Journal of Engineering and Technological Sciences

Coolant Bleed Flow Modeling and Prediction for Aircraft Engines Based on TSO-RF Algorithm

Enlai Zhang, Jiaqiang Jiang*, & Zijuan Ren

School of Mechanical and Automotive Engineering, Xiamen University of Technology, No.600 Ligong Road, Jimei District, Xiamen, Fujian Province, 361024, China

*Corresponding author: jjq16816899@163.com

Abstract

Coolant bleed flow forecasting is critical for the engineering health of aircraft engines. Using the datasets from Commercial Modular Aero-Propulsion System Simulation (C-MAPSS), this paper aims to establish high-precision coolant bleed flow models by the random forest (RF) algorithm combined with tuna swarm optimization (TSO). 17 sensor variables that are moderately or highly correlated with the low-pressure turbine (LPT) and high-pressure turbine (HPT) are selected as inputs with LPT and HPT as outputs. After being trained and validated on the FD002 and FD004 datasets, the TSO-RF model significantly reduces mean squared error, mean absolute error and root mean squared error, and improves the determination coefficient R^2 compared to other RF models. It verifies the superiority of the TSO-RF model in predicting engine coolant bleed flow, providing reliable technical support for subsequent evaluation and management in aircraft engine bleed air system.

Keywords: *aircraft engines; C-MAPSS; coolant beleed flow; engine performance; TSO-RF.*

Introduction

The aircraft engine, as a core component of modern aero-systems, directly impacts flight safety, operational efficiency, and reliability (Lei et al., 2024). With the growing emphasis on sustainable development, engine coolant bleed flow has emerged as a critical concern in the aircraft industry (Yuhas, A. J. & Ray, R. J., 2012). In particular, the low-pressure turbine (LPT) and high-pressure turbine (HPT) play key roles in the generation process of coolant bleed flow. It is recognized that these coolant bleed flows may exert potential impacts on terrestrial and atmospheric environments (Liu et al., 2023). However, existing research has primarily utilized open-source data from Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) for predicting remaining useful life of aircraft engines (Selvara et al., 2025; Duan et al., 2024), while no studies have yet reported on coolant bleed flow. Since the coolant bleed flow parameters directly affect the engine thrust output and performance, it is of great significance to investigate the coolant bleed flow modeling for aircraft engines using C-MAPSS. This facilitates accurate prediction of coolant bleed flow distribution across the LPT and HPT, thereby providing a basis for the optimized design and health management of aircraft engines.

Currently, various methods have been employed to predict engine performance, such as physical modeling and machine learning (Cai et al., 2024). Physical modeling focuses on describing the effect of chemical reactions during engine combustion. This is achieved by integrating the laws of mass conservation, momentum conservation, and energy conservation with chemical kinetics principles to establish governing equations (Guan et al., 2024). Stephen et al. (2022) emphasized the integrated application of chemical reaction kinetics with computational fluid dynamics simulations. Their work centered on employing the mechanically driven numerical simulation to analyze the formation mechanisms. Obviously, the main challenges of this method are the huge amounts of computation, storage resources and time required.

The second type of method is data-driven, primarily relying on large amounts of data to establish predictive models. For example, Tian et al. (2024) employed random forest (RF) and extreme gradient boosting to predict the impact of hydrogen enrichment on diesel engines. Subsequently, Madziel et al. (2024) utilized RF and gradient boosting methods to develop emission models for hydrocarbons and NOx in vehicles equipped with start-stop technology. It is evident

ISSN: 2337-5779
J. Eng. Technol. Sci. Vol. 58, No. 4, 2026, 540-554
DOI: 10.5614/j.eng.technol.sci.2026.58.4.6

DOI : 10.5614/j.eng.technol.sci.2026.58.4.6

that RF has found extensive application in predicting diesel engine performance, but no reports exist regarding its use in forecasting coolant bleed flow for aircraft engines, and establishing a more accurate coolant bleed flow model for aircraft engines is an ongoing task. RF operates through a collection of multiple decision trees, as detailed by Waqas et al. (2023) in the review of rainfall forecasting based on artificial intelligence technologies. Since the performance of RF models is highly sensitive to parameters such as the number of trees and minimum leaf node count, it is an essential research direction to explore intelligent algorithms for optimizing RF model parameters. In recent reports over the past three years, Yang et al. (2023) optimized a RF model using the tuna swarm optimization (TSO) algorithm for application in dual feature extraction system of ship-radiated noise, achieving the identification rate as high as 97.5%. It is worth noting that TSO has the advantages of simple structure and lower computational complexity (Cristina, 2023; Zhu et al., 2023; Xie et al., 2021).

Therefore, based on existing publicly available data from C-MAPSS, this paper investigates the establishment of engine coolant bleed flow models by combining TSO and RF to enhance the prediction accuracy, aiming to provide a critical engineering technology for aircraft engine health management.

RF Algorithm

The core concept of RF is to integrate multiple decision trees into a machine learning model (Khan et al., 2025). Specifically, it integrates a fixed number of decision trees with diverse characteristics, thereby enhancing predictive accuracy and robustness. As the term "forest" suggests, RF consists of numerous decision trees constructed from the training data. The collection of these trees forms the random forest. When input data is provided, each decision tree processes the data and generates a classification result. A voting mechanism is then employed, where the most frequently occurring result is selected as the final classification (Lin et al., 2020). Figure 1 illustrates the structure of the RF model.

In a classification task involving N samples and M feature values, each sample is associated with a corresponding label Y_i. Subsequently, a RF model is constructed by building N decision trees, where each tree is trained on a randomly selected subset of the samples. During the construction of each tree, only a subset of features is considered for node splitting. Each feature is termed an independent variable, collectively represented as $\{x_1, x_2, ..., x_M\}$. The core principle lies in selecting the optimal splitting feature. The construction process of the RF model primarily involves two steps, which are broken down as follows:

(1) Build random forest trees: the prepared dataset is divided into two parts, i.e. the training set and the prediction set. Bootstrap sampling is applied to the training set, which involves randomly selecting samples from existing data with replacement. This ensures that each decision tree is trained on a different subset of data, thereby increasing the model's diversity.

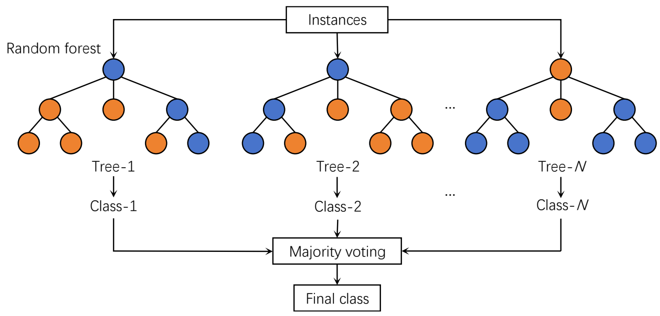


Figure 1 Structure diagram of RF model.

(2) Random selection of features: for each node in every decision tree, only a randomly selected subset of features is considered to determine the optimal splitting scheme. This prevents specific features from dominating the model, thereby enhancing its robustness.

RF was verified to be effective in reducing overfitting and enhancing generalization when constructing decision tree models (Kumar et al., 2024). Let $H(x)$ represent the random forest model, and $h_1(x)$, $h_2(x)$,..., $h_m(x)$ are multiple decision trees, where x denotes the feature attributes of samples, Y is the class attributes of samples, and I is the indicator function. The predictions for classification and regression are defined as follows.

The prediction result for RF classification is calculated as:

$$H(x) = argmax \sum I(h_m(x)) = Y \tag{1}$$

The regression estimates of RF are obtained by minimizing the objective function of Eq. (2), which controls the variance reduction between bootstrap samples.

$$H(x) = avg(\sum h_m(x)) \tag{2}$$

Regression prediction exhibited high prediction accuracy and flexibility in dealing with complex nonlinear relationships (Yu et al., 2018). Therefore, the regression method of RF is utilized to predict the coolant bleed flow for aircraft engines.

