

A research on improving mean value engine model accuracy based on the random forest algorithm

Haiyan Wang; Zhihui Li*; Kelin Wu

Merchant Marine College, Shanghai Maritime University,
Shanghai 201306, China.

*Corresponding Author: Zhihui Li

Merchant Marine College, Shanghai Maritime University,
Shanghai 201306, China.

Tel: 8618817502960; Email: 202230110114@stu.shmtu.edu.cn

Received: Oct 05, 2024

Accepted: Nov 15, 2024

Published Online: Nov 22, 2024

Website: www.joaiar.org

License: © Li Z (2024). This Article is distributed under the terms of Creative Commons Attribution 4.0 International License.

Volume 1 [2024] Issue 2

Abstract

The indicated thermal efficiency is a core parameter that significantly affects the calculation accuracy of the fuel consumption rate in the diesel Mean Value Engine Model (MVEM). In order to improve the accuracy of the mean value engine model, a method based on Random Forest Regression Model (RF-R) was proposed to predict the indicated thermal efficiency. Using experimental universal characteristic data, the random forest regression model for predicting indicated thermal efficiency was constructed. Three-dimensional surface and contour plots of the diesel engine's fuel consumption rate were created and the model was evaluated. The mean square errors of the training and test sets were 0.0056 and 0.0059, respectively, and the mean absolute errors were 0.0038 and 0.0043, respectively. The mean value engine model using random forest to predict the indicated thermal efficiency is simulated and verified according to the load characteristics, the results show that the mean value engine model using random forest indication thermal efficiency regression prediction has high simulation accuracy across the entire operating range of diesel engine, and compared with the model using MAP diagram calculation method, the simulation results of fuel consumption rate under each working condition are closer to the measured value, and the simulation error is less than 5%.

Keywords: Universal characteristic; Indicted thermal efficiency; Random forest regression; Simulation accuracy.

Introduction

The mean value engine model, a simple model capable of describing diesel engine operating characteristics, is widely used in simulating diesel engine performance prediction and fault diagnosis. Therefore, further improving the simulation accuracy of the model has significant theoretical and engineering value. Many scholars have researched enhancing the simulation accuracy of the mean value engine model, primarily through optimizing modeling and improving core parameter prediction accuracy. Based on the Matlab/Simulink, Zhou Dong, Su Tiexiong, and others [1] considered the heat transfer in the intake system and established a diesel engine intake system model.

They concluded that under steady-state conditions, simplifying the model to be adiabatic is appropriate, while under transient conditions, using an intake manifold heat exchange model improves simulation accuracy. Haitao Zhou, Ping Yan, and others [2] used a quasi-dimensional combustion model to establish a neural network model for predicting the indicated thermal efficiency of a common rail diesel engine based on five variables: speed, injection timing, excess air coefficient, injection duration, and injection pressure. This approach addressed the lack of suitable virtual prototypes and the low accuracy of the mean value engine model in electronic control system development. Chuanlei Yang, Wenle Zhang, and others [3] employed GT-power software to build a one-dimensional steady-state diesel engine

Citation: Wang H, Li Z, Wu K. A research on improving mean value engine model accuracy based on the random forest algorithm. *J Artif Intell Robot.* 2024; 1(2): 1013.

mean value simulation model. They proposed a dual response surface fitting method to solve the problem of handling turbo-charger characteristic parameters. Haiyan Wang, Yihuai Hu, and others [4] proposed a state space model with four variables for turbocharged diesel engines, ensuring real-time simulation accuracy while simplifying the sub-models, thereby laying the foundation for improving the overall model accuracy. Yuanyuan Tang and others [5], through an analysis of the airflow path in diesel engines, suggested adding a cylinder block sub-model between the scavenging manifold and cylinder sub-models. Based on AVL BOOST software, they established a two-stroke ultra-long stroke marine diesel engine model. Simulation analysis revealed that after adding the cylinder block sub-model, the simulation accuracy of various performance parameters of the diesel engine improved significantly. Notable improvements were seen in exhaust manifold temperature, scavenging manifold temperature, and main engine power simulation accuracy. Yanwu Ge, Ying Huang, and others [6] proposed a closed-loop correction method for the torque estimation of a turbocharged diesel engine mean value model based on engine speed using a Kalman Filter (KF) on the Matlab/Simulink. After correction with the KF, the steady-state simulation error of the model was reduced, and the dynamic simulation accuracy was improved to 90%.

Choosing a complex intake system model provides more comprehensive simulation results, but it increases the difficulty of model parameter calibration. Using a neural network to establish a predictive model for indicated thermal efficiency with five variables yields high prediction accuracy. However, the network structure greatly influences the model's accuracy, its robustness is difficult to ensure, and the interpretability of the neural network's internal mechanisms is weak. Establishing a one-dimensional mean value engine model enhances accuracy but also increases computational complexity. Utilizing a dual response surface to handle turbocharger parameters simplifies the model processing and improves simulation accuracy, but the model simplification can lead to reduced accuracy. Adding a cylinder block sub-model improves model accuracy and applicability, but it increases model complexity, requiring more computational resources and time. Applying a Kalman Filter for closed-loop correction of torque estimation provides more accurate torque estimates and real-time updates. However, it demanded a high level of model completeness, had high computational complexity, and requires experience and tuning for selecting filter parameters.

