].

In recent years, greater attention has been paid to face-texture design, performance trade-offs, and data-driven prediction. For example, Jiang et al. [21] investigated the effects of different surface texture forms on opening performance, gas-film stability, and

wear resistance based on a superposed groove model, and compared the steady-state and dynamic characteristics of different textured structures. Chen et al. [22], on the basis of conventional gas-film pressure equation analysis, combined Latin hypercube sampling with a radial basis function neural network to predict the performance of spiral groove dry gas seals and optimize their structural parameters. These studies indicate that research on spiral groove dry gas seals has begun to extend from purely mechanistic analysis toward a direction integrating numerical simulation, surrogate modeling, and parameter optimization. Nevertheless, most existing studies still rely on a single modeling route and usually establish predictive models based on a single data source, with insufficient exploitation of the complementary strengths between CFD data and experimental data [21,22]. Meanwhile, review studies in the fields of multi-fidelity modeling and engineering surrogate modeling have shown that multi-source and multi-fidelity data fusion can maintain good predictive capability while reducing the demand for high-fidelity samples, especially for engineering problems characterized by high cost, strong nonlinearity, and imbalanced data [23,24]. However, according to the publicly available literature, there is still a lack of systematic studies that place representative data fusion methods, such as weighted fusion, transfer learning, deep neural networks, and CoKriging, within a unified framework and comprehensively compare their performance and applicability in predicting the opening force and leakage rate of spiral groove dry gas seals. The present study is intended to address this gap.

Against this background, the use of data fusion methods for performance prediction of spiral-groove dry gas seals has become a feasible approach to improve prediction efficiency and expand the scope of analysis. In recent years, methods such as machine learning and surrogate models have gradually been applied to performance prediction in fluid machinery, friction seals, and related engineering problems, showing good nonlinear fitting capability and advantages in rapid prediction. However, for spiral-groove dry gas seals, relying solely on data from a single source still has obvious limitations [12]. On the one hand, CFD data are relatively easy to obtain and can be generated in large quantities, allowing a wide parameter range to be covered. However, the results are inevitably affected by factors such as model simplification and boundary conditions [25]. On the other hand, experimental data are closer to actual operating conditions and therefore generally more reliable, but they are costly and time-consuming to obtain, and the number of available samples is limited, making it difficult for them to independently support large-range modeling with multiple parameters [26]. Therefore, how to combine the advantages of these two types of data and construct a performance prediction model that accounts for both coverage and predictive reliability has become a natural and necessary issue.

Data fusion methods provide a new way to address the above issues. Their core idea is to comprehensively exploit the complementary information among data from different sources with different levels of fidelity or different sample sizes, so as to expand the range of sample utilization while improving the model's capability for the target problem [23,27–29]. At present, the concept of data fusion has attracted widespread attention in areas such as multi-fidelity modeling, transfer learning, surrogate modeling, and prediction of complex engineering systems, and has shown promising application potential in several related fields [23,28–31]. For the performance prediction of spiral-groove dry gas seals, the introduction of data fusion methods offers two main advantages. First, it enables full use of large-sample CFD data, thereby enhancing the model's ability to learn the overall patterns of parameter variation. Second, experimental data can be used to calibrate the model, making the prediction results closer to actual sealing operating conditions. Therefore, compared with modeling based on a single data source, multi-source

data fusion methods are more promising for achieving efficient and accurate performance prediction of spiral-groove dry gas seals.

However, data fusion can be implemented in various ways, and different methods show clear differences in modeling mechanisms, modes of information utilization, and applicable conditions. According to existing studies, relevant methods can generally be classified into several categories, including weighted fusion, transfer learning, deep learning, and multi-fidelity statistical modeling. Weighted fusion methods achieve the direct integration of multi-source information by assigning different weights to different data sources, and their formulation is relatively simple [32]. Transfer learning methods emphasize the transfer and utilization of features across different data domains and are suitable for scenarios in which sample distributions differ [32,33]. Deep learning methods are capable of handling complex input–output relationships because of their strong nonlinear mapping capability. Multi-fidelity statistical modeling methods focus on exploring the correlations among data with different levels of fidelity [23,34]. Although these methods each have distinct characteristics in principle, publicly available studies that systematically apply these representative data fusion methods to the performance prediction of spiral-groove dry gas seals are still limited. Therefore, it is necessary to select representative methods and conduct a unified analysis of their applicability to this problem.

Based on the above considerations, this study selects four representative data fusion methods: an uncertainty-weighted fusion algorithm, TrAdaBoost, MFDNN, and CoKriging. The uncertainty-weighted fusion algorithm directly defines weighting coefficients according to the uncertainty of different data sources, which facilitates direct investigation of how the confidence levels of data from different sources affect the prediction results. As a typical transfer learning method, TrAdaBoost can iteratively adjust sample weights and strengthen the influence of samples that are beneficial to the target task, making it representative in multi-source data fusion modeling. MFDNN relies on the strong nonlinear representation capability of deep neural networks and is suitable for handling complex mappings between parameter inputs and outputs. CoKriging can achieve collaborative prediction by exploiting the correlations among data with different levels of fidelity and is highly representative in engineering prediction problems. These four methods correspond to different data fusion strategies. They are both representative and clearly comparable, and are therefore well-suited as the subjects of the comparative study in this paper. The technical roadmap of this study is shown in Figure 1.

Overall, this study focuses on the prediction of opening force and leakage rate for spiral-groove dry gas seals. By integrating CFD data and experimental data, four representative data fusion models, namely an uncertainty-weighted fusion algorithm, TrAdaBoost, MFDNN, and CoKriging, are established. On this basis, the relationships between structural parameters and sealing performance are systematically investigated, and the predictive performance and applicability of the four methods are compared using a unified dataset. The results provide a clearer understanding of the characteristics and differences in different data fusion methods in spiral-groove dry gas seal performance prediction and identify their respective applicable scenarios. This work is expected to provide support for the transition of dry gas seal research from single-model-based analysis to efficient and intelligent prediction based on multi-source information fusion. The main contributions of this study are as follows:

A multi-source dataset for spiral groove dry gas seals was constructed by integrating numerical simulation data with experimental data.

Within a unified framework, four representative multi-source data fusion methods, namely UWF, TrAdaBoost, MFDNN, and CoKriging, were implemented and compared.

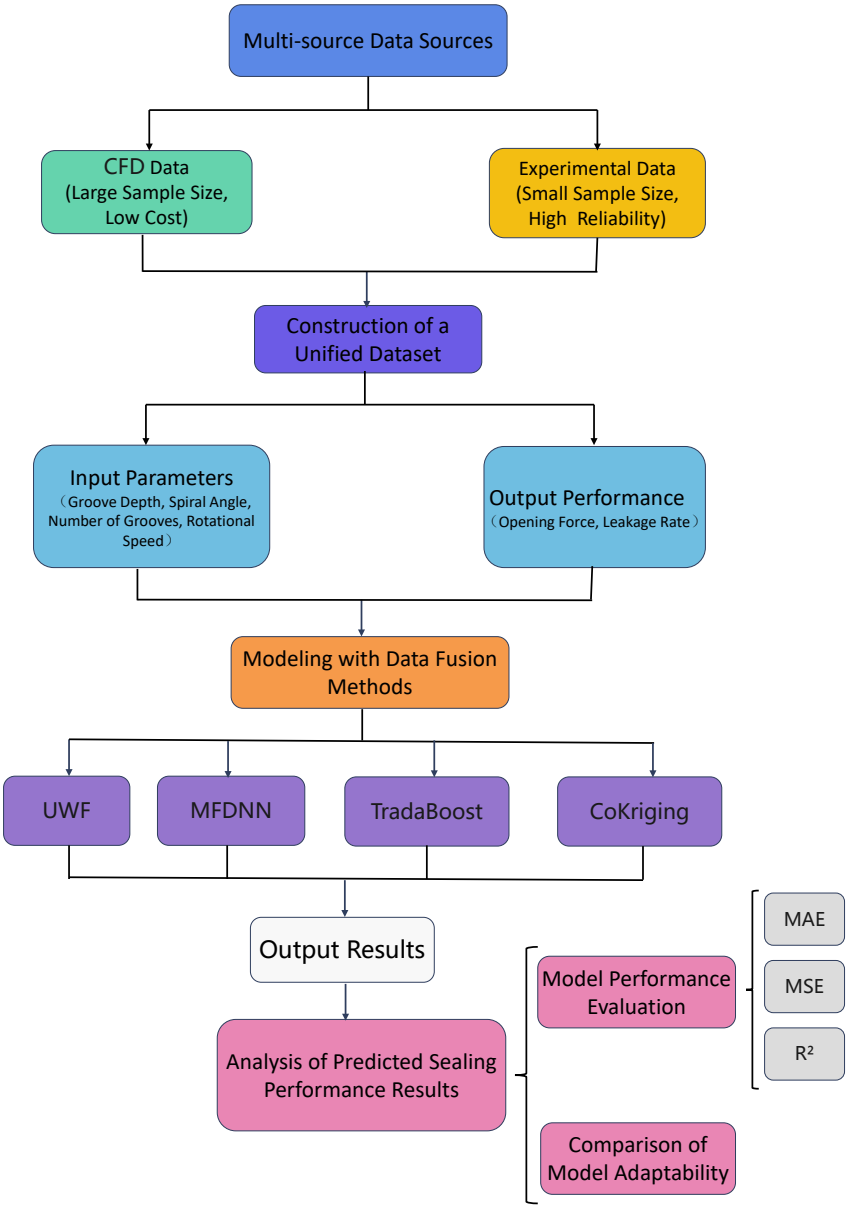


Figure 1. Technical roadmap of this study.

For two key performance indicators, namely opening force and leakage rate, the predictive capability of different fusion methods under different parameter combinations was systematically evaluated.

Considering the nonlinear characteristics of different prediction tasks, the applicability and relative advantages of the different fusion methods were analyzed.

2. Data Acquisition and Processing

2.1. Definition of Geometric Parameters of Spiral-Groove Dry Gas Seals

A dry gas seal consists of a rotating ring and a stationary ring. Its core principle is that the micron-scale spiral grooves on the rotating ring generate hydrodynamic gas pressure during high-speed rotation, thereby forming a gas film that is also on the micron scale. This gas film separates the rotating ring from the stationary ring and thus enables sealing under high-speed rotating conditions. Therefore, the spiral-groove structure on the rotating ring is a key factor affecting the sealing performance of the gas film. Figure 2 shows the schematic structure of the dry gas seal.

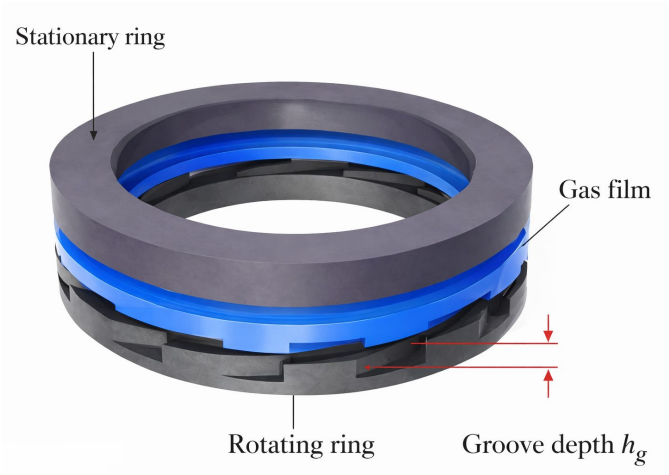


Figure 2. Schematic structure of the dry gas seal.

In this study, groove depth, spiral angle, and groove number are selected as structural design variables, while rotational speed is treated as the operating condition variable. Opening force and leakage rate are taken as the performance response variables. These variables are chosen because they constitute the most critical parameters in the performance analysis of spiral-grooved dry gas seals. The structural parameters determine the geometric pumping capability of the gas film and the mechanism of pressure generation, while the rotational speed directly influences the hydrodynamic effect of the gas film. Together, these factors govern the load-carrying capacity and leakage performance of the seal.

The geometric definitions of the structural parameters of the spiral grooves in the dry gas seal are given in Figure 3 and Table 1. The parameterization method of the spiral-groove structure and the corresponding mathematical formulation are available in Ref. [7] and are not repeated here.