TSO Algorithm

TSO is a swarm intelligence algorithm inspired by the cooperative hunting behavior of tuna swarms (Aldehim et al., 2023). This algorithm primarily simulates the spiral and parabolic foraging behaviors of tuna to solve and optimize real-world problems (Li et al., 2024). It is characterized by robust optimization capabilities and rapid convergence speed. During spiral foraging, tuna swarms achieve coordinated perception among neighboring individuals through information exchange. At this stage, the population position is represented as:

$$X_i^{t+1} = \begin{cases} \alpha_1 \cdot (X_{rand}^t + \beta \cdot |X_{rand}^t - X_i^t|) + \alpha_2 \cdot X_i^t, & i = 1 \\ \alpha_1 \cdot (X_{rand}^t + \beta \cdot |X_{rand}^t - X_i^t|) + \alpha_2 \cdot X_{i-1}^t, & i = 2,3,\dots,N \\ \alpha_1 \cdot (X_{best}^t + \beta \cdot |X_{best}^t - X_i^t|) + \alpha_2 \cdot X_i^t, & i = 1 \\ \alpha_1 \cdot (X_{best}^t + \beta \cdot |X_{best}^t - X_i^t|) + \alpha_2 \cdot X_{i-1}^t, & i = 2,3,\dots,N \end{cases} \tag{3}$$

$$\alpha_1 = a + (1-a) \cdot \frac{t}{t_{max}} \qquad \alpha_2 = (1-a) - (1-a) \cdot \frac{t}{t_{max}} \tag{4}$$

In the equation, α_1 and α_2 are weight coefficients controlling the tuna's tendency toward the optimal position and the position of the previous generation, respectively. Here, t represents the current iteration, t_{max} is the maximum number of iterations, and N is the population size. X_i^{t+1} denotes the position of the $i-th$ individual after the $(t+1)-th$ iteration, while X_{best}^t and X_{rand}^t are the current best individual and random individual, respectively. β is an exploitation parameter related to the optimal individual.

In addition to foraging through spiral formations, tuna also engage in parabolic cooperative foraging. They form a parabolic shape with the food source as a reference point while simultaneously searching for food in the surrounding area. To enhance the global and local performances of the TSO algorithm, it is assumed that the selection probability for both methods is 50%. The specific mathematical model is described as:

$$X_i^{t+1} = \begin{cases} X_{best}^t + rand \cdot (X_{best}^t - X_i^t) + TF \cdot p^2 \cdot (X_{best}^t - X_i^t), & rand < 0.5 \\ TF \cdot p^2 \cdot X_i^t, & rand \geq 0.5 \end{cases} \tag{5}$$

In the equation, *TF* is a random number taking the value of 1 or -1, and determines the direction of population exploitation. *p* is a nonlinear decreasing coefficient that influences the intensity of population evolution. Tuna are generated cooperatively through the above two methods, and constantly update their positions until the termination condition is met.

TSO-RF and Its Implementation process

In predicting coolant bleed flow for aircraft engines, the RF model demonstrates distinct advantages in regression tasks due to its exceptional nonlinear modeling capabilities and robust generalization performance. However, the performance of the model depends not only on the data itself, particularly the correlation between inputs and outputs, but also on its structural parameters. Key structural parameters include the number of trees and the minimum number of samples required for leaf nodes (Sandhu et al., 2024). These parameters directly influence the model's generalization, robustness, and adaptability to new data. Nevertheless, optimizing these parameters typically demands substantial computational resources and time. Manual parameter tuning is both tedious and time-consuming, prone to getting stuck in local optima. To address this issue, integrating optimization algorithms to automatically tune these parameters is particularly crucial. TSO, a novel swarm intelligence optimization algorithm, offers significant advantages such as fast convergence and efficient search capability. By simulating the foraging behavior of tuna swarms, this algorithm effectively searches for global optima while avoiding local optima traps (Xie et al., 2021). In this section, an ensemble

DOI : 10.5614/j.eng.technol.sci.2026.58.4.6

approach combining TSO and RF (TSO-RF) is presented, and its principle involves using TSO to search for the optimal structural parameters of the RF model, specifically the number of decision trees and the minimum number of samples required at a leaf node, thereby minimizing its prediction error for HPT and LPT cooling bleed air flow and ultimately achieving the required accuracy in practical engineering applications.

The following content is the specific workflow to establish a prediction model of the coolant bleed flow for aircraft engines using the TSO-RF algorithm, as illustrated in Figure 2. The specific implementation process is described below.

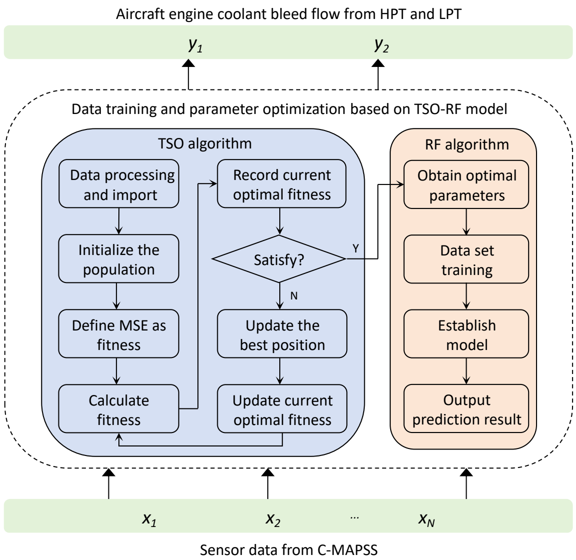


Figure 2 Flowchart of TSO-RF.

Step 1: Data processing and import. Calculate the correlation coefficients between the former 19 sensors and HPT/LPT from the C-MAPSS database. Take sensor data with moderate or higher correlation as input variables [x_1, x_2, ..., x_N], and [Y_1, Y_2] as output variables, then divide them into training and test data.

Step 2: Initialize the population. First, determine the size and the maximum number of iterations. Then, initialize the positions and velocities of all tuna individuals, and set the upper and lower bounds for these parameters.

Step 3: Define and calculate fitness. Each individual conducts an initial exploration within the search space to identify its optimal solution, and the fitness function value for each individual is computed. The mean squared error (MSE) is defined as the fitness function (Ayazi et al., 2024):

$$F(\theta) = \frac{1}{MSE(\theta)+\varepsilon} \tag{6}$$

MSE is employed as the criterion to evaluate the RF regression model, where $F(\theta)$ represents the fitness value of the current individual θ, and ε serves as an infinitesimal constant introduced to prevent division by zero (Xu et al., 2024).

Step 4: Search for the optimal solution. Local optimization is performed for potential food source individuals, updating their positions based on the current best solution. The adjustment is defined as:

$$\theta_{best} = argmax_{\theta} F(\theta) \tag{7}$$

Here, θ represents each individual tuna.

Step 5: Fitness evaluation and update. The fitness function value is recalculated for each individual, and the historical best solution for each individual is updated to move closer to the global optimal solution. The update formula is defined as:

$$\theta_{new} = \theta + \alpha \cdot (F(\theta_{best}) - F(\theta)) \tag{8}$$

Here, α is the factor controlling the pace of updates, and $F(\theta_{best})$ represents the fitness of the current best individual.

Step 6: Termination condition evaluation. Determine whether the maximum number of iterations has been reached or whether a satisfactory solution has been found. If yes, the optimal parameters are obtained, and the optimization process is concluded, yielding the optimal parameter combination for the RF model. If no, the current best fitness is updated, and the algorithm proceeds to the next iteration.