To address overfitting and lack of transparency in neural networks for predicting indicated thermal efficiency, this paper proposes using the random forest algorithm for prediction. The random forest algorithm is a powerful ensemble learning method where the regression prediction result is the simple average of multiple learners. The decision tree nodes are determined by formulas, giving the method high prediction accuracy, strong generalization ability, high interpretability of internal computations, and excellent capability to capture nonlinear relationships. Compared to optimization modeling methods, the random forest algorithm can improve model accuracy without increasing computational complexity. Additionally, it requires fewer parameter adjustments and simplifies the adjustment process.

Using the random forest algorithm, the indicated thermal efficiency was calculated with universal characteristic data obtained from experiments [7]. A random forest regression model was established using Matlab/Simulink. First, the performance of the random forest indicated thermal efficiency prediction model was evaluated. Next, the random forest regression model was used to validate the diesel engine's universal characteristics. Finally, the random forest indicated thermal efficiency prediction model replaced the MAP interpolation module in the original diesel engine mean value engine model for simulation verification.

Material and methods

Based on the random forest algorithm [8], the indicated thermal efficiency was fitted as shown in (Figure 1). First, the indicated thermal efficiency was calculated, and the resulting data was then normalized. During model training, hyperparameters were determined based on the error. The model stopped training when the number of samples was less than the number per leaf, and no pruning was performed on the tree models within the random forest, allowing each tree to grow fully. After the model training was completed, it was saved into the mean value engine model, replacing the original indicated thermal efficiency calculation module. The inputs of this module were speed and power, and the output was the indicated thermal efficiency.

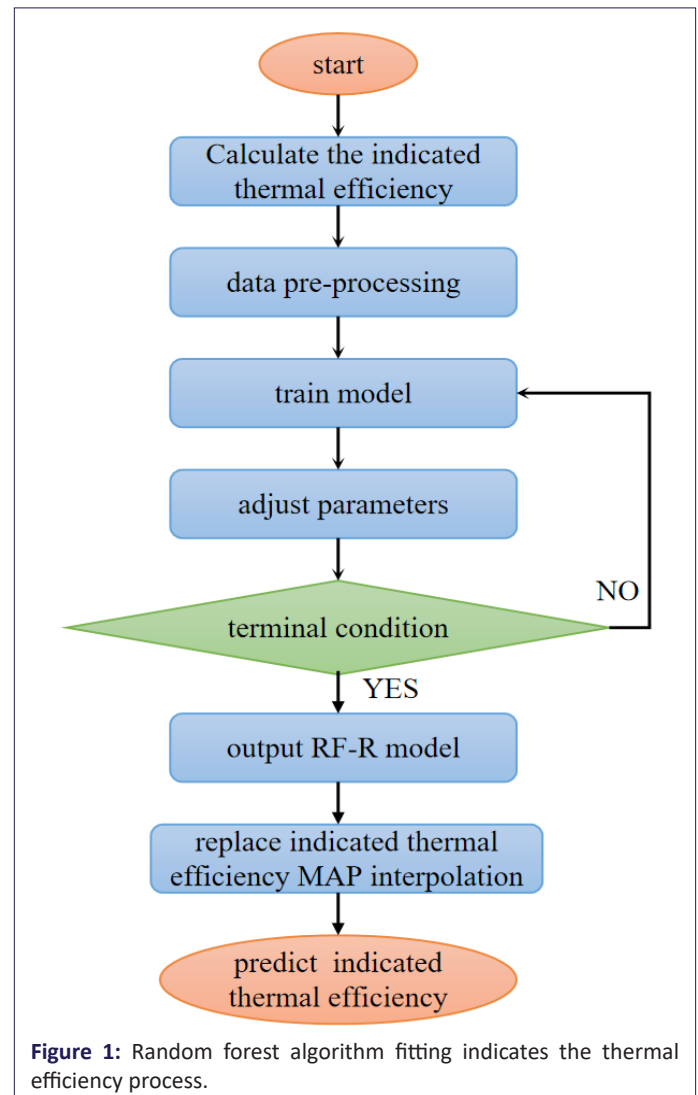


Figure 1: Random forest algorithm fitting indicates the thermal efficiency process.

Random forest algorithm

The Random Forest (RF) algorithm, proposed by Leo Breiman and other scientists, was a machine learning ensemble algorithm based on the bagging (Bootstrap Aggregating) concept [9-14]. Compared to bagging, Random Forest introduces feature randomness, significantly improving prediction accuracy without notably increasing computational load during large-scale data operations. Random Forest had a very strong learning capability, suitable for both classification and regression tasks. This paper employed a Random Forest regression model, which introduced feature randomness on top of bagging when handling regression tasks. Therefore, the Random Forest regression model had low sensitivity to outliers and missing values in the dataset and was less prone to overfitting. It was now widely used in various fields such as economics, medicine, and statistics.

The process of the Random Forest regression model [15,16] could be roughly described as follows: First, bootstrap resampling was used to extract subsets from the sample set. For a given sample set with M data points, M subsets were obtained through M resampling iterations with replacement, resulting in data subsets each containing M data points. The probability of any given data points not being selected in the sample set was $\lim_{M \rightarrow \infty} \left(1 - \frac{1}{M}\right)^M \approx \frac{1}{e} \approx 0.368$ (1)

$$\lim_{M \rightarrow \infty} \left(1 - \frac{1}{M}\right)^M \approx \frac{1}{e} \approx 0.368 \quad (1)$$

In this formula, e is the base of the natural logarithm.