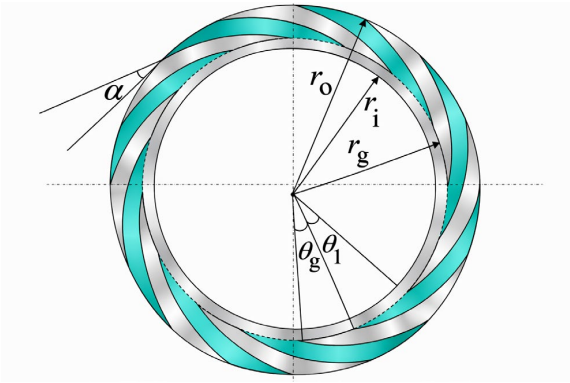


Figure 3. Structural parameters of the spiral-groove dry gas seal.

Table 1. The geometry parameters.

Parameters	Definitions
r_i	Inner radius of the rotating ring
r_o	Outer radius of the rotating ring
r_g	Groove root radius
α	Inlet spiral angle
θ_g	Circumferential angle of the groove region
θ_1	Circumferential angle of the land region
N_g	Number of spiral grooves
h_g	Groove depth

2.2. Numerical Calculation Method

Figure 4 shows the numerical calculation procedure used in this study. Detailed descriptions of the mesh generation, boundary conditions, and validation of the numerical method are available in Section 3 of our previous study [35].

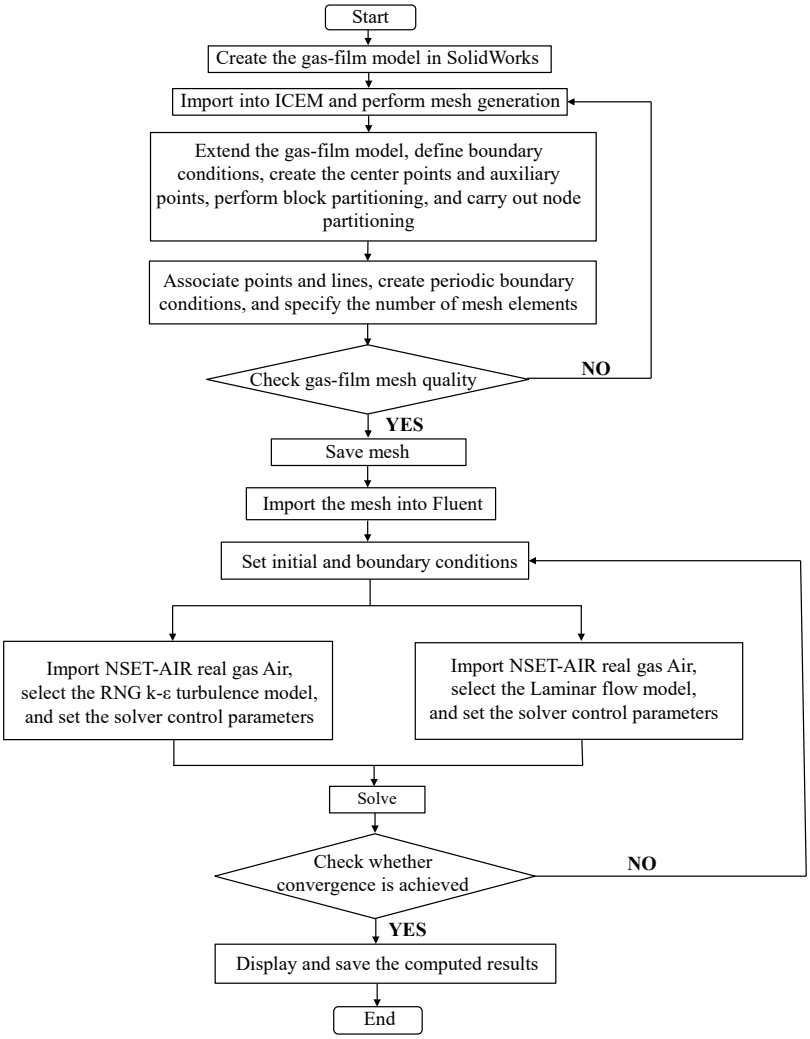


Figure 4. Flowchart of the numerical calculation procedure.

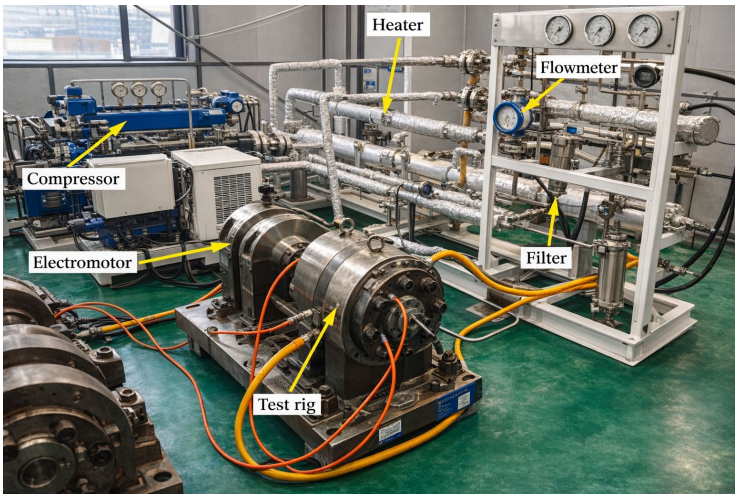
2.3. Experimental Method

The experiments and data acquisition in this study were conducted using a high-speed seal test rig. Figure 5a,b show a photograph of the test rig and its schematic diagram, respectively. Detailed specifications of the test rig and the data acquisition procedure are available in Section 3.2.1 of Ref. [7] and are not repeated here.

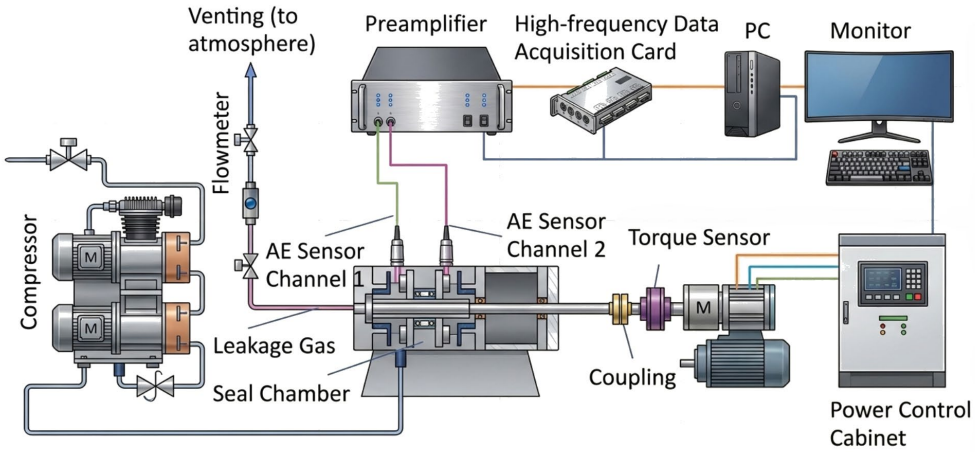
The structural parameters of the seal specimen selected for the experiments are listed in Table 2.

Table 2. The Structural parameters of the spiral groove specimen.

Parameter	Unit	Type 1	Type 2	Type 3	Type 4
α	degree	15.0	18.0	16.0	17.0
θ_g	degree	10.0	15.0	15.0	13.0
θ_1	degree	10.0	15.0	7.5	12.7
h_g	μm	6.0	5.0	7.0	6.0



(**a**) Photograph of the experimental setup



(**b**) Schematic diagram of the experimental setup

Figure 5. Experimental setup.

2.4. Dataset Generation

In the dataset, groove depth, spiral angle, and groove number were selected as the structural design variables, while rotational speed was taken as the operating-condition design variable. Opening force and leakage rate were used as the response variables. The ranges of the design variables are listed in Table 3. These ranges take into account both the representativeness of parameter variations in engineering practice and the availability of existing experimental conditions and numerical samples.

Table 3. Ranges of the design variables.

Design Variable	Parameter Range
groove depth	5–8 μm
spiral angle	5–20°
number of spiral grooves	5–30
RPM	10,000–40,000

In this study, MATLAB R2023 was used to develop programs for the four data fusion methods mentioned above, and the sample dataset was then processed using these programs. The dataset contains 271 sample points from numerical simulations and 62 sample points from experiments. Due to the inherent characteristics of the fusion algorithms, all 333 sample points are not directly used for training the surrogate models. In general,

the dataset is divided into training, validation, and test sets. The training set is used for model training, calibration, and learning; the validation set is used for hyperparameter selection and training-process control; and the test set does not participate in any training or parameter tuning, but is used only for the final performance evaluation.

Different methods use the dataset in different ways. For the uncertainty-weighted fusion method, the low-fidelity data and the high-fidelity training set are generally used to establish the low-fidelity and high-fidelity models, respectively, and the fusion weights are determined according to the predictive uncertainty of the models. The validation set is used to examine whether the uncertainty estimation is reasonable, while the test set is used for final evaluation. For the MFDNN method, the low-fidelity data are used for network pretraining and low-fidelity branch learning, while the high-fidelity training set is used for high-fidelity branch training and model fine-tuning. The validation set is used to select the network architecture and the number of training epochs, and the test set is used to evaluate the high-fidelity prediction accuracy. For the TrAdaBoost method, the low-fidelity data are treated as source-domain data, and the high-fidelity training set is treated as target-domain training data. Both are involved in the iterative training process, and sample-weight adjustment is used to reduce the influence of low-fidelity samples that differ considerably from the target domain. The high-fidelity test set is always kept independent. For the CoKriging method, the low-fidelity data are used to construct the low-fidelity surrogate model, while the high-fidelity training set is used to estimate the correlation between the high- and low-fidelity responses and the discrepancy correction term. The test set is then used to examine the final high-fidelity prediction results.

This data-partitioning strategy avoids interference from low-fidelity samples in the test results and ensures a fair comparison of different methods on the same high-fidelity test samples.

3. Data Fusion Methods

3.1. Uncertainty-Weighted Fusion Algorithm (UWF)

The core idea of the uncertainty-weighted fusion algorithm is to use Gaussian process regression (GPR) to model and quantify the predictive uncertainty of different data sources at each input location, which is usually characterized by the predictive variance or standard deviation. The weights are then determined based on this uncertainty for weighted fusion. In this way, adaptive weighted fusion of multi-source information can be achieved, leading to more robust and reliable fusion results.

3.1.1. Gaussian Process Regression

A Gaussian process is a stochastic process and has good applicability to complex problems involving small samples, nonlinearity, and high dimensionality [36]. In essence, it is a regression algorithm based on Bayesian optimization. Given a dataset D, a set of variables f(x) with a joint Gaussian distribution is defined in D, which consists of a mean function and a covariance function:

$$f(x) \sim GP(\mu(x), k(x, x')) \tag{1}$$

Considering noise, the general model of Gaussian process regression can be written as:

$$y = f(x) + \varepsilon \tag{2}$$

where ε is independent Gaussian white noise.

According to the Bayesian principle, Gaussian process regression establishes a prior function of y on the dataset $\boldsymbol{D}$. Then, the joint Gaussian distribution of the sample points and the new data point f_* can be expressed as:

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

where $\boldsymbol{K}$, $\boldsymbol{K}_*$, and $\boldsymbol{K}_{**}$ are defined as follows:

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix} \tag{4}$$

$$\mathbf{K}_* = [k(x_*, x_1) \;\; k(x_*, x_2) \;\; \cdots \;\; k(x_*, x_n)] \tag{5}$$

$$\mathbf{K}_{**} = k(x_*, x_*) \tag{6}$$

According to the properties of the Gaussian process, the posterior probability distribution of $f_*, f_* | X, y, x_*$ also follows a Gaussian distribution:

$$f_* \mid X, y, x_* \sim N(\mu_*(x), \Sigma_*) \tag{7}$$

where

$$\mu_*(x) = \mathbf{K}_* \left(\mathbf{K} + \sigma_n^2 \mathbf{I}_n\right)^{-1} y \tag{8}$$

$$\Sigma_* = \mathbf{K}_{**} - \mathbf{K}_* \left(\mathbf{K} + \sigma_n^2 \mathbf{I}_n\right)^{-1} \mathbf{K}_*^T \tag{9}$$

where $\mu_*(x)$ is the predicted output value of the new data point x_*, and Σ_* is the variance of the predicted output. In this way, the distribution characteristics of the predicted value and the uncertainty of the dataset can be obtained.