Step 7: Output the optimal parameters. Return the optimal solution and terminate the algorithm.

Performance testing

To validate the superior performance of the TSO algorithm, commonly used intelligent algorithms such as particle swarm optimization (PSO) and genetic algorithm (GA) were introduced for comparative testing. The benchmark parameters were set as follows: population size and maximum iteration count were 50 and 3000, respectively. The Ackley and Rosenbrock functions were employed for comparative evaluation (Amo et al., 2012; Zhang & Gao, 2019), with convergence curves depicted in Figures 3 and 4.

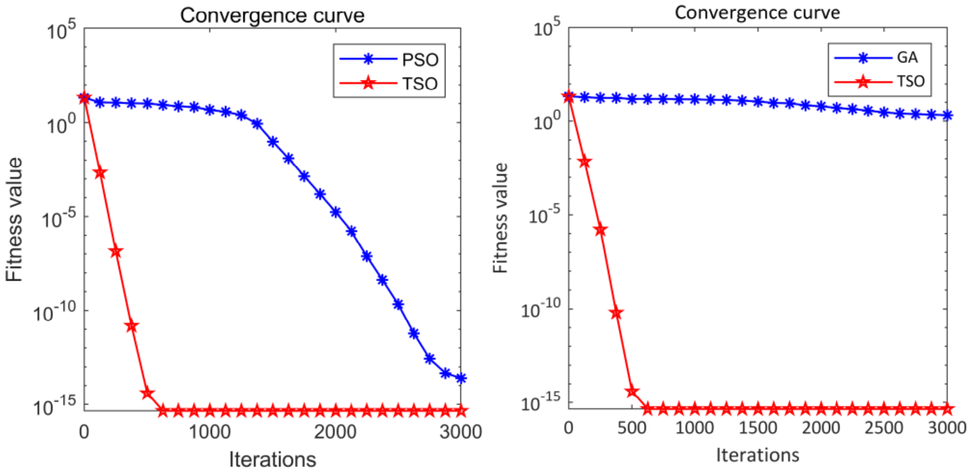


Figure 3 Performance comparison between TSO and PSO/GA with the Ackley function.

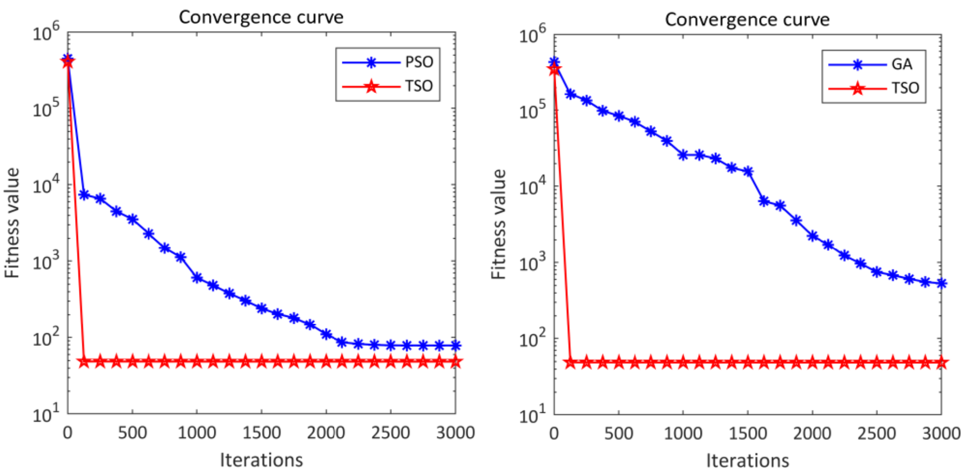


Figure 4 Performance comparison between TSO and PSO/GA with the Rosenbrock function.

Figures 3 and 4 display that, compared to PSO and GA, TSO demonstrates significantly faster convergence for the Ackley function, achieving the minimum fitness at the 608th iteration, while PSO and GA failed to converge. For the Rosenbrock function, TSO converged at the 126th iteration, whereas PSO gradually converged only at the 2175th iteration, and GA failed to converge throughout the iterations. Ultimately, the solution realized by TSO was closest to the optimal fitness value.

DOI : 10.5614/j.eng.technol.sci.2026.58.4.6

Coolant bleed flow modeling and prediction for aircraft engines

The aircraft engine coolant bleed flow data used in this study were obtained from C-MAPSS, a simulation platform widely used in the aerospace field for the research and design of aircraft engine performance. The dataset comprises four subsets: FD001 to FD004, and the coolant bleed flow is influenced by engine operating parameters. The dataset comprises six distinct operational conditions, and each subset includes training and test sets. Among these, FD001 and FD003 represent single-condition scenarios, capturing engine states during the cruise phase where operating parameters (altitude, Mach number, and thrust lever angle) are typically assumed constant. Conversely, FD002 and FD004 represent multi-condition scenarios, reflecting parameter variations across different operational phases such as taxiing, takeoff, cruise, and landing. Specific details are outlined in Table 1.

Table 1 Sample Information for the datasets.

Datasets	FD001	FD002	FD003	FD004
Number of engines in the training set	100	260	100	249
Total number of samples in the training set	20631	53579	24720	61249
Number of engines in the test set	100	259	100	148
Total number of samples in the test set	13096	33991	16596	41214

Within each dataset, 21 distinct sensors provided comprehensive monitoring data, including engine identification numbers, cycle indices, and operating parameters. Detailed data for these 21 sensors are presented in Table 2 (Saxena et al., 2008).

Table 2 Information related to 21 sensors.

Serial number	Instruction	Unit
x_1	Temperature of fan inlet	°R
x_2	Total temperature of LPC outlet	°R
x_3	Total temperature of HPC outlet	°R
x_4	Total temperature of LPT outlet	°R
x_5	Pressure at fan inlet	psia
x_6	Total pressure in bypass-duct	psia
x_7	Total pressure at HPC outlet	psia
x_8	Physical fan speed	rpm
x_9	Physical core speed	rpm
x_{10}	Engine pressure ratio(P50/P2)	-
x_{11}	Static pressure at HPC outlet	psia
x_{12}	Ratio of fuel flow to Ps30	pps/psi
x_{13}	Corrected fan speed	rpm
x_{14}	Corrected core speed	rpm
x_{15}	Bypass Ratio	-
x_{16}	Burner fuel-air ratio	-
x_{17}	Bleed Enthalpy	-
x_{18}	Demanded fan speed	rpm
x_{19}	Demanded corrected fan speed	rpm
y_1	HPT coolant bleed	lbm/s
y_2	LPT coolant bleed	lbm/s

the variables collected by the C-MAPSS sensors have different units, normalization was performed before coolant bleed flow modeling, and the following expression is employed (Mahmoud et al., 2024):

$$X = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{9}$$

The Pearson correlation coefficient is a commonly used index for describing the strength of linear relationships between two variables, offering the advantages of conceptual visualization and efficient computation, and its computational formula is given as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{10}$$

After the Pearson correlation coefficients were calculated using Eq. (10), the correlation strengths between variables were classified according to the ranges listed in Table 3.

Table 3 Correlation strength characterized by the Pearson coefficient.

$\lvert r \rvert$	The strength of correlation
1.0 - 0.8	Very strong correlation
0.8 - 0.6	Strong correlation
0.6 - 0.4	Moderate correlation
0.4 - 0.2	Weak correlation
0.0 - 0.2	Very weak or negligible correlation

Table 4 lists the Pearson coefficients between the 19 sensor data from FD001 to FD004 and the variables Y_1 and Y_2 representing HPT and LPT, respectively. It indicates that FD002 and FD004 exhibit high correlations with Y_1 and Y_2 for all variables except X_{13} and X_{19}, while FD001 and FD003 have mostly low correlations according to Table 3. As is well known, the correlation between inputs and outputs directly impacts modeling accuracy. To establish a high-precision predictive model for coolant bleed flow, the sensor data for FD001, FD003, X_{13} and X_{19} in FD002 and FD004 were excluded. Consequently, 17 sensors from FD002 and FD004 with higher correlations to HPT and LPT were ultimately selected as the data sources for this case.