According to the formula (1), the probability that a given sample did not participate in constructing a decision tree was approximately 36.8%. This meant that 36.8% of the samples in the dataset did not contribute to constructing the decision tree; this portion of the data was known as Out-of-Bag (OOB) data. OOB data could be used to test the model's generalization performance and tuned its hyperparameters. Evaluating the model's performance with OOB data was referred to as OOB estimation. For each decision tree, an OOB error estimation could be obtained. By summing the errors of all decision trees in the forest and then averaging them, the generalized error of the random forest (PE*) was obtained. Breiman demonstrated through experiments that OOB estimation was an unbiased estimate [17-19]. Compared to cross-validation, OOB estimation reduced the computational burden of the algorithm, improved efficiency, and produced validation results that were approximately equivalent to those of cross-validation.

The extracted M data subsets were used to train M decision trees. In the Random Forest algorithm, the default type of decision tree [20] was the CART (Classification and Regression Tree) regression tree. CART regression trees started growing from the root node, each tree grew fully without requiring pruning, and growth stopped when the tree met its termination conditions. These conditions typically included:

When the number of data points at the current node was less than a specified threshold.

When the mean squared error (MSE) at the current node was less than a predefined threshold, indicating no further split was necessary.

When the tree reached a specified depth limit.

During the growth process from root to leaf nodes, feature selection at each node was determined by calculating its mean

squared error (MSE). The optimal node for splitting was selected by minimizing the MSE, as shown in the following formula:

$$m = \min_{A,s} \left[\min_{c_1} \sum_{x_j \in D_1(A,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_j \in D_2(A,s)} (y_i - c_2)^2 \right] \quad (2)$$

In the formula: s represents the entire training dataset at the current node. A represents the subset of features selected for the current node.

The training set was partitioned into subsets D1 and D2 based on feature A. By iterating over the values of A, the sum of the minimum mean squared errors of the output values y for subsets D1 and D2 was calculated. Eventually, by iterating through all attributes, the minimum Mean Squared Error (MSE) was determined along with the corresponding attribute and its value, providing growth information for that node. This process was repeated for each generated child node until the termination conditions were met.

After training the model, when using Random Forest for regression prediction, the final prediction result was the arithmetic average of the predicted values from each decision tree. The workflow of the Random Forest algorithm is illustrated in (Figure 2).

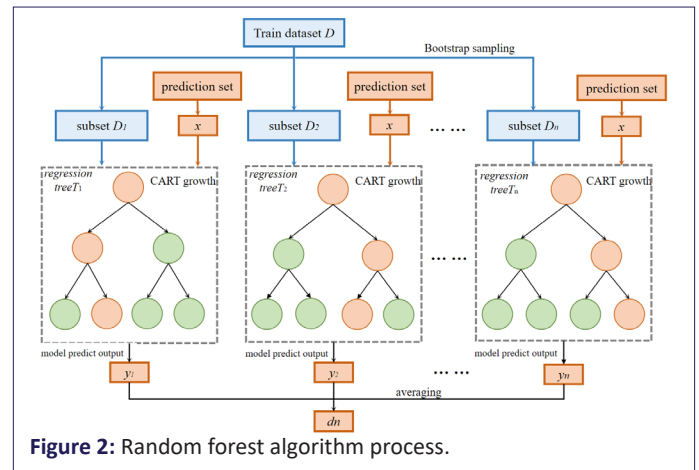


Figure 2: Random forest algorithm process.

Construction and evaluation of the indicated thermal efficiency random forest prediction model

Calculation and preprocessing of indicated thermal efficiency: The indicated thermal efficiency of an engine could not be directly obtained from experimental data. However, the universal characteristic curve of the engine provided performance parameters such as speed, effective power, and effective fuel consumption rate. Therefore, it was necessary to extract parameters from the universal characteristic curve of the engine to calculate the indicated thermal efficiency. For a diesel engine, given the indicated power and the hourly fuel consumption, the formula for calculating the indicated thermal efficiency could be derived based on its definition, as shown in Equation (3):

$$\eta_i = \frac{3600(P_e + P_f)}{G_T H_u} \quad (3)$$

In this formula: η_i the indicated thermal efficiency; P_e is the effective power, obtained from the universal characteristic curve derived from experiments; P_f is the mechanical loss power, obtained from the mechanical loss model, which will be introduced in the following model; 3600 represents the energy equivalent of 1 kWh; H_u is the lower heating value of the fuel

used. For ship engines, the commonly used fuel is heavy oil, and the standard lower heating value for heavy oil in China is 42,000 kJ/kg when calculating fuel consumption rate.

Moreover, the effective fuel consumption rate could be described as $g_e = G_T / P_i$

the following equation could be derived (4):

$$\eta_i = \frac{3600(P_e + P_f)}{g_e P_e H_u} \quad (4)$$

After calculating the indicated thermal efficiency, the data is normalized according to Equation (5).

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (5)$$

Parameter selection

In the process of using the random forest regression model to predict indicated thermal efficiency, it was first necessary to determine the model parameters. In a random forest regression model, parameters that needed to be set included the minimum number of leaf nodes (minleaf) and the number of decision trees (trees). The more decision trees, the more accurate the model prediction, but the computational load significantly increased. Therefore, in this study, the Out-of-Bag (OOB) data was used for estimation. The model parameters were determined based on the OOB error estimation. The curve of OOB error estimation as a function of the number of decision trees was shown in (Figure 3). The curve of OOB error estimation as a function of the number of leaf nodes was shown in (Figure 4).