3.1.2. Weighted Fusion Algorithm

For aerodynamic data from multiple sources, Gaussian process regression is first employed to obtain the mean $\mu_{\mathrm{GPR}i}$ and variance ${\sigma^2}_{\mathrm{GPR}i}$ of each individual data source. Then, an uncertainty-based weighted fusion algorithm is further applied for data fusion. The specific flowchart is shown in Figure 6.

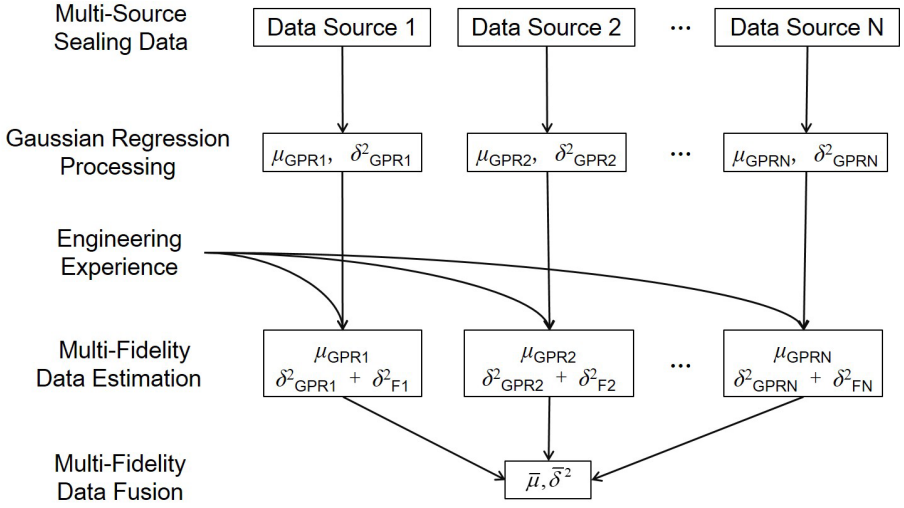


Figure 6. Schematic diagram of the uncertainty-based data fusion algorithm.

As shown in Figure 6, Gaussian process regression is first used to perform regression analysis on seal data samples from different sources, so as to obtain the uncertainty of the model. Since each data source is derived from calculation or measurement and deviates from the true value, it is necessary to determine the fidelity of each type of data according to prior information, such as experimental experience and engineering judgment. Fidelity reflects the uncertainty between the seal data samples and the true value. The higher the fidelity, the more accurate the data. The fidelity function of data source i is defined as σ^2_{Fi}. Then, the model uncertainty and the data uncertainty are combined to obtain the total uncertainty μ_{Ti} and σ^2_{Ti} of each data source, as shown in Equations (10) and (11). Finally, the total uncertainty of each type of data is fused and estimated according to the weighted fusion algorithm.

$$\mu_{T_i}(\boldsymbol{x}) = \mu_{GPR_i}(\boldsymbol{x}) \tag{10}$$

$$\sigma^2_{T_i}(\boldsymbol{x}) = \sigma^2_{GPR_i}(\boldsymbol{x}) + \sigma^2_{F_i}(\boldsymbol{x}) \tag{11}$$

For high-fidelity data, the model uncertainty is high because the number of sample points is small, whereas the uncertainty of the high-fidelity data itself is low. For low-fidelity data, the model uncertainty is low because the number of sample points is large, whereas the uncertainty of the low-fidelity data itself is high. Through weighted fusion, data with higher accuracy and lower uncertainty can be obtained. The weighted fusion algorithm assigns a different weighting coefficient w to each data source and then combines them by summation. When the data sources are mutually independent, the weighting coefficient is inversely proportional to the uncertainty variance of the data source. That is, the higher the variance, the lower the data accuracy, and thus the smaller the weighting coefficient. The specific expression is as follows:

$$w_i = \frac{\dfrac{1}{\sigma^2_{T_i}(\boldsymbol{x})}}{\sum\limits_{i=1}^{N} \dfrac{1}{\sigma^2_{T_i}(\boldsymbol{x})}} \tag{12}$$

$$\mu(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \mu_{T_i}(\boldsymbol{x}) \tag{13}$$

$$\sigma^2(\boldsymbol{x}) = \left(\sum_{i=1}^{N} \frac{1}{\sigma^2_{T_i}(\boldsymbol{x})} \right)^{-1} \tag{14}$$

3.2. Data Fusion Approach Based on Transfer Learning

The data fusion approach based on transfer learning (DFATL) refers to the integration of low-fidelity and high-fidelity data to improve the performance and generalization capability of predictive models built on high-fidelity data. Transfer learning is used to learn transferable features or model parameters from low-fidelity data and transfer them to the high-fidelity domain, thereby alleviating the modeling difficulty caused by the scarcity of high-fidelity data. Data fusion, in turn, jointly models low-fidelity and high-fidelity data within a unified framework, so as to improve the prediction accuracy and generalization performance of the model.

The DFATL method mainly includes sample-level fusion and model-level fusion.

3.2.1. Multi-Fidelity Fusion Method Based on Sample Features

The sample-based transfer-learning data fusion method mainly considers fusing the features of low-fidelity and high-fidelity data to obtain a unified feature space. Based on this idea, a feature-weighted fusion method is adopted in this study. CFD simulation data

are taken as the low-fidelity data (superscript l), and seal experimental data are taken as the high-fidelity data (superscript h), to construct the fusion loss function L:

$$L = \sum \frac{1}{m}\left(y_i^h - y_i^{yc}\right)^2 + \sum_i w_i^l \frac{1}{n}\left(y_i^l - y_i^{yc}\right)^2 \tag{15}$$

where w_i is the weighting coefficient; m is the number of seal experimental data samples; y_i is the sample value; and n is the number of CFD simulation data samples.

Each CFD sample is introduced into the evaluation of the loss function and assigned a corresponding weight. Through continuous iteration until the loss function is minimized, feature fusion is achieved.

The most critical issue in achieving sample-level fusion is the rapid acquisition of the optimal feature weighting coefficients. In this study, the TrAdaBoost algorithm is employed for this purpose [37]. The TrAdaBoost algorithm enhances the generalization capability of the model by introducing a weight adjustment process and combining multiple weak regression models into a strong regression model through an iterative procedure. In each iteration, the sample weights are adjusted according to the prediction results of the previous iteration, so that the samples with larger prediction errors in the previous iteration receive more attention in the current iteration.

Assume that the number of iterations is t = 1, 2, . . ., N. In the iterative process of the TrAdaBoost algorithm, the sample weights are first initialized, and the weight of each sample is set as:

$$w_i^t = \frac{1}{m+n} \tag{16}$$

During the iterative process, a weak regression model is trained using the current sample weights. The trained weak regression model is then used for prediction, and the error eti of each sample is calculated.

The adjustment of sample weights is mainly divided into two stages. In the first stage, the weights of the low-fidelity data are gradually decreased until a critical point is reached, which is determined through cross-validation. In the second stage, the weights of the high-fidelity data are updated, and the weight is given by:

$$w_i^{t+1} = \begin{cases} \frac{w_i^t \beta_t}{P_t} & (1 \le i \le n) \\ \frac{w_i^t}{P_t} & (n+1 \le i \le n+m) \end{cases} \tag{17}$$

where

$$\beta_t = \frac{m}{m+n} + \frac{t}{N-1}\left(1 - \frac{m}{n+m}\right) \tag{18}$$

where Pt is the sum of the sample weights, and βt is the weighting factor.

Finally, during the iterative process, when eti reaches the minimum value or the maximum number of iterations is reached, the prediction results of each weak regression model are weighted and summed according to their corresponding weights to obtain the final prediction result of the regression model.

By continuously adjusting the sample weights and combining the prediction results of multiple weak regression models, the TrAdaBoost algorithm can not only achieve weighted fusion based on sample features but also effectively address problems such as overfitting and underfitting in regression tasks.

3.2.2. Multi-Fidelity Fusion Method Based on Model Features

The model-based transfer-learning data fusion method is a method that utilizes existing model features to assist the fusion of high-fidelity data. It can be trained on a

low-fidelity model and transfer the data features within the low-fidelity range to the target range, so as to improve the prediction performance in the high-fidelity range.

In this study, a high- and low-level neural network modeling strategy is adopted. A deep neural network model (MFDNN) is used to introduce multi-source data during model construction, and a hierarchical training strategy is employed to realize the data fusion method [31]. Figure 7 shows the high- and low-fidelity deep neural network fusion framework for CFD data and seal experimental data.

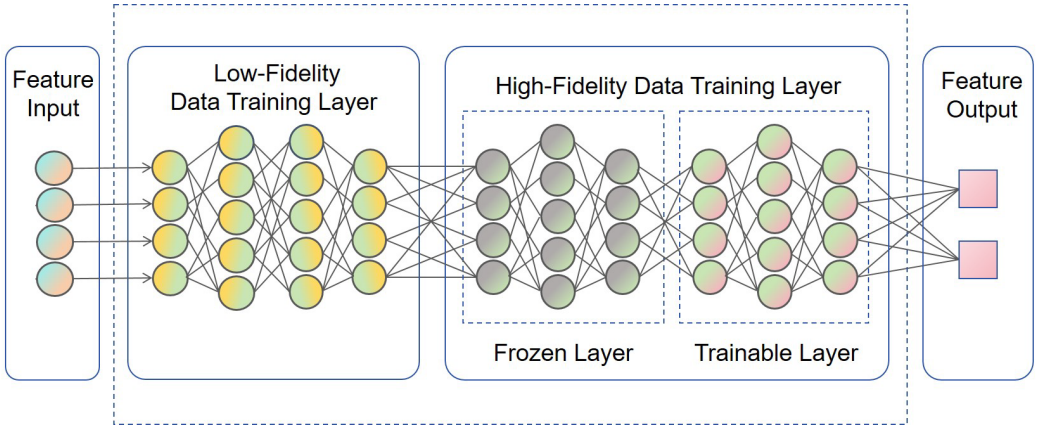


Figure 7. Neural network framework based on data fusion.

First, the CFD data and seal experimental data are organized into forms suitable for input into the neural network model, including preprocessing steps such as feature extraction and data normalization. Second, a deep neural network model is constructed and divided into a low-fidelity layer and a high-fidelity layer. Specifically, the CFD data are used to construct the low-fidelity layer, and the experimental data are used to construct the high-fidelity layer. Assume that the input neurons of the network layer are $\{x_1, \ldots, x_i\}$. The neural network generates the corresponding weight coefficients $\{w_1, \ldots, w_i\}$ through training, and the output neuron can be expressed as:

$$y = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{19}$$

where $\sigma(\cdot)$ is a nonlinear activation function.

Then, the high-fidelity layer is frozen so that the network weight coefficients of the high-fidelity layer remain fixed, while the low-fidelity layer participates in model construction. The CFD dataset is introduced to construct a low-fidelity neural network model and obtain the model output results. Finally, the high-fidelity layer is unfrozen, and the model output results are incorporated into the seal experimental data to build the high-fidelity neural network. The loss function of the model can be expressed as:

$$L = L_l + L_h + \lambda \sum w_i^2 \tag{20}$$

where L_l and L_h are the loss function values of the low-fidelity layer and the high-fidelity layer, respectively, and λ is the regularization coefficient.

A global model is ultimately formed after hierarchical training, and the final results are predicted according to the feature information. This neural network construction method can make full use of information from multi-source data and progressively integrate the features of different data sources through hierarchical training, thereby improving the generalization capability and prediction accuracy of the model. At the same time, the hierarchical structure and nonlinear mapping capability of the deep neural network can

effectively explore the complexity and correlation of high-fidelity and low-fidelity data, thereby enhancing the representational capability of the fusion results and the extraction capability of latent information.

3.3. Fusion Algorithm Based on the CoKriging Surrogate Model

A surrogate model refers to an approximate model established using a limited number of sample data. Its advantage lies in its ability to replace time-consuming physical analysis models during the processes of analysis and optimization. However, the prediction accuracy of a surrogate model based on a single data source is, to a large extent, constrained by the number of sample points. The larger the number of sample points, the easier it is to achieve good fitting and prediction results. However, obtaining accurate and highly reliable fitting for dry gas seals requires a large number of sample points, which makes the construction of a high-accuracy surrogate model based on a single data source costly and inefficient.