Table 4 Pearson correlation results for FD001 to FD004.

	FD001(Y_1)	FD002(Y_1)	FD003(Y_1)	FD004(Y_1)	FD001(Y_2)	FD002(Y_2)	FD003(Y_2)	FD004(Y_2)
x_1	0	0.9779	0	0.9774	0	0.9779	0	0.9774
x_2	0.4570	0.9629	0.1571	0.9624	0.4669	0.9629	0.1679	0.9625
x_3	0.4139	0.9182	0.0190	0.9137	0.4097	0.9182	0.0219	0.9173
x_4	0.5772	0.9479	0.1890	0.9474	0.5625	0.9479	0.1935	0.9474
x_5	0	0.9858	0	0.9856	0	0.9859	0	0.9856
x_6	0.1303	0.9964	0.2375	0.9963	0.1395	0.9964	0.2230	0.9963
x_7	0.5695	0.9992	0.7075	0.9991	0.5646	0.9992	0.7079	0.9991
x_8	0.5611	0.6569	0.2385	0.6539	0.5615	0.6568	0.2437	0.6540
x_9	0.0361	0.9118	0.4472	0.9108	0.0283	0.9118	0.4431	0.9108
x_{10}	0	0.9108	0.4297	0.9071	0	0.9108	0.4265	0.9071
x_{11}	0.6075	0.7866	0.1377	0.7853	0.6005	0.7865	0.1392	0.7853
x_{12}	0.5914	0.9991	0.7027	0.9991	0.5922	0.9991	0.7034	0.9991
x_{13}	0.5560	0.2764	0.2378	0.2737	0.5602	0.2763	0.2439	0.2737
x_{14}	0.0807	0.4867	0.5228	0.4849	0.0952	0.4866	0.5177	0.4849
x_{15}	0.5107	0.6628	0.7005	0.6607	0.5147	0.6627	0.6994	0.6607
x_{16}	0	0.8558	0	0.8528	0	0.8558	0	0.8528
x_{17}	0.4514	0.9202	0.0345	0.9196	0.4482	0.9202	0.0258	0.9196
x_{18}	0	0.6564	0	0.6536	0	0.6563	0	0.6536
x_{19}	0	0.2761	0	0.2737	0	0.2761	0	0.2737

To enhance the robustness and generalization ability of the coolant bleed flow model, the maximum possible amount of data was used for training, using the common 8:2 ratio to partition the selected dataset into training and test sets (Wang et al., 2023; Xiao & Zeng, 2024), yielding the sample configurations listed in Table 5. Taking FD002 as an example, the training set comprises the first 27,193 samples, while the remaining 6,792 samples form the test set.

Table 5 Data partition results for the samples.

	Data for training	Data for test
FD002	1 - 27193	27194 - 33991
FD004	1 - 32971	32972 - 41214

In the RF framework, the number of decision trees (a) and the minimum sample size per leaf node (b) are critical parameters affecting model performance. Preliminary tests employed default structural parameters with the optimization range set around these default values (Breiman, 2001). The constraint ranges for the parameters are specified in Table 6. An excessively small value of a can lead to underfitting, while an overly large value of b may cause excessive simplification of the model. It is noted that all subsequent optimized RF models were trained within the same parameter ranges described above, including the boundaries.

DOI : 10.5614/j.eng.technol.sci.2026.58.4.6

Table 6 Parameter optimization ranges for the RF model.

Structural parameter	Physical meaning of parameters	Optimization range
a	The number of decision trees	(100, 800)
b	The minimum sample size per leaf node	(1, 12)

Based on the engine coolant bleed flow modeling process using the TSO-RF algorithm in Figure 2, and to validate the algorithm's superior predictive accuracy, PSO and GA were introduced to optimize the RF model. The running codes for the PSO-RF, GA-RF, and TSO-RF algorithms were developed on the MATLAB platform. By applying the sample datasets from Table 5 and parameter optimization ranges set in Table 6, the final structural parameter optimization results for these three models were obtained, as listed in Tables 7 and 8, respectively.

Table 7 Optimal structural parameters of three models for FD002.

Structural parameter	PSO-RF	GA-RF	TSO-RF
a	512	675	633
b	5	3	5

Table 8 Optimal structural parameters of three models for FD004.

Structural parameter	PSO-RF	GA-RF	TSO-RF
a	663	658	800
b	5	3	3

To evaluate the predictive performance of the established RF models, four metrics were employed: MSE, mean absolute error (MAE), root mean squared error (RMSE) and determination coefficient R^2 (Huang et al.,2025). Their calculation formulas are showed below:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2 \tag{11}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y_i}| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2} \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y_i})^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{14}$$

where N represents the number of samples in the test set, $\hat{y_i}$ denotes the model's predicted value, and y_i stands for the actual value. The lower the values of the first three metrics, the better the predictive accuracy of the model. In addition, R^2 serves as a measure of how well the regression model explains the variation in observed data (Waqas et al. 2024). A higher R^2 value indicates greater accuracy.

Under the structural parameters listed in Tables 7 and 8 above, the predicted results of the engine coolant bleed flow for the four RF models are presented in Tables 9 to 12, respectively. Among them, the fitting results of y1 as an example are shown in Figures 5 to 8.

The conclusions are drawn from the following tables and figures: regardless of the datasets from FD002 and FD004 or Y_1 and Y_2, the standard RF model exhibits the lowest prediction and fitting accuracy; in contrast, the optimized TSO-RF model demonstrates the best predictive performance, while the GA-RF model outperforms the PSO-RF model overall. Therefore, the TSO-RF models established with the structural parameters listed in Tables 7 and 8 can be regarded as effective engineering tools for predicting aircraft engine coolant bleed flow.

Table 9 Prediction results of FD002 for Y_1.

Model	MSE	MAE	RMSE	R^2
RF	0.06431	0.1532	0.2536	0.9956
PSO-RF	0.00382	0.0492	0.0618	0.9982
GA-RF	0.00271	0.0426	0.0521	0.9981
TSO-RF	0.00013	0.0093	0.0115	0.9989

Table 10 Prediction results of FD004 for Y_1.