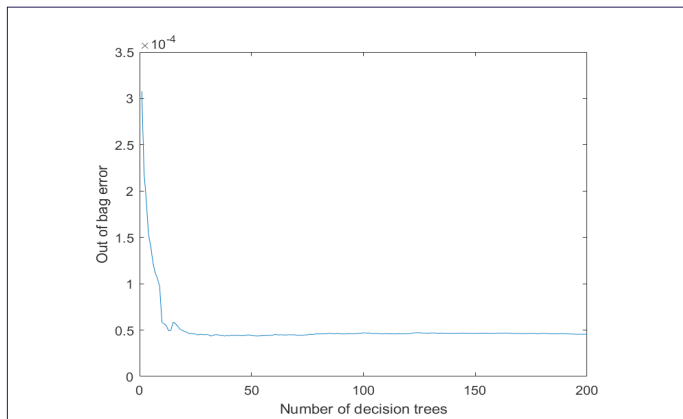


Figure 3: Variation curve of out-of-bag error with the number of decision trees.

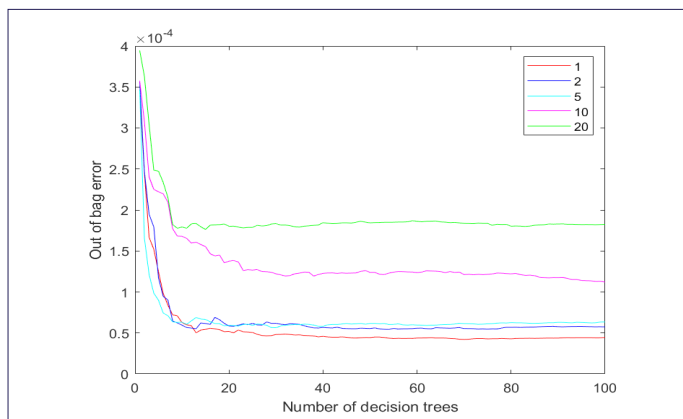


Figure 4: Effect of minleaf node number on the out-of-bag error.

From (Figure 3), it was observed that the error decreased sharply at first and then stabilized when the number of decision trees reached 80. Therefore, to maintain computational efficiency, the number of decision trees was set to 100. From (Figure 4), it was seen that the final error was minimized when the minimum number of leaf nodes was 1. Thus, the parameter minleaf was set to 1.

RF-R evaluations and analysis

To evaluate the performance of the random forest regression algorithm for predicting indicated thermal efficiency, three metrics were selected: Root Mean Square Error (RMSE), correlation coefficient R-squared (R²), and Mean Absolute Error (MAE). RMSE and MAE were used to measure the error between the predicted values and the actual values, while R² was used to assess the correlation between the predicted values and the actual values. The formulas were as follows.

$$RMSE = \sqrt{\sum_{i=1}^n (h_i - y_i)^2 / n} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (h_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

$$MAE = \left(\sum_{i=1}^n \left| 1 - \frac{h_i}{y_i} \right| \right) / n \quad (8)$$

In this formula: h_i is predicted thermal efficiency indicator from model output. y_i is measured thermal efficiency indicator. n is the logarithm of predicted and observed values. \bar{y} is the average value of observed indicated thermal efficiency.

The data was divided into training and testing sets in an 8:2 ratio. Using Out-of-Bag (OOB) estimation, the model's predicted and observed values of indicated thermal efficiency were shown in (Figure 5). Based on the evaluation metrics, the training and testing sets exhibit mean squared errors of 0.0056 and 0.0059 respectively, mean absolute errors of 0.0038 and 0.0043 respectively, and R² values of 0.9684 and 0.9612 respectively. These results indicated that the model demonstrated good generalization performance without overfitting. According to these assessment results, the model exhibited high predictive accu-

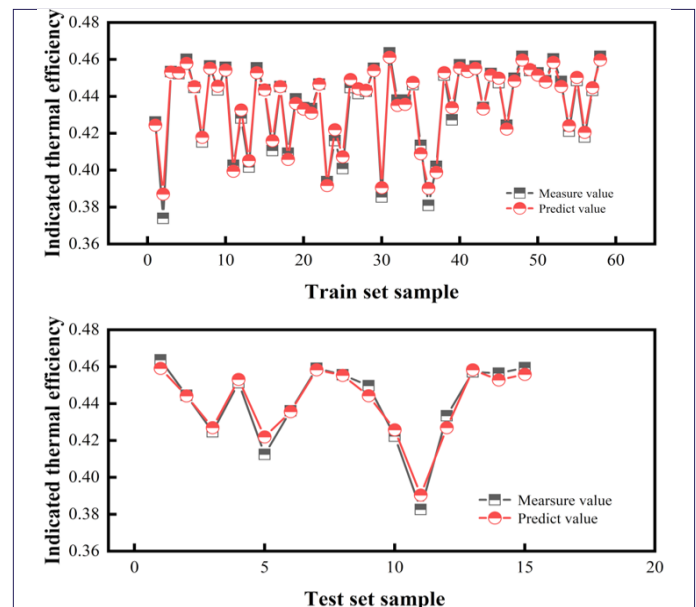


Figure 5: The result of the predicted value is compared with the measured value.

accuracy and strong correlation between observed and predicted values, accurately forecasting the indicated thermal efficiency of a full-load diesel engine. This performance was primarily attributed to the robust learning capability of the random forest algorithm, which effectively captured nonlinear relationships between different features.