To this end, researchers have proposed in recent years an improved data fusion model based on the Kriging model, namely the CoKriging model [38]. The CoKriging surrogate model is a multi-source extension of the Kriging model. It can be constructed using a small number of highly reliable samples together with a large number of easily obtainable low-reliability samples, thereby reducing the cost of surrogate model construction while maintaining prediction accuracy.

For a problem with m design variables, high-reliability and low-reliability samples are extracted, respectively:

$$\begin{cases} \boldsymbol{X}_e = [\, \boldsymbol{x}_e^{(1)} \;\; \boldsymbol{x}_e^{(2)} \;\; \cdots \;\; \boldsymbol{x}_e^{(n_e)} \,]^T \\ \boldsymbol{X}_c = [\, \boldsymbol{x}_c^{(1)} \;\; \boldsymbol{x}_c^{(2)} \;\; \cdots \;\; \boldsymbol{x}_c^{(n_c)} \,]^T \end{cases} \tag{21}$$

where the subscripts "e" and "c" denote the high-reliability and low-reliability samples, respectively; n is the number of samples (assuming ne < nc).

The response values corresponding to the high-reliability and low-reliability samples are:

$$\begin{cases} \boldsymbol{y}_e = [\, y_e^{(1)} \;\; y_e^{(2)} \;\; \cdots \;\; y_e^{(n_e)} \,]^T \\ \boldsymbol{y}_c = [\, y_c^{(1)} \;\; y_c^{(2)} \;\; \cdots \;\; y_c^{(n_c)} \,]^T \end{cases} \tag{22}$$

Let the Gaussian stochastic processes Ze(·) and Zc(·) represent the characteristics of the high-reliability and low-reliability data, respectively. Then, we have:

$$Z_e(\boldsymbol{x}) = \rho Z_c(\boldsymbol{x}) + Z_d(\boldsymbol{x}) \tag{23}$$

where ρ is a hyperparameter, and Zd(x) is the Gaussian stochastic process of the difference between Ze(x) and ρZc(x).

The difference vector between the high-reliability data and the low-reliability data is defined as:

$$\boldsymbol{d} = \boldsymbol{y}_e - \rho \boldsymbol{y}_c(\boldsymbol{X}_e) \tag{24}$$

where yc(Xe) is the value of the low-reliability model at point Xe.

The covariance matrix of the CoKriging model can be expressed as:

$$C = \begin{bmatrix} \sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{X}_c) & \rho\sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{X}_e) \\ \rho\sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_e, \boldsymbol{X}_c) & \rho^2\sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_e, \boldsymbol{X}_e) + \sigma_d^2 \boldsymbol{R}_d(\boldsymbol{X}_e, \boldsymbol{X}_e) \end{bmatrix} \tag{25}$$

where σ_c^2 and σ_d^2 are the mean square variances of the random variables yc(x) and d(x), respectively, and Rd and Rc are the covariance matrices of the corresponding stochastic processes, respectively.

When the correlation function is chosen as the Gaussian exponential model, the correlation function of the sample points can be expressed as:

$$\boldsymbol{R}_{ij} = \exp\left(-\sum_{k=1}^{m} \theta_k \left|\left| \boldsymbol{x}_k^{(j)} - \boldsymbol{x}_k^{(i)} \right|\right|^{p_k} \right) \tag{26}$$

The CoKriging surrogate model has two different correlation functions, and the hyperparameters θc, θd, pc, pd, and ρ need to be estimated. Assuming that the two levels of reliability data are mutually independent, the estimation of θc and pc is the same as that in the standard Kriging method, whereas θd, pd, and ρ can be estimated using numerical optimization methods similar to those employed in standard Kriging.

The predicted value of the CoKriging model at an unknown point x* can be expressed as:

$$\hat{y}_e(\boldsymbol{x}^*) = \hat{\mu} + \boldsymbol{c}^T \boldsymbol{C}^{-1}(\boldsymbol{y} - 1\hat{\mu}) \tag{27}$$

where $\hat{\mu}$ is the mean value of ye(x).

$$\hat{\mu} = \frac{1^T \boldsymbol{C}^{-1} \boldsymbol{y}}{1^T \boldsymbol{C}^{-1} 1} \tag{28}$$

c is the column vector of the covariance between the unknown point x* and the known sample points:

$$\boldsymbol{c} = \begin{bmatrix} \rho \hat{\sigma}_c^2 \boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{x}^*) \\ \rho^2 \hat{\sigma}_c^2 \boldsymbol{R}_c(\boldsymbol{X}_e, \boldsymbol{x}^*) + \hat{\sigma}_d^2 \boldsymbol{R}_d(\boldsymbol{X}_e, \boldsymbol{x}^*) \end{bmatrix} \tag{29}$$

The mean square variance estimate of the predicted value can be expressed as:

$$\hat{s}^2(\boldsymbol{x}^*) = \rho^2 \hat{\sigma}_c^2 + \hat{\sigma}_d^2 - \boldsymbol{c}^T \boldsymbol{C}^{-1} \boldsymbol{c} \tag{30}$$

4. Analysis and Evaluation of Model Results

4.1. Discussion of Results

4.1.1. Effect of Groove Depth

Figure 8 presents the three-dimensional fitted surfaces of opening force as functions of groove depth and rotational speed for the four data fusion methods, namely UWF, TrAdaBoost, MFDNN, and CoKriging. The overall variation trends in the four figures are generally consistent. In all cases, the opening force increases with increasing groove depth, h_g, and rotational speed, RPM, indicating that both variables have a significant positive effect on the opening force. The minimum values of all models appear in the region of low groove depth and low rotational speed, namely near $h_g \approx 5$ μm and RPM $\approx 10 \times 10^3$ r/min, where the opening force is about 150 N. The maximum values are all located in the region of high groove depth and high rotational speed, namely near $h_g \approx 8$ μm and RPM $\approx 40 \times 10^3$ r/min. Among them, the peak values of UWF and MFDNN are about 340–350 N, that of CoKriging is about 310–320 N, and that of TrAdaBoost is about 300 N. Considering the overall trends of the four figures, when the rotational speed increases from 10×10^3 r/min to 40×10^3 r/min, the opening force usually increases by about 120–180 N. When the groove depth increases from 5 μm to 8 μm, the opening force increases by approximately 40–80 N.

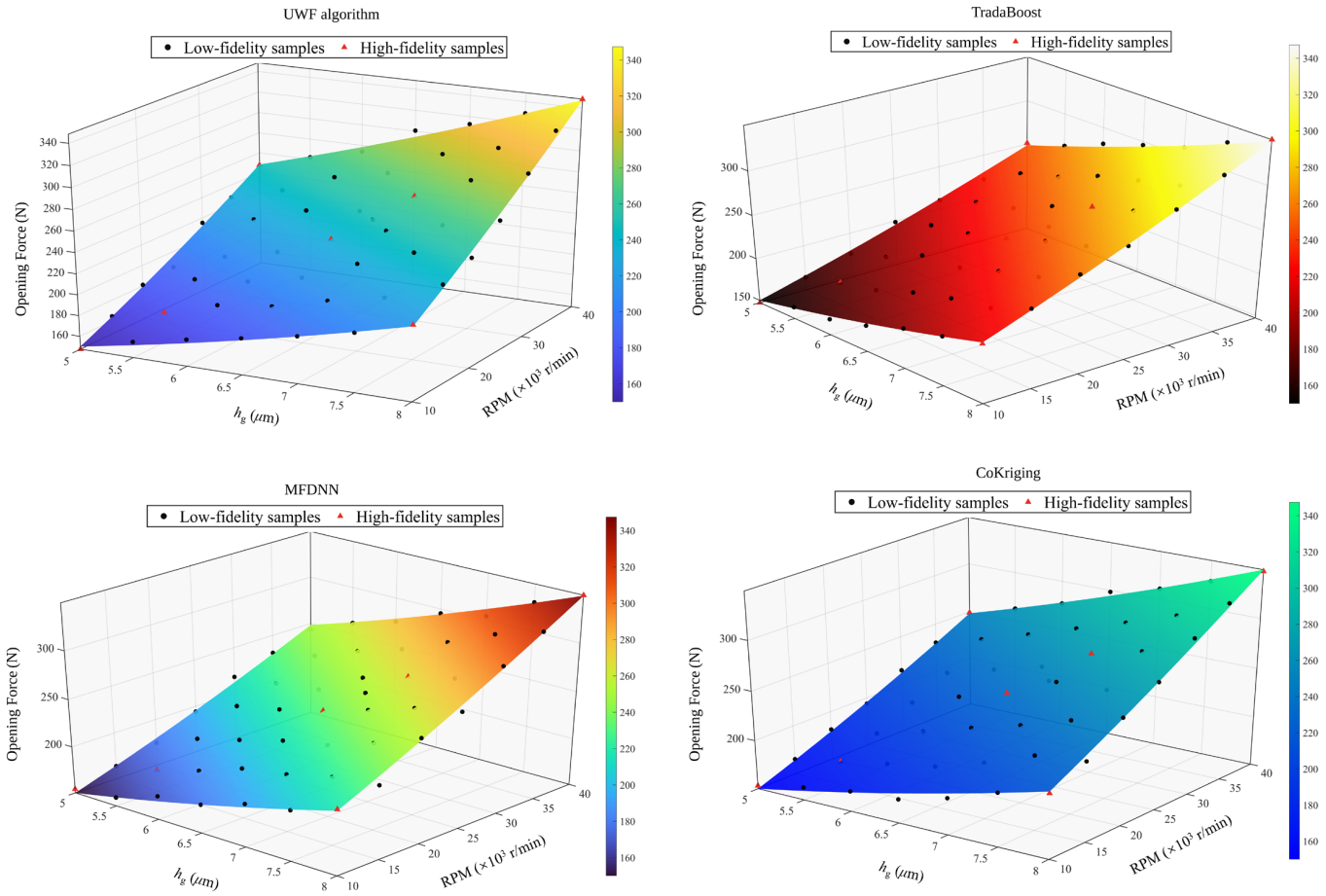


Figure 8. Data-fusion-algorithm-fitted response surface of opening force as a function of groove depth and rotational speed.

The fitted surfaces obtained by the four methods all rise continuously from the region of low groove depth and low rotational speed to the region of high groove depth and high rotational speed. The magnitude of variation along the rotational-speed direction is generally greater than that along the groove-depth direction, indicating that, within the investigated parameter range, the opening force is more sensitive to rotational speed. In particular, in the region of high rotational speed and large groove depth, the surface slope increases significantly, suggesting that the opening force grows most rapidly in this region, which is also where the fitted surfaces of the four models vary most sharply. In the medium- and low-parameter regions, the surfaces are relatively gentle, indicating that under low rotational-speed conditions, the response of opening force to parameter variations is comparatively limited.

From the characteristics of each individual figure, all four methods can clearly reflect the overall trend that the opening force increases monotonically with increasing rotational speed and groove depth. The fitted surface obtained by UWF is relatively accurate in capturing local variations. The surface continuity of TrAdaBoost is slightly better than that of UWF. MFDNN exhibits good smoothness throughout the entire parameter space and provides a relatively natural description of variations in the intermediate region. Similar to MFDNN, CoKriging also maintains good fitting characteristics.

Figure 9 presents the three-dimensional fitted surfaces of leakage rate as functions of groove depth and rotational speed for the four data fusion methods, namely UWF, TrAdaBoost, MFDNN, and CoKriging. Overall, the variation trends obtained by the four methods are generally consistent, with leakage rate increasing as groove depth h_g and

rotational speed RPM increase. In all four figures, the minimum leakage rate appears in the region of low groove depth and low rotational speed, approaching 0 mL/s, whereas the maximum values all occur in the region of high groove depth and high rotational speed, where the leakage rate is about 4.0–4.2 mL/s. From the overall trends of the four figures, when the rotational speed increases from 10×10^3 r/min to 40×10^3 r/min, the leakage rate usually increases by about 2.5–4.0 mL/s. When the groove depth increases from 5 μm to 8 μm, the leakage rate increases by approximately 0.5–1.2 mL/s. By comparison, the increase in rotational speed leads to a larger rise in leakage rate, indicating that, within the investigated range, rotational speed has a more pronounced effect on leakage rate. A closer examination shows that, in the high-rotational-speed region, the fitted surfaces rise much more rapidly, especially when the rotational speed increases from 30×10^3 r/min to 40×10^3 r/min, where the increase in leakage rate is the most significant. In contrast, in the low-rotational-speed region, the fitted surfaces are relatively gentle, and the differences in leakage rate under different groove depths are comparatively small.