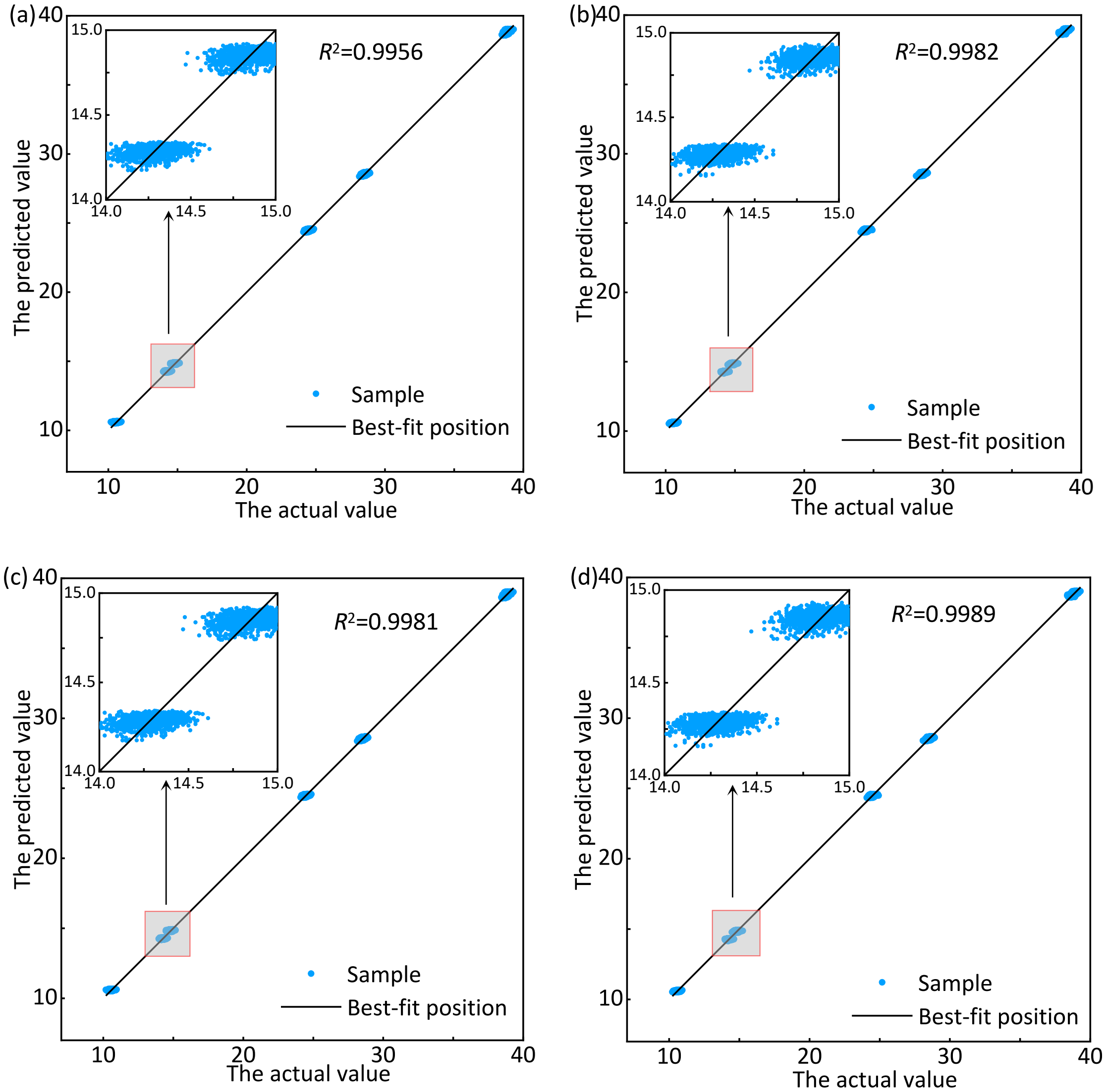
Model	MSE	MAE	RMSE	R^2
RF	0.07901	0.1782	0.2811	0.9976
PSO-RF	0.00271	0.0485	0.0521	0.9988
GA-RF	0.00261	0.0412	0.0511	0.9986
TSO-RF	0.00015	0.0101	0.0123	0.9992

Table 11 Prediction results of FD002 for Y_2.

Model	MSE	MAE	RMSE	R^2
RF	0.06431	0.1529	0.2536	0.9976
PSO-RF	0.00357	0.0479	0.0598	0.9991
GA-RF	0.00358	0.0431	0.0599	0.9989
TSO-RF	0.00016	0.0102	0.0128	0.9998

Table 12 Prediction results of FD004 for Y_2.

Model	MSE	MAE	RMSE	R^2
RF	0.07907	0.1821	0.2812	0.9978
PSO-RF	0.00271	0.0483	0.0521	0.9989
GA-RF	0.00225	0.0399	0.0475	0.9986
TSO-RF	0.00014	0.0093	0.0118	0.9991



DOI : 10.5614/j.eng.technol.sci.2026.58.4.6

Figure 5 The fitting accuracy of FD002 for y1: (a) RF; (b) PSO-RF; (c) GA-RF; (d) TSO-RF.

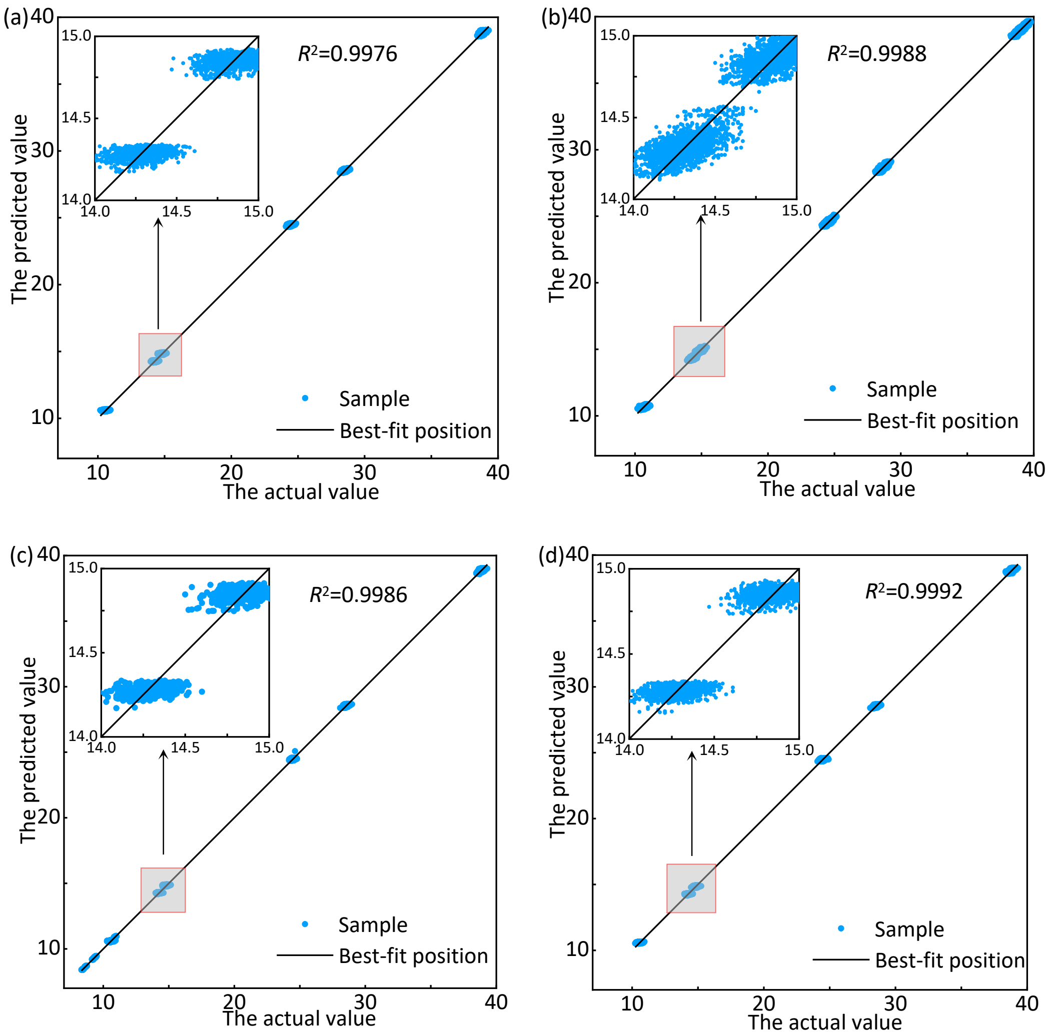


Figure 6 The fitting accuracy of FD004 for y1: (a) RF; (b) PSO-RF; (c) GA-RF; (d) TSO-RF.