Modules related to the mean value engine model

The mean value engine model [21-24] was a zero-dimensional mechanism model, known for its fast calculation speed and accurate external characteristic calculations. It had been widely used in the dynamic simulation of diesel engines. The mean value engine model combined the characteristics of the quasi-static model and the volume method model. Referring to the volume method model, the diesel engine was divided into several relatively independent modules. The working principle was shown in (Figure 6). The modules were linked by parameters such as pressure, flow, and temperature to form the overall diesel engine model. The mean value model simplified the in-cylinder process of the diesel engine. It treated the in-cylinder process as stable, ignoring fluctuations in temperature and pressure. The first law of thermodynamics was directly used to calculate the indicated power of the diesel engine based on indicated thermal efficiency, thereby calculating the average indicated torque [25,26].

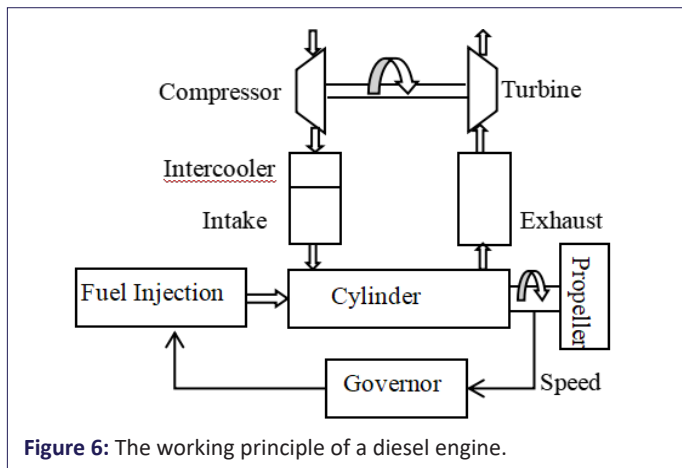


Figure 6: The working principle of a diesel engine.

The governor adjusted the amount of circulating fuel injected into the cylinder according to the speed feedback. After the oil mist of the injection cylinder is fully mixed with the air, it exploded near the top dead center of compression, pushed the piston to do work, and converted the chemical energy of the fuel into mechanical energy with a certain thermal efficiency. The average indicated torque generated by the diesel engine could be calculated by the following formula (9):

$$T_i = \frac{30}{\pi} \frac{P_i}{n_e} = \frac{30}{\pi} \frac{\eta_i H_u m_f}{n_e} \quad (9)$$

In this formula: m_f is the average mass flow rate of fuel, H_u is the low calorific value of fuel, η_i is the indicated thermal efficiency, n_e is the speed of diesel engine, P_i is the indicated power;

This paper only introduced the calculation method of related sub-models, and the rest can be referred to.

Indicated thermal predict model

In this article, the indicated thermal efficiency was considered as a function of rotational speed and power, as shown in formula (10):

$$\eta_i = f(n_e, P) \quad (10)$$

The trained random forest regression model was encapsulated in the module, and the indicated thermal efficiency MAP calculation module was replaced. The indicated thermal efficiency was predicted according to the input speed and power, and output to the next sub-module. The indicated thermal efficiency was no longer calculated by the MAP interpolation of excess air coefficient and speed, but by inputting speed and power, which was predicted by random forest regression.

Mechanical loss model

The heat loss power during the diesel engine cycle could not be ignored. Mechanical loss was considered the primary loss power of the diesel engine and was crucial for calculating its indicated thermal efficiency. However, the mechanical loss power could not be obtained by direct measurement. Therefore, this paper obtained it through the mechanical loss torque calculation sub-model in the diesel engine mean value model. The research object of this paper was a four-stroke turbocharged diesel engine, and its mechanical loss pressure could be expressed as:

$$p_f = c_0 + c_1 n_e + c_2 n_e^2 \quad (11)$$

In this formula: P_f is mechanical loss pressure; n_e is engine speed; C_1, C_2, C_3 is parameter related to the specific diesel engine;

The mechanical loss torque generated by the mechanical loss pressure was calculated according to the same mode of converting the in-cylinder gas force into the output torque. The calculation formula is:

$$T_f = F_g R \left[\sin\left(\frac{n\pi}{30}\right)t + \frac{\lambda \sin 2\left(\frac{n\pi}{30}\right)t}{2\sqrt{1-\lambda^2 \sin^2\left(\frac{n\pi}{30}\right)t}} \right] \quad (12)$$

In the formula: F_g the mechanical loss pressure is the pressure on the piston according to the gas force mode; R the crank radius; λ for the crank connecting rod ratio;

The mechanical loss torque was obtained by the mechanical loss module according to formula (12), and then the mechanical loss power was obtained according to formula (13).

$$P_f = \frac{\pi \cdot n_e \cdot T_f}{30} \quad (13)$$

In the formula: P_f is mechanical loss power; T_f is mechanical loss torque;

Table 1: Diesel engine specifications.

Parameters	Value
Bore *stroke (mm)	170*185
Cylinder Center Distance (mm)	214
Compression Ratio	14.5:1
Maximum Combustion Pressure /MPa	17.5
Rated Power/Speed (kW/ r/min)	647/1500
Idle Speed (r/min)	900±20
Maximum Exhaust Temperature Before Turbocharger (°C)	≤700
Fuel Consumption Rate at Rated Condition (g/kW·h)	≤210
Overload Total Power/Speed (kW/ r/min)	800/1500

Results and discussion

The simulation model of the diesel engine in this study is built based on a turbocharged diesel engine. The main technical parameters are shown in (Table 1).