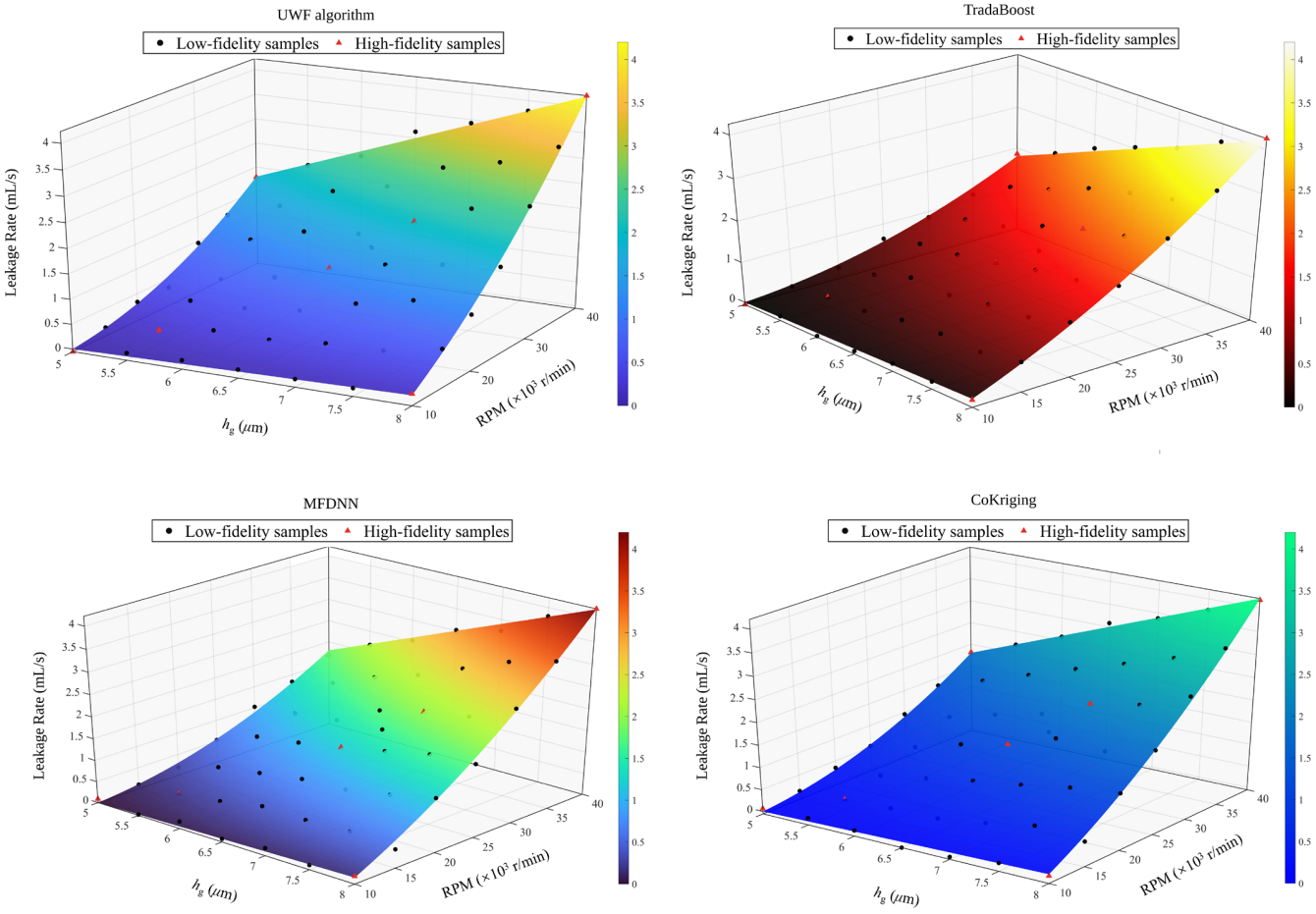


Figure 9. Data-fusion-algorithm-fitted response surface of leakage rate as a function of groove depth and rotational speed.

From the shapes of the fitted surfaces obtained by each method, the UWF surface is generally smooth and can clearly reflect the continuously increasing trend of leakage rate with increasing groove depth and rotational speed. The CoKriging surface also maintains a monotonically increasing trend, but its fitting performance is slightly inferior to that of UWF when the groove depth is less than 6.5 μm. Both the TrAdaBoost and MFDNN surfaces exhibit good overall continuity and better capture the variation characteristics reflected by the distribution of the sample points.

Overall, the four methods all show good-fitting characteristics with respect to groove depth, and no obvious superiority or inferiority can be identified among them. From the perspective of gas dynamic theory, the above observations can be explained as follows: increasing the groove depth enhances the gas dynamic pressure effect within the sealing face, thereby simultaneously increasing the opening force and the leakage rate through the sealing-face clearance.

4.1.2. Effect of Spiral Angle

Figure 10 presents the three-dimensional fitted surfaces of opening force as functions of spiral angle and rotational speed for the four methods, namely UWF, TrAdaBoost, MFDNN, and CoKriging. Overall, the surface shapes are relatively complex, but the variation trends exhibited by the four methods are generally consistent. In all cases, the opening force first increases and then decreases with increasing spiral angle, and the maximum value appears at a spiral angle of about 10°. When the rotational speed remains constant, increasing the spiral angle from 12° to 20° leads to a sharp decrease in the opening force of about 40%. Within the investigated parameter range, the maximum opening force is about 240–250 N, whereas the minimum value is about 110–120 N.

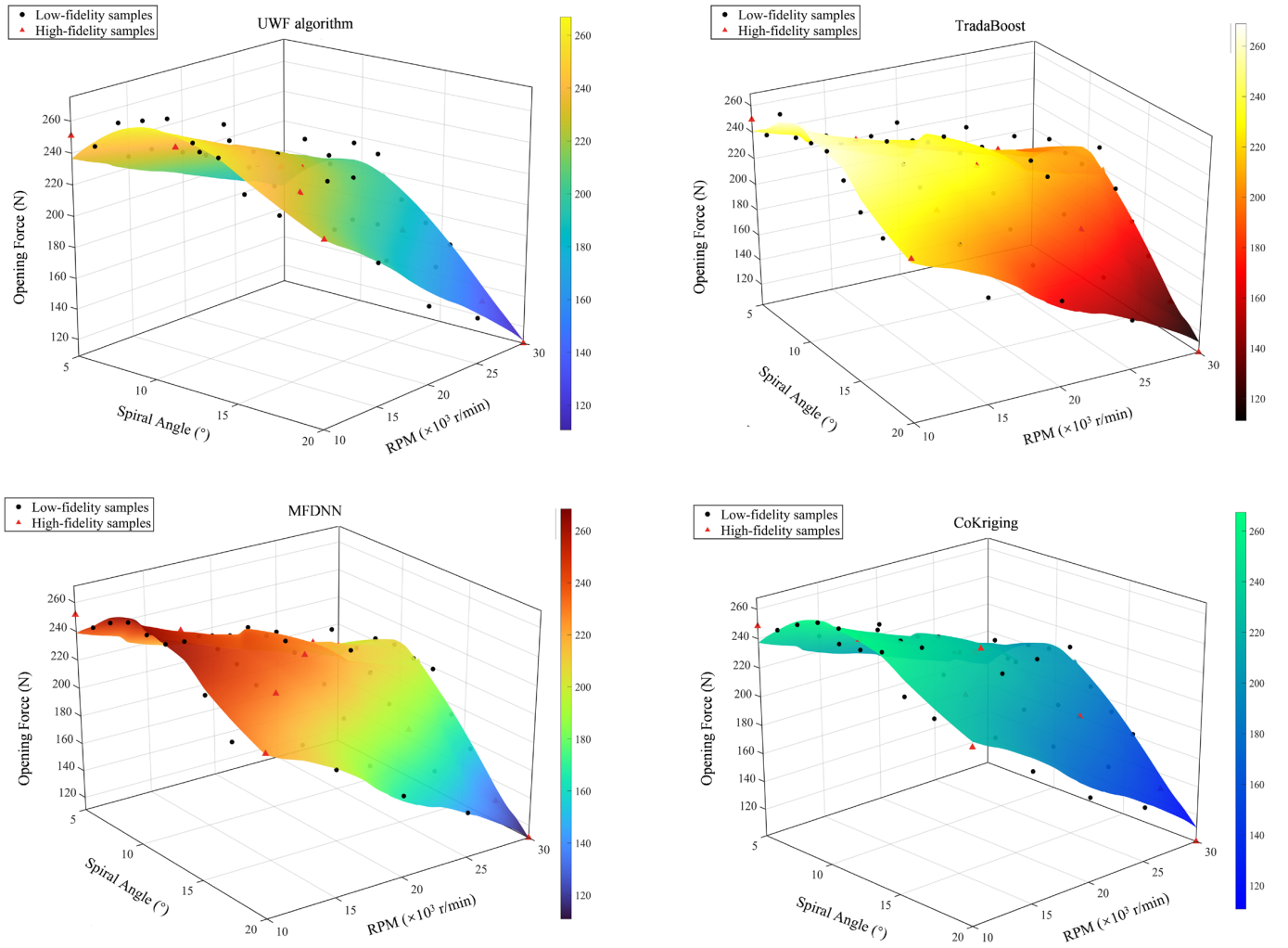


Figure 10. Data-fusion-algorithm-fitted response surface of opening force as a function of spiral angle and rotational speed.

From the shapes of the fitted surfaces obtained by the different methods, MFDNN provides the most adequate description of the overall decreasing trend, and the corre-

spondence between the fitted surface and the sample-point distribution is relatively good. TrAdaBoost is slightly inferior to MFDNN, but its fitted surface is still relatively smooth and shows good continuity. The CoKriging surface can also reflect the overall variation trend well, although its local details are slightly weaker than those of the former two methods in some regions. The UWF surface is relatively gentle as a whole. Although it can capture the overall decreasing trend of opening force with changes in the two parameters, its description of local features is relatively limited.

Figure 11 presents the three-dimensional fitted surfaces of leakage rate as functions of spiral angle and rotational speed for the four methods, namely UWF, TrAdaBoost, MFDNN, and CoKriging. Overall, the variation trends predicted by the four methods are generally consistent. In all cases, the leakage rate first decreases and then increases with increasing spiral angle, while it increases approximately linearly with rotational speed. The maximum leakage rate appears near a spiral angle of about 20° and RPM $\approx 30 \times 10^3$ r/min, with a value of about 1.8–2.0 mL/s. The minimum value occurs near a spiral angle of about 12° and RPM $\approx 10 \times 10^3$ r/min, approaching 0.2 mL/s. When the spiral angle remains nearly constant, increasing the rotational speed from 10×10^3 r/min to 30×10^3 r/min leads to an increase in leakage rate of about 30–80%. When the rotational speed remains nearly constant, increasing the spiral angle from 12° to 20° generally causes the leakage rate to increase by more than 200%. These results indicate that, compared with rotational speed, spiral angle has a more significant effect on leakage rate.