According to the comparative analysis of the optimization mechanisms of TSO-RF, PSO-RF, and GA-RF in Tables 9 to 12, the superior performance of TSO-RF over PSO-RF and GA-RF is not merely reflected in a marginal increase in R^2 or a slight reduction in MAE. Rather, its superiority stems from a fundamental alignment between its algorithmic mechanics and the core characteristics of coolant bleed flow from HPT and LPT.

The spiral foraging strategy (Eqs. 3-4) of the TSO algorithm is particularly adept at accurately capturing the relationships inherent in the C-MAPSS sensor data related to cooling bleed flow. Concurrently, its parabolic foraging strategy (Eq. 5) enables rapid correction of localized prediction errors during transitions between multiple operational regimes. In contrast, the PSO-RF model relies on velocity updates where particles follow the global and individual historical best solutions. This often leads to the premature convergence of parameters during a stable regime like cruise, resulting in significant prediction bias when operational conditions shift. The GA-RF model, dependent on population crossover and random mutation to generate new individuals, requires a full population fitness evaluation with each iteration. This computational inefficiency makes GA-RF less suitable for the real-time prediction of engine coolant bleed flow.

Consequently, the TSO-RF model proves to be a more effective solution for the modeling challenges in predicting coolant bleed flow from HPT and LPT. It not only ensures predictive accuracy but also enhances the interpretability of the model, thereby providing a novel and robust tool for coolant bleed flow of aircraft turbofan engines.

Impact analysis of parameter weights for the TSO-RF model

To investigate the influence of various factors on turbine engine coolant bleed flow, the Gini coefficient inherent in the RF algorithm (Deng & Runger, 2012) was employed using FD004 as an example to calculate the impact weights of 17 input parameters on both HPT and LPT. The specific implementation steps are as follows:

Utilizing RF's bootstrap sampling, a subset of samples was randomly selected. Following the discretization of the continuous coolant flow values, the initial Gini impurity at the root node was computed to serve as the splitting criterion, with the following formula:

$$G(t) = 1 - \sum_{k=1}^{K} p_{t,k}^2 \tag{15}$$

Where t denotes an individual tree in the forest, and $p_{t,k}$ represents the proportion of samples belonging to the k-th flow interval.

For each of the 17 input parameters, the left and right child nodes $N_{t,L}/N_{t,R}$ resulting from a split are identified. The total Gini impurity after the split, $G(t,j,v)$, is then calculated as the weighted average of the impurities of these child nodes, with the weights being the proportion of samples in each node. The formula is described as:

$$G(t,j,v) = \frac{N_{t,L}}{N_t} \cdot G(t,L) + \frac{N_{t,R}}{N_t} \cdot G(t,R) \tag{16}$$

Where v is the split threshold using the mean value as the threshold; $N_{t,L}$ and $N_{t,R}$ are the number of samples in the left and right child nodes, respectively, and $G_{t,L}$ and $G_{t,R}$represent the Gini impurity of the left and right child nodes.

The importance contribution of a parameter in the current tree, denoted as $Imp(t,j)$, is obtained by subtracting the minimum total impurity $min_v[G(t,j,v)]$, corresponding to the parameter's optimal split threshold from the initial Gini impurity of the single tree. Its mathematical expression is given as:

$$Imp(t,j) = G(t) - min_v[G(t,j,v)] \tag{17}$$

One hundred decision trees were empirically determined to be sufficient for capturing the distributional characteristics of the dataset. To maintain computational efficiency, the importance contributions of the 17 input parameters were averaged across these 100 trees to obtain the global raw importance, denoted as $Importance(j)$, namely:

$$Importance_{raw}(j) = \frac{1}{T} \sum_{m=1}^{T} Imp(t,j) \tag{18}$$

where $Imp(t,j)$ denotes the importance contribution of parameter j in the t-th tree.

The global raw importance scores of all 17 parameters were summed and normalized. This process enables a direct comparison of the relative priority of each parameter's influence.

$$Importance(j) = \frac{Importance_{raw}(j)}{\sum_{m=1}^{17} Importance_{raw}(m)} \tag{19}$$

Ultimately, the influence weight analysis based on the Gini coefficient was performed for the 17 input variables in the model. Figure 7 and Figure 8 present the parameter weighting results for the coolant bleed flows from HPT and LPT, respectively.

DOI : 10.5614/j.eng.technol.sci.2026.58.4.6

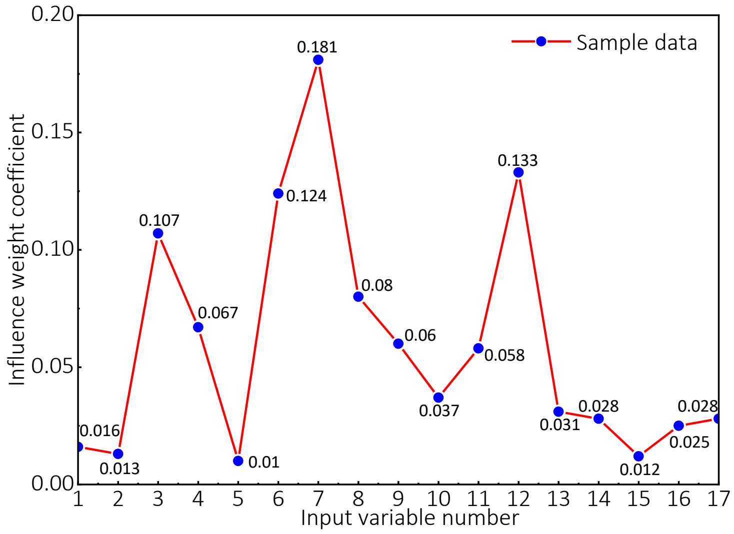


Figure 7 Parameter weighting results for HPT.

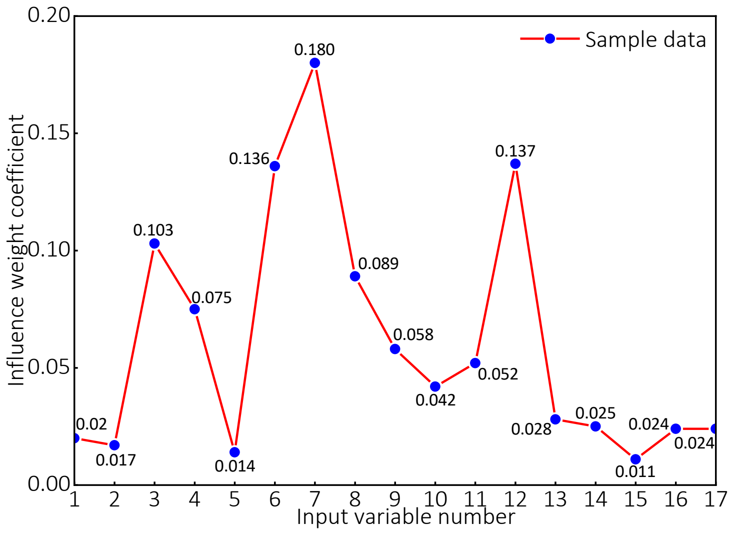


Figure 8 Parameter weighting results for LPT.

The analytical findings presented in Figures 7 and 8 demonstrate that the total exhaust pressure of the high-pressure compressor (HPC) (corresponding to parameter x_7 in Table 2) is the parameter exerting the most pronounced influence on the system output. An elevation in this pressure augments the pressure differential between the HPC outlet and the high-pressure/low-pressure turbines (HPT/LPT) (Xiu et al., 2025), subsequently escalating the leakage flow rate of the cooling medium. Consequently, managing the HPC discharge total pressure is a critical factor for assessing and controlling coolant leakage in aircraft turbine engines. However, given that turbine engine components operate in a highly synergistic manner, it is imperative to further investigate the extent to which the interaction between this variable and other parameters influences the output. In practical terms, this understanding carries significant implications for engine design and aviation emissions regulations.