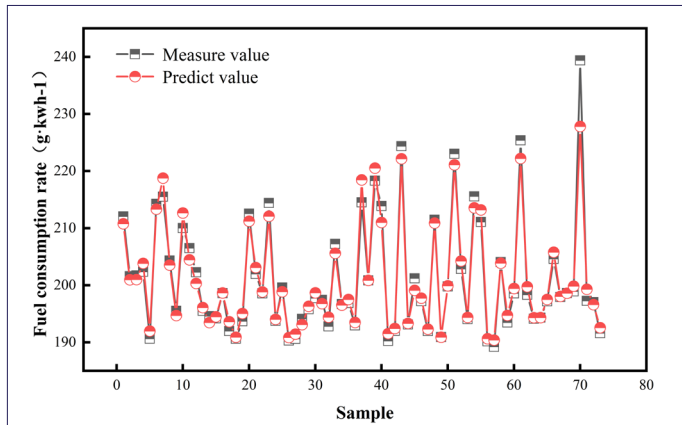


Figure 7: Comparative analysis of actual and projected fuel consumption rate in a line graph.

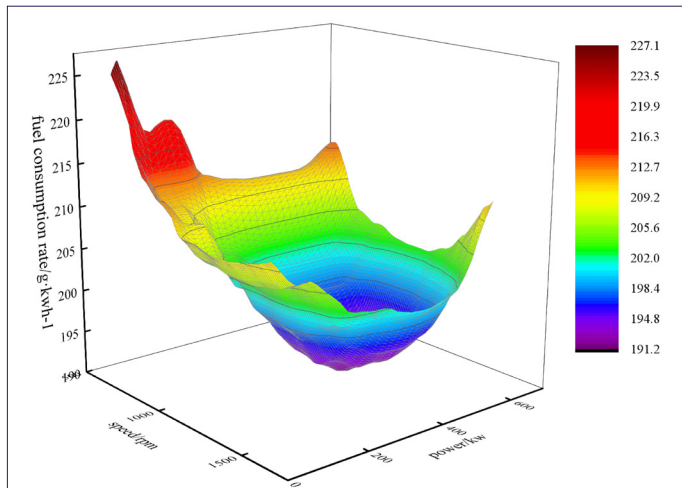


Figure 8: Three-dimensional surface plot of fuel consumption rate.

Fuel consumption rate experiment verification

Based on the random forest regression model, the fuel consumption rate [27-30] prediction model was established.

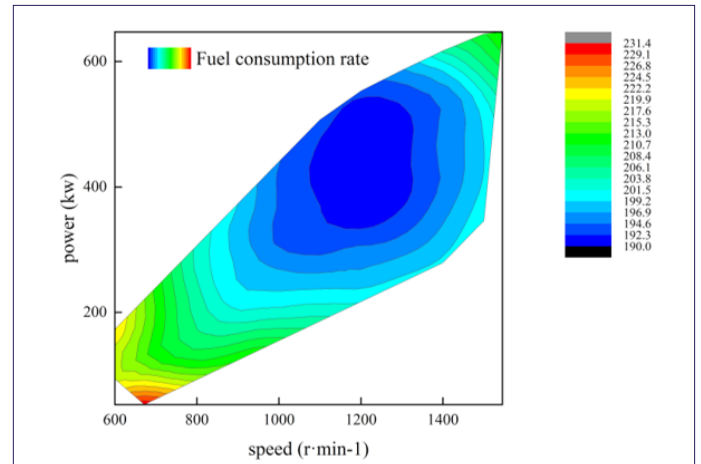
The fuel consumption rate of each speed test point was predicted, and the measured values were compared. The measured values and the random forest predicted values were compared.

The line was shown in (Figure 7). The fuel consumption rate data across the full experiment speed and power range were predicted, and three-dimensional surface and two-dimensional contour diagrams of the fuel consumption rate were drawn. (Figure 8) showed the three-dimensional surface diagram of the universal characteristics, and (Figure 9) showed the two-dimensional contour comparison diagram of the universal characteristics. The (Figures 7,8,9) showed that the universal characteristics predicted by the random forest closely matched the measured universal characteristics curve.

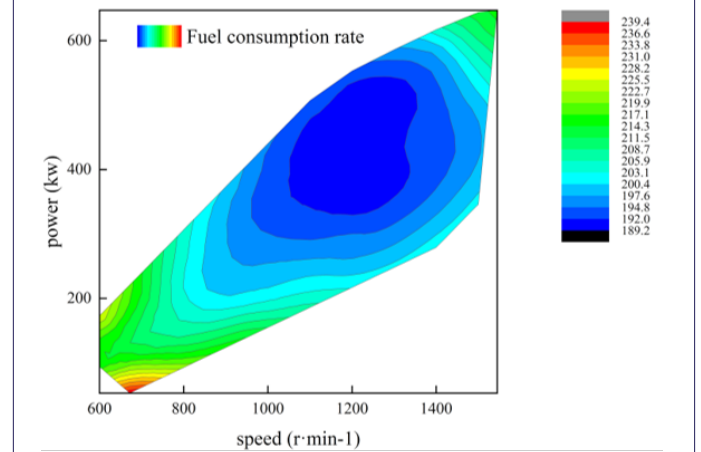
Simulation model verification

The trained random forest prediction model of indicated thermal efficiency was implemented in the

Matlab function module of Matlab/Simulink, replacing the original data interpolation calculation module. Based on the



(a) Random forest predicts universal characteristic curve



(b) The measured data plots the universal characteristic curve

Figure 9: Comparison between the fitted universal characteristic curve using random forest regression and the measured data.