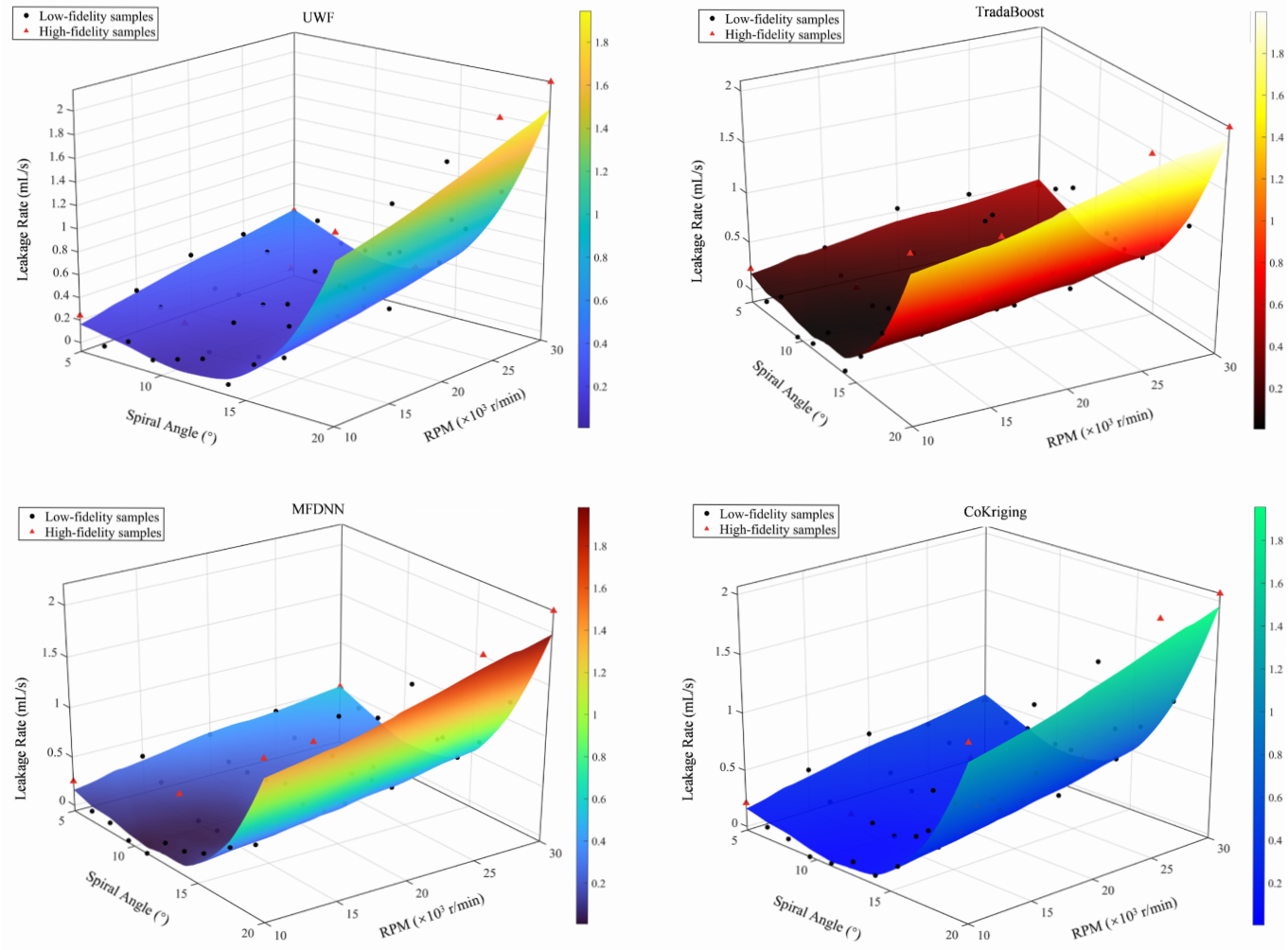


Figure 11. Data-fusion-algorithm-fitted response surface of leakage rate as a function of spiral angle and rotational speed.

From the shapes of the fitted surfaces obtained by the different methods, TrAdaBoost provides the clearest representation of the overall increasing trend, and the sample points do not deviate substantially from the fitted surface. The fitted surface produced by MFDNN is also relatively satisfactory. The performance of CoKriging is slightly weaker than that of the former two methods in the region of large spiral angles. The UWF surface captures the correct overall variation trend, but it is relatively simple and less capable of describing local variations than the other three methods.

Overall, for sealing performance, when the spiral angle is lower than 8° or higher than 20°, both opening force and leakage rate deteriorate simultaneously. Therefore, in engineering practice, a spiral angle in the range of 10° to 14° is preferable, as it can balance gas-film load-carrying capacity and leakage rate. From a mechanistic point of view, this can be explained as follows. An excessively large spiral angle allows more gas to enter the spiral grooves directly, which weakens the circumferential pumping and guiding effects of the grooves on the gas. As a result, the hydrodynamic effect of the gas film is reduced, and the opening force decreases. At the same time, the effective flow-restricting capability of the sealing face becomes poorer, leading to an increase in leakage rate. In the fitting evaluation under varying spiral angle and rotational speed, the TrAdaBoost method shows the best performance, while the MFDNN and CoKriging methods are generally comparable and belong to the second tier, and the UWF method performs slightly worse than the other three.

4.1.3. Effect of Number of Grooves

Figure 12 presents the three-dimensional fitted surfaces of opening force as functions of groove number and rotational speed for the four methods, namely UWF, TrAdaBoost, MFDNN, and CoKriging. Overall, the variation trends obtained by the four methods are generally consistent. When the groove number is less than 15, the opening force increases with increasing groove number and rotational speed. Once the groove number reaches 15, further increases in groove number lead to almost no change in opening force. Within the investigated parameter range, the maximum opening force is about 240–250 N, whereas the minimum value is about 70–80 N. When the groove number remains nearly constant, increasing the rotational speed from 10×10^3 r/min to 30×10^3 r/min causes the opening force to increase by more than 80%. When the rotational speed remains constant, increasing the groove number from eight to 15 causes the opening force to increase by more than 100%.

From the shapes of the fitted surfaces obtained by the different methods, the fitted surface of UWF is relatively gentle, with many local details missing overall, and it shows relatively large deviations from the sample points. TrAdaBoost provides the clearest representation of the variation process characterized by a rapid rise followed by a gradual stabilization. MFDNN misses some local details in the region of small groove number and low rotational speed, but shows good fitting performance in the other regions. CoKriging is less distinct than TrAdaBoost in distinguishing the plateau region from the transition region, but performs better than MFDNN.

Figure 13 presents the three-dimensional fitted surfaces of leakage rate as functions of groove number and rotational speed for the four methods, namely UWF, TrAdaBoost, MFDNN, and CoKriging. From an overall perspective, the variation trends obtained by the four methods are generally consistent. The leakage rate first increases, then decreases, and finally tends to stabilize with increasing groove number. The maximum value appears near a groove number of nine and a rotational speed of 30×10^3 r/min, with a value of about 2.3–2.6 mL/s. When the groove number remains nearly constant, increasing the rotational speed from 10×10^3 r/min to 30×10^3 r/min causes the leakage rate to increase

by about 1.5–2.3 mL/s. When the rotational speed remains constant, increasing the groove number from eight to about 15 leads to a noticeable increase in leakage rate, whereas further increases in groove number result in a more gradual variation. These results indicate that, compared with rotational speed, groove number affects leakage rate more noticeably only within a limited range.

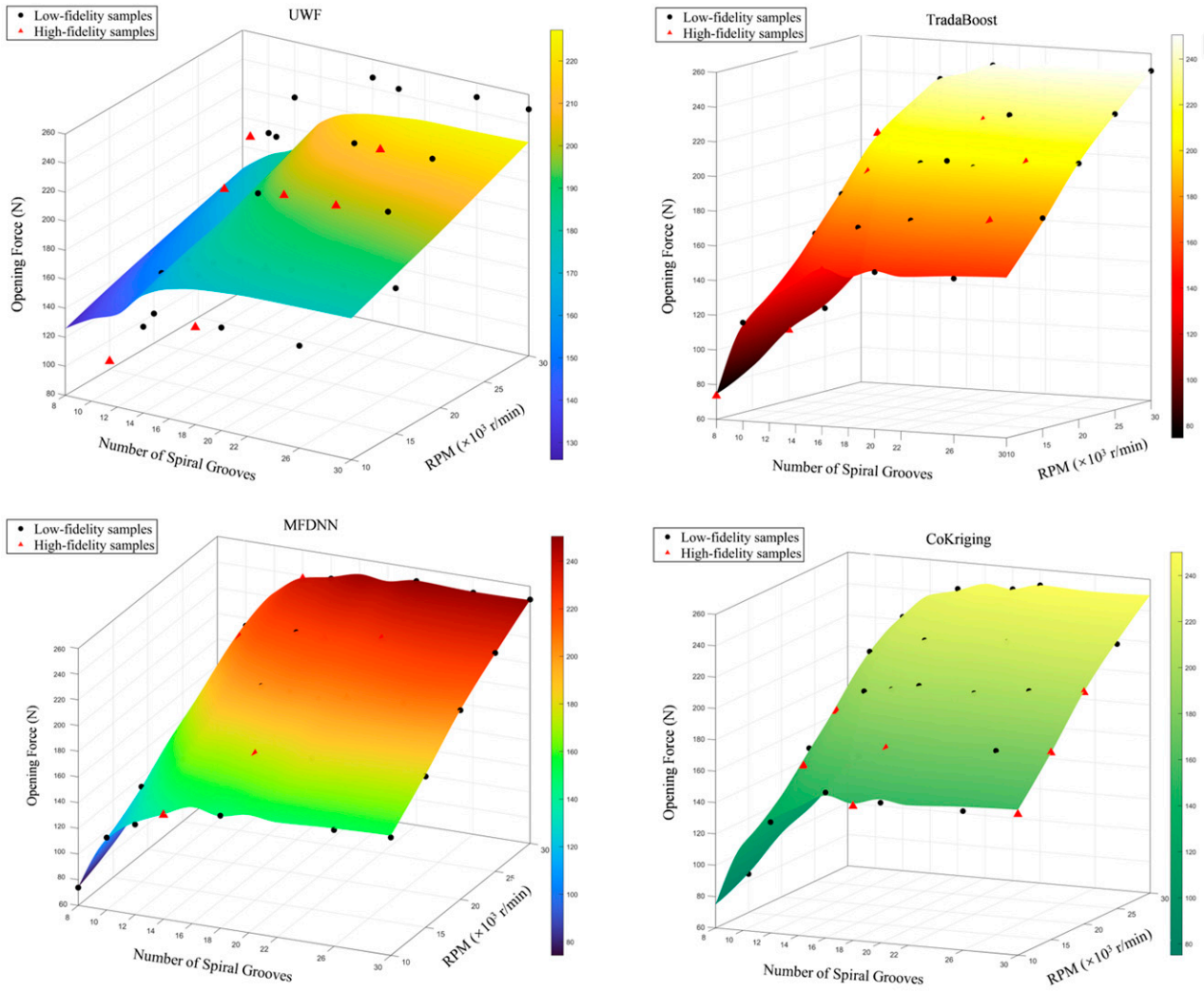


Figure 12. Data-fusion-algorithm-fitted response surface of opening force as a function of groove number and rotational speed.

From the perspective of fluid dynamic mechanisms, increasing the number of grooves changes not merely a single geometric parameter, but the entire circumferential flow structure of the gas film on the sealing face. As the groove number increases, the number of hydrodynamic pumping units per unit circumference rises, and the pressure distribution evolves from a relatively discrete pattern to a more continuous and uniform one, leading to an overall increase in the opening force. At the same time, however, the circumferential scale of each individual groove decreases, which increases the flow resistance within the groove and intensifies the pressure interference between adjacent grooves. As a result, the incremental contribution of newly added grooves to the load-carrying capacity gradually diminishes. Therefore, the opening force does not increase linearly with groove number, but instead exhibits a nonlinear trend characterized by a rapid increase at lower groove numbers followed by a progressively slower growth at higher groove numbers. The variation in leakage rate can be explained by the competition between two underlying mechanisms. When the groove number is small, the circumferential hydrodynamic effect

on the sealing face is still insufficient. As the groove number increases, the number of flow passages grows, the circumferential connectivity and shear-driven pumping effect are enhanced, and the fluid's hydrodynamic pressure-generating capability is strengthened, resulting in an increase in leakage rate. However, when the groove number becomes excessively large, the width of each individual groove decreases, which not only increases the flow resistance but also reduces the effective flow area, thereby suppressing fluid leakage. Under the combined influence of these competing effects, the leakage rate does not continue to increase at the same rate; instead, it exhibits a pronounced non-monotonic behavior, reaching a peak and then gradually weakening until it approaches a stable state.