Discussion

Tables 9 to 12 demonstrate that, in comparison with the standard RF model, the proposed TSO-RF model exhibits only a slight advantage in terms of the R^2 metric. To further conduct a statistical comparison of the performance between the two models, this section employs non-parametric tests on the predicted values from the FD002 and FD004 datasets, with a significance level of α = 0.05 (Bonnini & Borghesi, 2024). The non-parametric tests were used to assess whether the observed improvements were statistically significant by comparing the distributions of prediction residuals between model groups. The results of this significance analysis for FD002 and FD004 are presented in Table 13.

Table 13 Results of significance analysis.

Dataset	Prediction target	Comparison	p
FD002	Y_1(HPT)	TSO-RF vs RF	0.023
FD002	Y_2(LPT)	TSO-RF vs RF	0.018
FD004	Y_1(HPT)	TSO-RF vs RF	0.031
FD004	Y_2(LPT)	TSO-RF vs RF	0.027

As shown in Table 13, the p-values for the comparative groups are all less than 0.05. This indicates that the improvement in model performance was not attributable to random fluctuations, but rather represented a stable enhancement achieved through the TSO optimization algorithm. For an aero-engine's cooling bleed air system, this predictive accuracy can directly influence both engine thrust control and health condition assessment.

As shown in Figures 7 and 8, the feature importance analysis based on the Gini coefficient not only validates the profound physical implications of the coolant bleed flow mechanism but also provides an engineering-level explanation for the high prediction accuracy of the TSO-RF model. The high-pressure compressor (HPC), as the dominant factor in the system, exhibits a strong correlation with the coolant bleed flow. This correlation enables the spiral foraging mechanism of the TSO algorithm to more accurately capture the nonlinear variation patterns of the core variables. In addition, secondary influential features such as total pressure in the bypass-duct (corresponding to parameter x_6 in Table 2) and the ratio of fuel flow to Ps30 (corresponding to x_{12} in Table 2) reflect the interconnectedness of the engine system. The total pressure in the bypass-duct influences the overall pressure field of the engine, while the ratio of fuel flow to Ps30 is related to combustion efficiency and heat generation, indirectly regulating the demand for coolant bleed flow (Xiu et al., 2025). The interdependencies among these features highlight that coolant bleed flow is not an isolated parameter, but rather the result of coordinated interactions among multiple engine components, offering a theoretical basis for targeted sensor monitoring and parameter adjustment.

Regarding the sensitivity comparison between the FD002 and FD004 datasets, the applicability and inherent characteristics of the model are further clarified. Both datasets represent multi-condition scenarios, where sufficient sample sizes and extensive operating condition coverage effectively enhance the model's generalization capability. However, the TSO-RF model performs poorly in predicting Y_1 and Y_2 on single-condition datasets (FD001 and FD003). This is not necessarily a limitation of TSO-RF itself, but rather a consequence of the fact that the Pearson correlation coefficients between most of the 19 sensors in FD001/FD003 and Y_1/Y_2 are below 0.4, indicating a lack of core input information that effectively reflects changes in coolant bleed flow. Moreover, the model's strong reliance on standardized C-MAPSS simulation data makes it susceptible to unmodeled disturbances in real flight scenarios, such as measurement errors due to sensor aging or inlet airflow fluctuations under extreme weather conditions. These disturbances can disrupt the relationship between input features and coolant bleed flow, leading to prediction biases and subsequently misleading adaptive bleed air control strategies. Excessive cooling may result in thrust loss, while insufficient cooling can accelerate turbine component wear, directly reducing the engine's RUL. Additionally, although the model can capture the overall dynamic changes in coolant bleed flow, it cannot distinguish between normal flow fluctuations and abnormal leakage caused by early-stage faults such as seal wear. Accurate identification of early-stage faults is a core prerequisite for extending engine RUL. This limitation hinders its deep integration into engine health management systems.

Conclusion

This paper aims to employ the TSO-RF algorithm for establishing high-accuracy engine coolant bleed flow models, an emerging application in the aircraft field. Key research findings are summarized as follows:Within the available data from C-MAPSS, not all datasets exhibit moderate correlation with LPT and HPT, and the correlation between 17 sensor variables in FD002 and FD004 and LPT/HPT are acceptable, with the Pearson coefficients exceeding 0.4. This may be associated with their structural forms and operational mechanisms.Regarding the enhancement of predictive performance, compared with the standard RF model, the TSO-RF model achieved the greatest reduction in MSE, MAE, and RMSE, along with the highest improvement in $\boldsymbol{R^2}$, and the GA-RF model followed, with the PSO-RF model ranking last. This fully demonstrates that the constructed TSO-RF model possesses a significant advantage in predicting the coolant bleed flow for aircraft engines, which provides an effective technical solution for engine bleed air system management and evaluation.

Acknowledgements

The work was supported by Major Educational Research Project of Fujian Province (FBJY20230154).

DOI : 10.5614/j.eng.technol.sci.2026.58.4.6

Data Availability Statement

These public data are derived from the NASA-C-MAPSS:
https://phm-datasets.s3.amazonaws.com/NASA/6.+Turbofan+Engine+Degradation+Simulation+Data+Set.zip.

Compliance with ethics guidelines

The authors declare they have no conflict of interest or financial conflicts to disclose.

This article contains no studies with human or animal subjects performed by the authors.

References

Aldehim, G., Asiri, M.M., Aljebreen, M., Mohamed, A., Assiri, M. & Ibrahim, S. S.(2023). Tuna swarm algorithm with deep learning enabled violence detection in smart video surveillance systems. *IEEE Access*, *11*, 95104–95113. https://doi.org/10.1109/ACCESS.2023.3310885

Amo, D. G. I., Pelta, A. D., Gonzalez, R. J., & Masegosa, D. A.(2012). An algorithm comparison for dynamic optimization problems. *Applied Soft Computing Journal*, *12*(10), 3176–3192. https://doi.org/10.1016/j.asoc.2012.05.021

Ayazi, F. M., Singh, M., & Kumar, R. (2024). Prediction and modelling marshall stability of modified reclaimed asphalt pavement with rejuvenators using latest machine learning techniques. *Engineering Research Express*, *6*(3), 035004-035004. https://doi.org/10.1088/2631-8695/AD65B7

Breiman, L.(2001). Random forests. *Machine Learning*, https://doi.org/10.1023/A:1010933404324

Bonnini, S., Borghesi, M.(2024). Nonparametric test for logistic regression with application to italian enterprises' propensity for innovation. *Mathematics*, *12*(18), 2955–2955. https://doi.org/10.3390/MATH12182955

Cai, W., Zhao, Y. P., Zhu, Y., Yin, J. & Xu, Z. Y.(2024). A novel hybrid aeroengine modeling method for combining data-driven modules. *Journal of Aerospace Engineering*, *37*(5). https://doi.org/10.1061/JAEEEZ.ASENG-5531

Cristina, M. R.(2023). Theory and foundations of modeling and simulation. *Simulation-Transactions of The Society for Modeling and Simulation International*, *99*, 431-432. https://doi.org/10.1177/00375497231171140

Duan, Y. T., Zhang, T., & Shi, D.(2024). Anomaly detection and remaining useful life prediction for turbofan engines with a key point-based approach to secure health management. *Sensors*, *24*, 8022–8022. https://doi.org/10.3390/S24248022

Deng, H. T., Runger, C. G.(2012). Feature selection via regularized trees. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE.