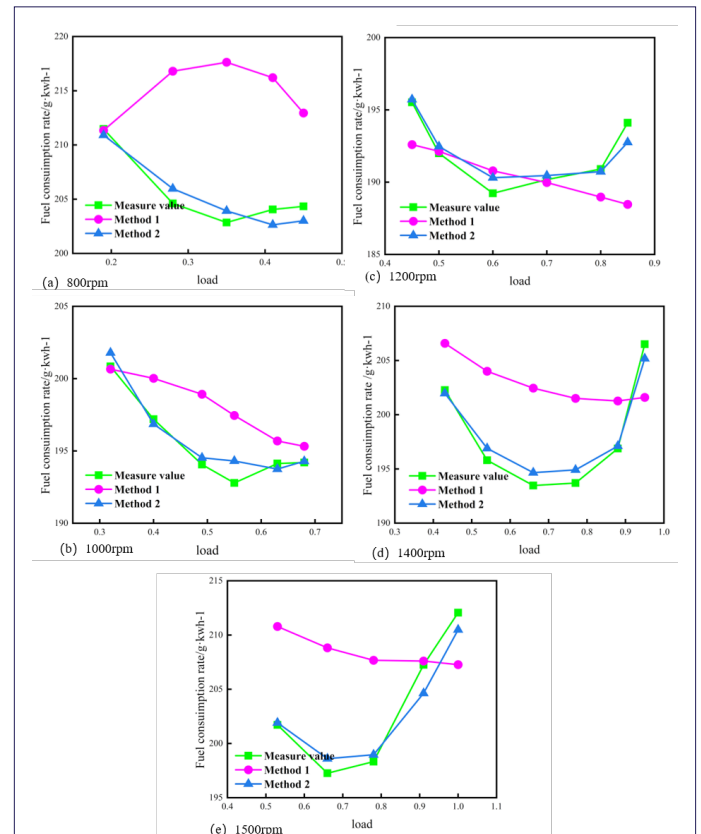


Figure 10: Simulation comparison under different working conditions.

load characteristics, fuel consumption rate data under different working conditions were selected for comparison of simulation results. To test the performance of the random forest algorithm across the full operating range, simulation tests were conducted at operating loads of 800 rpm, 1000 rpm, 1200 rpm, 1400 rpm, and 1500 rpm. The simulation results were shown in (Figure 10). In the (Figure 10), the error between the simulated and the measured value of the fuel consumption rate under different working conditions was compared. Method 1 used the MAP diagram to calculate the MVEM fuel consumption rate simulation result of the indicated thermal efficiency, while Method 2 used the random forest algorithm to predict the MVEM fuel consumption rate simulation result of the indicated thermal efficiency.

It can be seen from (Figure 10) that the fuel consumption rate calculated by the MAP diagram matched the measured rate only at low loads. At 800 rpm, it do not match the measured data at all. At 1400 rpm, the overall trend of the fuel consumption rate calculated by the MAP diagram is roughly the same as the measured value. However, the sensitivity to changes in the fuel consumption rate is insufficient, leading to a relatively large error.

Conclusion

In this paper, according to the universal characteristic data of diesel engine, the random forest prediction model of indicated thermal efficiency of diesel engine is successfully constructed. The model is embedded into the mean value engine model of real-time simulation, and the following conclusions are drawn:

1) The R^2 values of the training set and test set of the random forest indicated thermal efficiency prediction model are 0.9684 and 0.9612, respectively, indicating excellent fitting and strong correlation. In the low-speed and low-load area, the simulation error of the MVEM fuel consumption rate using the MAP diagram is slightly higher than 5%, reaching a maximum of 5.96%. The simulation error of the MVEM fuel consumption rate using the random forest algorithm to predict the indicated thermal efficiency is almost less than 5%, with a maximum of -1.45%. The remaining simulation values are closer to the measured values, indicating that the overall performance of the random forest algorithm is good. This demonstrates that using the random forest algorithm to improve the simulation accuracy of the model is feasible.

The prediction accuracy of fuel consumption rates at different loads and speeds using the MVEM calculated by the random forest is nearly higher than that using the MAP diagram. This method has the advantage of improving high prediction accuracy across the full load range.