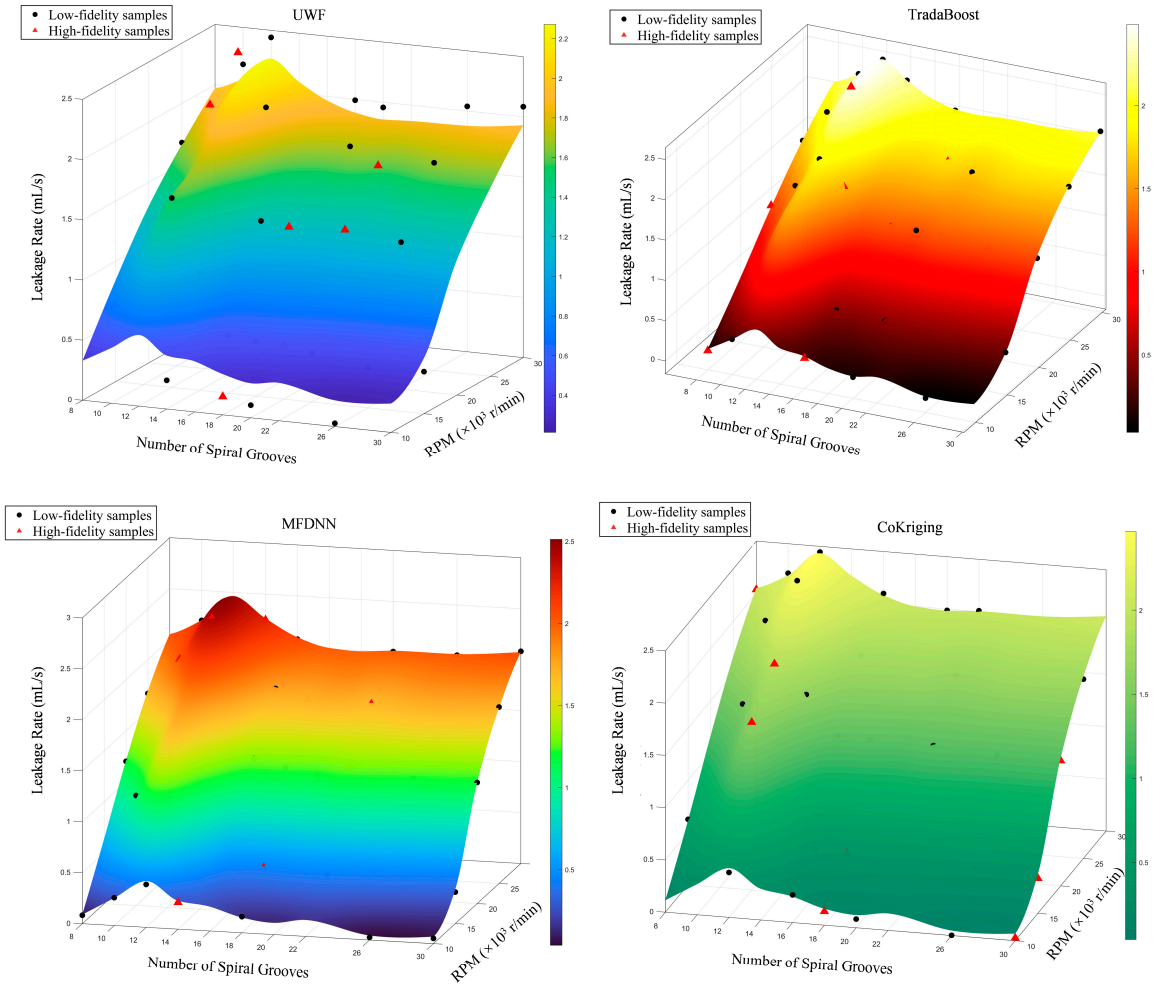


Figure 13. Data-fusion-algorithm-fitted response surface of leakage rate as a function of groove number and rotational speed.

From the characteristics of the individual fitted surfaces, UWF still shows large deviations from the sample points in the high-rotational-speed region, and its fitted values in the low-rotational-speed region are higher than those of the other three methods, indicating relatively poor overall performance. TrAdaBoost provides the most complete and clearest fitting of the leakage-rate variation process and describes the locations of variation more explicitly. The fitting capability of MFDNN at the variation locations is slightly weaker than that of TrAdaBoost. The CoKriging method shows some deviation from some sample points in the region of small groove number and high rotational speed, while the fitting performance in the other regions is good.

Overall, in terms of sealing performance, selecting a groove number in the range of 12–16 can provide a balance between leakage rate and opening force. Increasing the

number of grooves further will raise the manufacturing cost of the seal, while the improvement in sealing performance is limited. From the perspective of data fusion methods, the UWF method performs worse than the other three methods in fitting both opening force and leakage rate and should therefore be excluded in studies focusing on groove number. Among the remaining three methods, TrAdaBoost shows the best performance, whereas MFDNN and CoKriging are relatively close to each other, and both can be regarded as viable alternatives in engineering practice when the accuracy requirement is not particularly high.

4.2. Evaluation of Model Performance

4.2.1. Definition of Evaluation Metrics

To comprehensively evaluate the fitting performance of the four data fusion models, this study selected the mean absolute error (MAE), mean square error (MSE), and coefficient of determination (R^2) as model performance evaluation metrics, and used these three indicators to analyze the effectiveness of the data fusion methods [39,40].

If the total number of samples in the validation set is n, y_i is the true value of the i-th sample, and $\hat{y}_i$ is the predicted value of the i-th sample, then:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{31}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{32}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right)^2} \tag{33}$$

Here, MAE is used to measure the mean absolute deviation between the predicted values and the true values and can directly reflect the average prediction error of the model. MSE is used to measure the mean of the squared prediction errors. It is more sensitive to large errors and can effectively characterize the model's ability to control abnormal deviations. R^2 is used to evaluate the degree to which the model fits the variation pattern of the true data and reflects the overall goodness of fit of the model. For MAE and MSE, smaller values indicate higher prediction accuracy. For R^2, a value closer to one indicates better fitting performance.

MAE is characterized by its intuitive physical meaning and straightforward interpretability; MSE is more sensitive to relatively large local deviations in the model predictions; and R^2 reflects the overall consistency between the predicted results and the true values. These three metrics correspond respectively to the average error level, sensitivity to larger errors, and overall fitting capability. When used in combination, they enable a more comprehensive evaluation of the predictive performance of different methods from multiple perspectives.

4.2.2. Comparative Analysis of Error Results

Tables 4 and 5 quantitatively present the evaluation results of the four data fusion methods for fitting the effects of groove depth and rotational speed on sealing performance. From the numerical results, no obvious difference can be identified among the four methods, and all three metrics, namely MAE, MSE, and R^2, show good performance. This indicates that the prediction models established for the study of groove-depth-related problems have relatively high reliability. For all four methods, the R^2 values are greater than 0.95,

while most MAE and MSE values are also within 3%, indicating that all results fall within a satisfactory range. The conclusions drawn from the quantitative evaluation are also consistent with the preceding analysis, suggesting that all four methods are applicable to the study of groove-depth-related problems. From a mechanistic point of view, this is because groove depth, rotational speed, leakage rate, and opening force exhibit a relatively strong linear relationship, and the complex nonlinear relationships that need to be handled by the models are relatively limited. Therefore, all four methods are able to maintain good predictive capability. From a practical perspective, when the performance of the models is similar, the uncertainty-weighted data fusion method is more recommended. Compared with the other three methods, it has lower implementation difficulty and lower cost, while still achieving favorable predictive performance.

Table 4. Comparison of error metrics for opening force under varying rotational speed and groove depth.

	MAE	MSE	R^2
UWF	0.0292	0.0312	0.9552
CoKriging	0.0278	0.0298	0.9641
MFDNN	0.0199	0.0275	0.9685
TrAdaBoost	0.0164	0.0196	0.9756

Table 5. Comparison of error metrics for leakage rate under varying rotational speed and groove depth.

	MAE	MSE	R^2
UWF	0.0308	0.0323	0.9514
CoKriging	0.0279	0.0316	0.9572
MFDNN	0.0213	0.0267	0.9611
TrAdaBoost	0.0240	0.0284	0.9589

Tables 6 and 7 quantitatively present the fitting evaluation results of the four data fusion methods for the rotational-speed–spiral-angle problem. Overall, in the study of spiral-angle-related problems, relatively clear performance differences can already be observed among the four methods. In terms of absolute values, except for the UWF method, most R^2 values are greater than 0.95. For the other three methods, most MAE and MSE values are also below 5%, indicating that they have good fitting performance. A closer examination shows that, for opening-force prediction, MFDNN performs the best, while TrAdaBoost is only slightly inferior to MFDNN. Both methods exhibit high prediction accuracy and fitting capability. CoKriging ranks next, whereas the UWF method performs the worst overall. For leakage-rate prediction, TrAdaBoost performs the best, and the gap between TrAdaBoost and MFDNN is very small. CoKriging ranks third, while the UWF method remains the weakest.

Table 6. Comparison of error metrics for opening force under varying rotational speed and spiral angle.

	MAE	MSE	R^2
UWF	0.0846	0.0518	0.9117
CoKriging	0.0613	0.0226	0.9528
MFDNN	0.0328	0.0124	0.9796
TrAdaBoost	0.0345	0.0135	0.9711

Table 7. Comparison of error metrics for leakage rate under varying rotational speed and spiral angle.

	MAE	MSE	R^2
UWF	0.0974	0.0638	0.8952
CoKriging	0.0721	0.0314	0.9378
MFDNN	0.0496	0.0187	0.9514
TrAdaBoost	0.0478	0.0179	0.9533

Overall, in the study of the effects of rotational speed and spiral angle on sealing performance, TrAdaBoost and MFDNN exhibit clearly better overall performance, indicating that these two methods have stronger feature extraction and prediction capabilities when dealing with this problem. CoKriging also shows good performance but remains overall inferior to the former two methods. The fitting capability of the UWF method is relatively limited. This conclusion is also reflected in the figures presented in the previous section. These results further indicate that data fusion methods with stronger learning capability have greater advantages in describing complex nonlinear relationships.

Tables 8 and 9 quantitatively present the fitting evaluation results of the four data fusion methods for the rotational-speed–groove-number problem. Overall, similar to the case of spiral angle, relatively clear differences can be observed in the performance of the four methods. In terms of the specific values of the three evaluation metrics, the UWF method performs slightly worse, whereas for the other methods, the R^2 values are all above 0.94, and the MAE and MSE values are all within 10%. Specifically, for opening-force prediction, TrAdaBoost performs the best, MFDNN is relatively close to it, CoKriging performs worse than MFDNN, and the UWF method performs the worst. For leakage-rate prediction, all four methods achieve relatively good results overall, but the same trend can still be observed, with TrAdaBoost performing the best, followed by MFDNN, then CoKriging, and finally UWF. Overall, in the study of the effects of rotational speed and groove number on sealing performance, TrAdaBoost and MFDNN exhibit higher prediction accuracy, CoKriging remains at an intermediate level, whereas the overall fitting capability of the UWF method is relatively weak.

Table 8. Comparison of error metrics for opening force under varying rotational speed and groove number.

	MAE	MSE	R^2
UWF	0.1186	0.0842	0.8157
CoKriging	0.0675	0.0284	0.9412
MFDNN	0.0628	0.0251	0.9463
TrAdaBoost	0.0584	0.0219	0.9511

Table 9. Comparison of error metrics for leakage rate under varying rotational speed and groove number.

	MAE	MSE	R^2
UWF	0.0837	0.0486	0.9156
CoKriging	0.0602	0.0284	0.9507
MFDNN	0.0461	0.0235	0.9683
TrAdaBoost	0.0418	0.0187	0.9744

The main reason for this phenomenon is that, similar to spiral angle, groove number affects sealing performance in a more nonlinear manner. Traditional fusion models lack sufficient fitting capability when dealing with complex nonlinear features over a wide parameter range and therefore perform worse than models with transfer learning capability. However, transfer-learning-based models are more difficult to implement and involve

higher construction and deployment costs. Therefore, in specific engineering practice, it is necessary to comprehensively consider sample characteristics, sample size, and model deployment difficulty when selecting an appropriate model, rather than indiscriminately choosing a high-performance model.

5. Conclusions

To address the performance prediction problem of spiral-groove dry gas seals, this study constructed a multi-source data system composed of numerical simulation data and experimental data. Within a research framework based on a unified sample set and unified evaluation metrics, four representative data fusion methods, namely the uncertainty-weighted fusion method, TrAdaBoost, MFDNN, and CoKriging, were systematically compared. The results show that:

1. Multi-source data fusion methods can effectively combine the respective advantages of low-cost, large-sample simulation data and high-reliability, small-sample experimental data, thereby improving the prediction accuracy of opening force and leakage rate. This indicates that the application of multi-source data fusion methods to dry gas seal research is both feasible and effective. Compared with traditional modeling approaches that rely on a single data source, multi-source data fusion methods exhibit clear advantages in terms of sample utilization efficiency and prediction accuracy.