Guan, L. R., Sun, H. J., Chen, F. J., Shen, W. & Cai, Y. Z.(2024). CFD investigation of structural effects of internal gas intake on powder conveying performance in fuel supply systems for aerospace engines. *Particuology*, 92140–92154. https://doi.org/10.1016/j.partic.2024.04.016

Huang, Y., Guo, S., Xie, H. L., Wen, J. & Wang, J. R.(2025). Evaluation of machine learning models for predicting performance metrics of aero-engine combustors. *Case Studies in Thermal Engineering*, *65*, 105627. https://doi.org/10.1016/J.CSITE.2024.105627

Khan, A., Sutanto, M. H., Khan, N., Baarimah, A. O., Wahab, M. M. A., Khan, M. I., Ramu, M. B., & Khan, R. (2025). Decision tree machine learning approach for the performance prediction of asphalt mixes modified with waste tyre metal fibre. *Journal of Engineering and Technological Sciences*, *57*(3), 303–314. https://doi.org/10.5614/j.eng.technol.sci.2025.57.3.2

Kumar, S. M., Nagarajan, L., & Vasudevan, B.(2024). Machinability study on abrasive waterjet drilling of YSZ-coated inconel 718 superalloy: optimization and modelling by machine learning. *Engineering Research Express*, *6*, 025425. https://doi.oth/10.1088/2631-8695/ad51d3

Liu, Y. L., Xu, G. Q., Fu, Y. C., Wen, J. & Huang, H. R.(2023). Thermal dynamic and failure research on an air-fuel heat exchanger for aero-engine cooling. *Case Studies in Thermal Engineering*, *42*. https://doi.org/10.1016/J.CSITE.2023.102715

Lei, B.Y., Shen, N., Ji, D.Y., Ouyang, Z. R., Yang, L., & Liu, M. Q. (2024). A comprehensive review of key technologies for enhancing the reliability of aero-engine. *IEEE Access*, *12*, 148315–148331. https://doi.org/10.1109/ACCESS.2024.3474737

Lin, R. H., Pei, Z. X., Ye, Z. Z., Wu, B. D., & Yang, D.(2020). A voted based random forests algorithm for smart grid distribution network faults prediction. *Enterprise Information Systems*, *14*, 496–514. https://doi.org/10.1080/17517575.2019.1600724

Li, L.L., Ji, B.X., Lim, M.K., Tseng, & L. M.(2024). Active distribution network operational optimization problem: A multi-objective tuna swarm optimization model. *Applied Soft Computing*, *150*, 111087. https://doi.org/10.1016/j.asoc.2023.111087

Madziel, M.(2024). Quantifying emissions in vehicles equipped with energy-saving start–stop technology: THC and NOx modeling insights. *Energies*, *17*(12), 2815–2815. https://doi.org/10.3390/EN17122815.

Mahmoud, A. E., Bassel, S., Bou, A. N., Mohammad Noor, I.(2024). Comparative study of ML models for IIoT intrusion detection: impact of data preprocessing and balancing. *Neural Computing & Applications*, *36*, 6955–6972. https://doi.org/10.1007/S00521-024-09439-X

Selvara, S., Thulasi, N. J., Vijayakumari, J. J. L. M., & Selvaraj, K.(2025). Deep residual ensemble model for predicting remaining useful life of turbo fan engines. *International Journal of Turbo & Jet-Engines*, *42*(2), 229–239. https://doi.org/10.1515/TJJ-2024-0025

Stephen D, S., Iohn, S., Alexis, H., Kevin, D. & Ariana, M.(2022). Rotating detonation combustion for advanced liquid propellant space engines. *Aerospace*, *9*(10), 581-581. https://doi.org/10.3390/AEROSPACE9100581

Sandhu, G., Singh, A., Bedi, S. P., Lamba, S. P. & Chaudhary, G.(2024). Improvement of random forest ensembling algorithm efficiency through cardinal tuning of n_estimators parameter. *Journal of Multiple-Valued Logic and Soft Computing*, *42*, 67–93.

Saxena, A., Goebel, K., Simon, D., Eklund, N.(2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *2008 International Conference on Prognostics and Health Management*, Denver, CO, USA: 1-9. https://doi.org/10.1109/PHM.2008.4711414

Tian, E. L.,Lu, G. N. & Li, Z. H.(2024). Evaluation of emission of the hydrogen-enriched diesel engine through machine learning. *Energy*, 307132303-132303. https://doi.org/10.1016/J.ENERGY.2024.132303

Waqas, M., Humphries, W. U., Hlaing, T. P. & Ahmad, S.(2024). Seasonal WaveNet-LSTM: a deep learning framework for precipitation forecasting with integrated large scale climate drivers. *Water*, *16*(22), 3194. https://doi.org/10.3390/W16223194

Waqas, M., Humphries, W. U., Wangwongchai, A., Dechpichai, P. & Ahmad, S.(2023). Potential of artificial intelligence-based techniques for rainfall forecasting in thailand: a comprehensive review. *Water, 15*, 2979. https://doi.org/10.3390/W15162979

Wang, H. R., Li, D. W., Li, D. J., Liu, C. Q., Yang, X. Q. & Zhu, G. F.(2023). Remaining useful life prediction of aircraft turbofan engine based on random forest feature selection and multi-Layer perceptron. *Applied Sciences, 13*,7186. https://doi.org/10.3390/APP13127186

Xie, L., Han, T., Zhou, H., Zhang, Z. R. & Han, B.(2021). Tuna swarm optimization: A novel swarm-based metaheuristic algorithm for global optimization. Computational Intelligence and Neuroscience, 9210050. https://doi.org/10.1155/2021/9210050

Xu,Y.Y., Xiao,H., Zhang, R.H., Wu, W. Z. & Wan, F.(2024). Matrix factorization with a sigmoid-like loss control. *Neurocomputing*, *576*, 127338-. https://doi.org/10.1016/J.NEUCOM.2024.127338

Xiao, L., & Zeng, Y.(2024). Remaining life prediction of turbofan engines based on RF-SA-SDCNN. *Agricultural equipment & vehicle engineering*. *62*(03), 153-158. https://doi.org/10.1038/s41598-022-10191-2

Xiu, X. Y., Ma, S. S., Guo, F. F., Liu, H. & Li, H. C.(2025). Performance analysis of a turbofan engine integrated with flame-assisted fuel cells for combined propulsion and power generation with more electric aircrafts. *Energy Conversion and Management*, *325*, 119335–119335. https://doi.org/10.1016/J.ENCONMAN.2024.119335

Yang, H., Yang, X.D. & Li, G. H. (2023). Dual feature extraction system for ship-radiated noise and its application extension. *Ocean Engineering, 285*(2), 115352. https://doi.org/10.1016/j.oceaneng.2023.115352

Yuhas, A. J. & Ray, R. J.(2012). Effects of bleed air extraction of thrust levels on the F404-GE-400 turbofan engine. https://doi.org/10.2514/6.1992-3092

Yu, B., Shu, W., Cao, C.(2018). A novel modeling method for aircraft engine using nonlinear autoregressive exogenous (NARX) models based on wavelet neural networks. *International Journal of Turbo & Jet-Engines*, *35*(2), 161–169. https://doi.org/10.1515/tjj-2017-0005

Zhu, Z., Wang, T., Hessam, S., Sarjoughian, H., Wang, W. P. & Zhao, Y.H.(2023). Knowledge-based and data-driven behavioral modeling techniques in engagement simulation. *Simulation-Transactions of The Society for Modeling and Simulation International*, *99*(10), 1069–1089. https://doi.org/10.1177/00375497231177123

Zhang, S. P. & Gao, D.(2019). Flower pollination algorithm based on dynamic adjustment and cooperative. Search Computer Engineering and Applications, *55*, 46–53.

Manuscript Received: 2 July 2025
1st Manuscript Revision Received: 4 October 2025
2nd Manuscript Revision Received: 22 November 2025
3rd Manuscript Revision Received: 2 February 2026
Accepted Manuscript: 4 May 2026