References

- Zhou Dong, Su Tiexiong, Li Chuang, et al. Simulation accuracy analysis of intake manifold model based on engine mean value model J. Science, Technology and Engineering. 2016; 16(32): 238-242.
- Zhou Haitao, Yan Ping, Wang Xinquan. Modeling of mean value model for electronically controlled diesel engine J. Diesel engine. 2010; 32 (02): 12-17.
- Yang Chuanlei, Zhang Wenle, Wang Hechun, et al. Modeling and simulation of mean value model of marine diesel engine J. Applied Science and Technology. 2024; 51(01): 105-111.
- Wang Haiyan, Zhang Jundong, Ren Guang. Linear Parameter Varying Model of Turbocharged Diesel Engine J. Journal of Harbin Engineering University. 2007; (09): 966-969,979.
- Tang Yuanyuan, Zhang Jundong, Jia Baozhu, et al. The influence of cylinder block on the accuracy of marine diesel engine model J. Journal of Harbin Engineering University. 2017; 38(12): 1836-1843.
- Ge Yanwu, Huang Ying, Su Guoqing, et al. Research on indicated torque estimation of diesel engine based on mean value model and Kalman filter C. Chinese Society of Automotive Engineering
- Sun Haowen, Sui Mingzheng, Dong Haiguo. Engine characteristic curve analysis and suggestion J. Equipment management and maintenance. 2022; (07): 139-142. DOI: 10.16621/j.cnki.issn1001-0599.2022.04.59.
- lv Hongyan, Feng Qian. Research review of random forest algorithm J. Journal of Hebei Academy of Sciences. 2019; 36(03): 37-41. DOI: 10.16191/j.cnki.hbxx.2019.03.005.
- Amirali Shateri, Zhiyin Yang, Jianfei Xie. Utilizing Artificial intelligence to identify an Optimal Machine learning model for predicting fuel consumption in Diesel engines J. Energy and Ai. 2024; 16: 100360.
- Mohammad A, Rezaei R, Hayduk C, et al. Physical-oriented and machine learning-based emission modeling in a diesel compression ignition engine: Dimensionality reduction and regression J. International Journal of Engine Research. 2023; 24(3): 904-918.
- Karunamurthy K, Janvekar A A, Palaniappan P L, et al. Prediction of IC engine performance and emission parameters using machine learning: A review J. Journal of Thermal Analysis and Calorimetry. 2023; 148(9): 3155-3177.
- Li Y, Jia M, Han X, et al. Towards a comprehensive optimization of engine efficiency and emissions by coupling Artificial Neural Network (ANN) with genetic algorithm (GA) J. Energy. 2021; 225: 120331.
- Yazar I, Anagun Y, Isik S. Predicting compressor mass flow rate using various machine learning approaches J. International Journal of Turbo & Jet-Engines. 2024.
- Ma J. A Comparative Study on Fuel Consumption Prediction Methods of Heavy-Duty Diesel Trucks Considering 21 Influencing Factors J. Energies. 2021; 14. DOI:10.3390/en14238106.
- Zengyou He, Jiaqi Wang, Mudi Jiang, et al. Random subsequence forests J. Information Sciences. 2024; 667: 120478.
- Junyu Guo, Xueping Zan, Lin Wang, et al. A random forest regression with Bayesian optimization-based method for fatigue strength prediction of ferrous alloys J. Engineering Fracture Mechanics. 2023; 293: 109714.
- Tri Wahyuningsih, Danny Manongga, Irwan Sembiring, et al. Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments J. Procedia Computer Science. 2024; 234: 349-356.
- Paulino José García Nieto, Esperanza García Gonzalo, Luis Alfonso Menéndez García, et al. Predicting the critical superconducting temperature using the random forest, MLP neural network, M5 model tree and multivariate linear regression J. Alexandria Engineering Journal. 2024; 86: 144-156.
- Chen, Zhou, Tang. Study on spatial interpolation of precipitation in Jingzhou City based on random forest algorithm J. Water conservancy technical supervision. 2024; (03): 199-202.
- Zechun Huang, Zipu Liu. A Complex Terrain Simulation Approach Using Ensemble Learning of Random Forest Regression J. Jour-

- nal of the Indian Society of Remote Sensing. 2022; 50: 2011-2023.
21. He Lidong, Wang Haiyan, Liu Sen. Summary of modeling methods for marine diesel engines J. Internal combustion engines and power plants. 2013; 30(06): 53-55.
 22. Zhang D, Shen Z, Xu N, et al. Development of a zero-dimensional model for a low-speed two-stroke marine diesel engine with exhaust gas bypass and performance evaluation J. Processes. 2023; 11(3): 936.
 23. Tang, Yuanyuan, Zhang J, et al. The accuracy of the mean value engine model affected little by the accuracy of compressor mass flow rate model for a marine diesel engine J. Journal of Marine Science and Technology. 2024. 29(1): 1-19.
 24. Hautala S, Mikulski M, Söderäng E, et al. Toward a digital twin of a mid-speed marine engine: From detailed 1D engine model to real-time implementation on a target platform J. International Journal of Engine Research. 2023; 24(12): 4553-4571.
 25. Yang Wei, Li Ying, Bai Sichun, et al. Research on torque control-based diesel engine oil-gas coordinated control strategy J. Internal combustion engine. 2023; 39 (3): 47-51.
 26. Kharroubi K, Söğüt O S. Modelling of the propulsion plant of a large container ship by the partition of the cycle of the main diesel engine J. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment. 2020; 234(2): 475-489.
 27. Ruan Z, Huang L, Wang K, et al. A novel prediction method of fuel consumption for wing-diesel hybrid vessels based on feature construction J. Energy. 2024; 286. DOI:10.1016/j.energy.2023.129516.
 28. Marques C H, Caprace J D, Belchior C R P, et al. An approach for predicting the specific fuel consumption of dual-fuel two-stroke marine engines J. Journal of Marine Science and Engineering. 2019; 7(2): 20.
 29. Gad M S, Alenany A. Comparative diesel engine performance and emission forecasting using extreme learning and quadratic regression techniques burning waste cooking biodiesel J. International Journal of Hydrogen Energy. 2024; 60: 94-106.
 30. Chen Y, Sun B, Xie X, et al. Short-term forecasting for ship fuel consumption based on deep learning J. Ocean Engineering. 2024; 301. DOI:10.1016/j.oceaneng.2024.117398.