2. In terms of the influence of structural parameters and operating conditions on sealing performance, groove depth, spiral angle, and groove number all significantly affect opening force and leakage rate, but their modes of action are not the same. For groove-depth-related problems, both opening force and leakage rate increase with increasing groove depth and rotational speed, and the overall influence of rotational speed is greater than that of groove depth. For spiral-angle-related problems, the opening force first increases and then decreases with increasing spiral angle, and the optimal region is approximately located near 10°. In contrast, the leakage rate first decreases and then increases with increasing spiral angle, indicating that there is a relatively distinct suitable range for spiral angle. For groove-number-related problems, the opening force increases rapidly when the groove number is small, but after reaching a certain value, the gain becomes significantly weaker with further increases in groove number. The leakage rate, by contrast, first increases, then decreases, and gradually tends to stabilize. In structural parameter design, the spiral angle should preferably be controlled within the range of 10–14°, and the groove number should preferably be controlled within the range of 12–16, so as to balance opening force and leakage rate. Blindly increasing the groove number or deviating from a reasonable spiral-angle range not only fails to further improve sealing performance but may also increase manufacturing cost or lead to deterioration of overall performance. Overall, the influences of different structural parameters on sealing performance show marked differences. Among them, the effects of spiral angle and groove number on sealing performance exhibit more pronounced nonlinear characteristics, whereas the influence of groove depth on sealing performance is relatively more direct.

3. From the comparative results of the four data fusion methods, it can be seen that different methods show clear differences in adaptability to different problems. For the groove-depth–rotational-speed problem, all four methods yield good results. The R^2 values for opening-force prediction are all greater than 0.95, and those for leakage-rate prediction also remain around 0.95, with only small differences among the models. This indicates that the variation pattern of this problem is relatively clear and that the modeling difficulty is relatively low. By contrast, for the spiral-angle and groove-number problems, the differences among the methods become much more pronounced. Among them, TrAdaBoost performs best overall, MFDNN is only slightly inferior to TrAdaBoost,

CoKriging ranks third, and UWF performs the worst. Considering the overall results of this study, it can be concluded that, for dry gas seal problems with more pronounced complex nonlinear characteristics, methods with stronger feature learning and transfer capability are more likely to achieve higher prediction accuracy. In contrast, for problems in which the variation patterns are relatively stable and the degree of nonlinearity is weaker, methods with simpler structures can also obtain good results.

4. From an engineering application perspective, the selection of a data fusion method should not be discussed in isolation from the specific problem but should instead be made through a comprehensive consideration of the parameter variation patterns, sample size, target accuracy, and model deployment cost. The present study demonstrates that data fusion methods have clear engineering value for dry gas seal performance prediction. When high-fidelity samples are limited, these methods can effectively incorporate the trend information contained in low-fidelity data to achieve relatively accurate predictions of opening force and leakage rate, thereby providing effective support for the rapid selection of dry gas seal structural parameters. In addition, for key parameters such as groove depth, groove number, and rotational speed, data fusion models can efficiently reveal performance variation patterns and parameter sensitivity characteristics, offering useful guidance for scheme screening, parameter matching, and the determination of reasonable parameter ranges during the preliminary design stage. Specifically, when the study mainly involves groove depth and the performance variation is relatively smooth, the uncertainty-weighted fusion method may be preferred because of its simple implementation and low computational cost. In contrast, when the problem involves parameters with stronger nonlinear effects, such as spiral angle or groove number, methods with stronger learning capability, such as TrAdaBoost or MFDNN, should be given priority. Compared with approaches that rely solely on extensive experiments or high-precision numerical simulations, data fusion methods can reduce design cost, decrease repetitive analysis effort, and improve the efficiency of dry gas seal structural optimization and preliminary engineering design.

Author Contributions: J.Y. (Jiashu Yu): Methodology, data analysis, software and writing. X.D.: Experimental equipment, review and editing. J.Y. (Jianping Yu): Review and revision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Gansu Provincial Department of Education Innovation Star Fund (2025CXZX-526) and the National Key R & D Program of China (No.2020YFB2010001).

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors are grateful to the reviewers for their careful and detailed comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gruenewald, M.; Wagner, W. Recent progress in compressor sealing. *Seal. Technol.* **2005**, *2005*, 6–8. [CrossRef]
2. Wang, Y.; Yang, H.; Wang, J.; Liu, Y.; Wang, H.; Feng, X. Theoretical analyses and field applications of gas-film lubricated mechanical face seals with herringbone spiral grooves. *Tribol. Trans.* **2009**, *52*, 800–806. [CrossRef]
3. Ruan, B. Numerical modeling of dynamic sealing behaviors of spiral groove gas face seals. *J. Tribol.* **2002**, *124*, 186–195. [CrossRef]
4. Zirkelback, N. Parametric study of spiral groove gas face seals. *Tribol. Trans.* **2000**, *43*, 337–343. [CrossRef]
5. Teng, L.; Jiang, J.; Peng, X.; Li, J.; Zheng, S. Study on Angular Free Vibration Stability and Parameters Influence of Dry Gas Seal Based on the Characteristic Equation. *Shock Vib.* **2022**, *2022*, 7378999. [CrossRef]
6. Xu, H.; Yue, Y.; Song, P. Analysis on the dynamic characteristics of spiral groove dry gas seal based on the gas film adaptive adjustment model. *Ind. Lubr. Tribol.* **2023**, *75*, 406–414. [CrossRef]
7. Yu, J.; Ding, X.; Chen, J.; Yu, J. Investigation of spiral-groove dry gas seal performance using an experimental data-driven Kriging surrogate model. *Lubricants* **2026**, *14*, 119. [CrossRef]

8. Wang, X.; Shi, L.; Huang, W.; Wang, X. A multi-objective optimization approach on the spiral grooves for gas mechanical seals. *J. Tribol.* **2018**, *140*, 041701. [CrossRef]
9. Zhang, L.; Ding, X.; Wang, S.; Zhang, S.; Cui, B. Lubrication characteristics of dry-gas seals with spiral grooves. *Machines* **2024**, *12*, 610. [CrossRef]
10. Beskok, A.; Karniadakis, G.E.; Trimmer, W. Rarefaction and compressibility effects in gas microflows. *J. Fluids Eng.* **1996**, *118*, 448–456. [CrossRef]
11. Lu, J.; Ding, X.; Zhang, W. The influence of spiral angle on lift-off speed of face gas seal under start-up. *SN Appl. Sci.* **2019**, *1*, 1068. [CrossRef]
12. Xu, H.; Song, P.; Mao, W.; Deng, Q. The performance of spiral groove dry gas seal under choked flow condition considering the real gas effect. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **2020**, *234*, 554–566. [CrossRef]
13. Blasiak, S.; Zahorulko, A.V. A parametric and dynamic analysis of non-contacting gas face seals with modified surfaces. *Tribol. Int.* **2016**, *94*, 126–137. [CrossRef]
14. Chen, Y.; Peng, X.; Jiang, J.; Meng, X.; Li, J. Experimental and theoretical studies of the dynamic behavior of a spiral-groove dry gas seal at high-speeds. *Tribol. Int.* **2018**, *125*, 17–26. [CrossRef]
15. Sneck, H.J.; McGrew, J.M. Analytical investigation of the spiral groove face seal. *J. Lubr. Technol.* **1973**, *95*, 499–510. [CrossRef]
16. Ding, X.; Lu, J. Theoretical analysis and experiment on gas film temperature in a spiral groove dry gas seal under high speed and pressure. *Int. J. Heat Mass Transf.* **2016**, *96*, 438–450. [CrossRef]
17. Shi, T.; Zhang, Y. Numerical Investigation of Gas Film Performance in Face Dry Gas Seals with Combined Micro-Textured Structures. *Lubricants* **2026**, *14*, 90. [CrossRef]
18. Gabriel, R. Fundamentals of spiral groove noncontacting face seals. *Lubr. Eng.* **1994**, *50*, 215–224.
19. Salant, R.F.; Homiller, S.J. Stiffness and leakage in spiral groove upstream pumping mechanical seals. *Tribol. Trans.* **1993**, *36*, 55–60. [CrossRef]
20. Faria, M.T.C. An efficient finite element procedure for analysis of high-speed spiral groove gas face seals. *J. Tribol.* **2001**, *123*, 205–210. [CrossRef]
21. Jiang, J.; Zhao, W.; Peng, X.; Li, J. A novel design for discrete surface texture on gas face seals based on a superposed groove model. *Tribol. Int.* **2020**, *147*, 106269. [CrossRef]
22. Chen, Y.; Zhuo, Z.; Li, Y.; Li, X.; Wang, B.; Jin, J.; Zhang, C. Prediction of spiral groove dry gas seal performance and intelligent optimization of structural parameters. *Tribol. Int.* **2024**, *193*, 109439. [CrossRef]
23. Fernández-Godino, M.G. Review of multi-fidelity models. *Adv. Comput. Sci. Eng.* **2023**, *1*, 351–400. [CrossRef]
24. Azarhoosh, Z.; Ilchi Ghazaan, M. A review of recent advances in surrogate models for uncertainty quantification of high-dimensional engineering applications. *Comput. Methods Appl. Mech. Eng.* **2025**, *433*, 117508. [CrossRef]
25. Jiang, J.; Peng, X.; Li, J.; Chen, Y. Leakage and stiffness characteristics of bionic cluster spiral groove dry gas seal. *Chin. J. Mech. Eng.* **2018**, *31*, 21. [CrossRef]
26. Jiang, J.; Peng, X.; Li, J.; Chen, Y. A comparative study on the performance of typical types of bionic groove dry gas seal based on bird wing. *J. Bionic Eng.* **2016**, *13*, 324–334. [CrossRef]
27. Liu, X.; Dong, Z.; Ji, H.; Yue, Z.; Kang, J. Uncertainty-based fusion method for structural modal parameter identification. *Sensors* **2025**, *25*, 4397. [CrossRef]
28. Kennedy, M.C.; O'Hagan, A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **2000**, *87*, 1–13. [CrossRef]
29. Peherstorfer, B.; Willcox, K.; Gunzburger, M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev.* **2018**, *60*, 550–591. [CrossRef]
30. Dai, W.; Yang, Q.; Xue, G.-R.; Yu, Y. Boosting for transfer learning. In *Proceedings of the 24th International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009*; Association for Computing Machinery: New York, NY, USA, 2009; pp. 193–200. [CrossRef]
31. Zhang, X.; Xie, F.; Ji, T.; Zhu, Z.; Zheng, Y. Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization. *Comput. Methods Appl. Mech. Eng.* **2021**, *373*, 113485. [CrossRef]
32. Goel, T.; Haftka, R.T.; Shyy, W.; Queipo, N.V. Ensemble of surrogates. *Struct. Multidiscip. Optim.* **2007**, *33*, 199–216. [CrossRef]
33. Bhosekar, A.; Ierapetritou, M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comput. Chem. Eng.* **2018**, *108*, 250–267. [CrossRef]
34. Jin, S.-S. Accelerating Gaussian Process surrogate modeling using Compositional Kernel Learning and multi-stage sampling framework. *Appl. Soft Comput.* **2021**, *104*, 106909. [CrossRef]
35. Liu, G.; Ding, X.; Xu, J.; Ding, J.; Zhang, Z. Study on the gas film lubrication characteristics of micro-textured ultra-high-speed dry gas seals. *Fluid Mach.* **2024**, *52*, 24–32. (In Chinese) [CrossRef]
36. Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **2018**, *85*, 1–16. [CrossRef]

37. Li, K.; Wei, B.; Tang, Q.; Liu, Y. A data-efficient building electricity load forecasting method based on maximum mean discrepancy and improved TrAdaBoost algorithm. *Energies* **2022**, *15*, 8780. [CrossRef]
38. Xu, C.; Han, Z.; Zhang, K.; Zeng, H.; Chen, G.; Zhou, Z. A novel multi-fidelity cokriging model assisted by multiple non-hierarchical low-fidelity datasets. *Struct. Multidiscip. Optim.* **2024**, *67*, 52. [CrossRef]
39. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
40. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef]

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