Fuel consumption prediction for the tractors augmented with GNSS recordings

Xiang Luo[1,2], Caicong Wu[1], Ying Chen[1*], Liwei Pan[3]

(1. *College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China*;
2. *Institute of Agricultural Engineering, Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China*;
3. *Inspur Group Company Limited, Jinan 250000, China*)

Abstract: With the rapid advancement of agricultural mechanization, fuel consumption in tractors has increased, leading to higher production costs. Selecting tractors with lower fuel consumption requires a method to evaluate fuel consumption in practical scenarios. In this study, we constructed a fuel consumption dataset based on a number of tractors operating in various regions in China, and we proposed a novel approach that can automatically predict the instant fuel flow rate of tractors. Firstly, we collected a large-scale fuel consumption dataset recorded by GNSS (Global Navigation Satellite System) and ECU (Electronic Control Unit) devices installed on 90 tractors with three tractor models. Then, we analyzed the interactions among a number of factors involved in their fuel consumption, and characterized each data point with the four parameters: two engine-based parameters (torque and engine speed) and two motion-based parameters (driving speed and acceleration). Based on the four parameters and a powerful machine learning method (Random Forest), we developed a fuel consumption prediction model, which predicts the fuel flow rate at each point. Finally, we made intensive experiments, which demonstrate that the proposed method achieves state-of-the-art performances in predicting fuel flow rate, yielding at least an average R^2 value of 0.88 on the three tractor models. Moreover, an in-depth analysis was made to examine the accuracy and transfer capability of the developed prediction models.

Keywords: fuel consumption, tractors, GNSS, fuel flow rate, ECU
DOI: 10.25165/j.ijabe.20261904.10646

Citation: Luo X, Wu C C, Chen Y, Pan L W. Fuel consumption prediction for the tractors augmented with GNSS recordings. Int J Agric & Biol Eng, 2026; 19(4): 184–190.

1 Introduction

Fuel consumption evaluation plays a crucial role in the agricultural sector, specifically in aiding operators in the selection of tractors that meet their needs while ensuring cost-effectiveness and energy efficiency. Tractors serve as the primary power source for a wide range of implements and agricultural machinery, making them essential in various farming operations[1,2]. Their extensive use in agriculture highlights the need for efficient fuel management, as fuel costs directly impact the overall economic performance of agricultural activities. Previous research emphasized that fuel consumption is the largest operational expense when using tractors, surpassing other factors such as machinery depreciation and labor costs[3]. This is largely due to the high energy demands required for these machines to perform their diverse operations across different terrains and under varying load conditions. Thus, a comprehensive fuel consumption evaluation for tractors is a key factor in the ongoing efforts to make agricultural practices more economically viable[4].

Since the amount of fuel consumed is dependent on the tractor's working condition, the fuel consumption evaluation should be conducted under such a practical scenario[5,6]. In the literature, various fuel consumption modeling methods have been proposed to point out factors involved in the practical fuel consumption of tractors. The most crucial factors are from the tractor configuration. Grisso et al.[7] have developed new equations to predict fuel consumption under different load conditions and engine speeds. Besides tractor configuration, implement specification usually plays an important role in fuel consumption[8]. For example, Damanauskas et al.[9] and Ekemube et al.[10] showed that the harrowing implement setting (e.g., disc angle, tillage depth) and the choice of working speed influence fuel consumption. Md-Tahir et al.[11] found that different tire types have a direct impact on fuel consumption during field operations. The study of Al-Sager et al.[12] showed that implement width significantly affected the fuel consumption. Moreover, previous studies also showed that operational behavior and the field environment often influence fuel consumption[13]. For example, Moinfar et al.[14], Tihanov and Ivanov[15], and Siddique et al.[13] showed the relationship between operation behavior (e.g., driving system, operation mode) and fuel consumption. Janulevičius et al.[16] revealed that in smaller fields, the time spent on turning operations accounted for a substantial portion, leading to a direct increase in fuel consumption. Battiato et al.[8] pointed out that the fuel consumption on different soils is influenced by various factors, including soil type, moisture content, slip rate, and tire pressure.

The following three problems exist in current fuel consumption modeling studies: 1) In order to examine only the considered factors, most of the previous fuel consumption modeling methods and their experiments were well-designed and tested only on one tractor in a field, leading to a lack of a comprehensive fuel

Received date: 2026-05-12 **Accepted date:** 2026-07-10
Biographies: Xiang Luo, PhD candidate, research interest: agricultural engineering, Email: 15152190986@163.com; **Caicong Wu**, PhD, Professor, research interest: agricultural machinery unmanned driving and autonomous operation, Email: wucc@cau.edu.cn; **Liwei Pan**, MEng, research interest: agricultural big data, Email: panliwei8886@163.com.
***Corresponding author: Ying Chen**, PhD, Associate Professor, research interest: agricultural big data mining. China Agricultural University, No.17 Qinghua East Road, Haidian District, Beijing 100083, China. Tel: +86-13520292799, Email: chenying@cau.edu.eu.

consumption evaluation for a tractor model. 2) Fuel consumption of a tractor is an overall result of many factors, but it is high-cost to collect information on all factors, particularly for experiments conducted in a wide range of regions. So, it is better to analyze the interaction of the factors and collect the information about some key factors using advanced sensors. 3) There is not much research on using machine learning for predicting agricultural machinery fuel consumption[12,17]. However, there is a lack of a comprehensive analysis of model performance, causing the inconvenience of the model's usage. To address the three problems, this study constructed large-scale fuel consumption data that provides tractor fuel consumption in a practical scenario, and proposed an automatic fuel consumption prediction method that uses four factors and a Random Forest (RF) to develop an instant fuel flow rate prediction model for the considered tractor model[18-20]. In the fuel consumption data, there are large-scale recordings collected during the operation of 90 tractors of three tractor models, and each point (or sample) in the recordings contains three types of information: the tractor's instant fuel flow rate, the engine working conditions, and the tractor's motion characteristics. As the recordings contain various operations performed across several provinces in China, they can represent the practical scenarios of the 90 tractors. On the other hand, although the practical scenario used in the data collection does not set a limitation on factors involved in tractor fuel consumption (e.g., tractor number, implement type, operation mode, and field shape), this study uses only two engine-based parameters (torque and engine speed) from ECU (Electronic Control Unit), which are fundamental indices of tractor configuration[7,12], and two motion-based parameters (driving speed and acceleration) from GNSS (Global Navigation Satellite System) devices, which are crucial indices reflecting tractor operation behavior. Our factor analysis has shown that the engine-based and motion-based parameters contain crucial information for fuel consumption prediction. Moreover, a comprehensive analysis has been made, which examines several factors (e.g., input parameters, data settings) influencing model performance and the model's transferring capability.

The goals of the study are to: 1) conduct a comprehensive fuel consumption evaluation for various tractor models by designing and executing experiments that incorporate multiple tractors operating under diverse conditions; 2) apply machine learning techniques for fuel consumption prediction and conduct a thorough analysis of model performance so as to evaluate various machine learning algorithms, compare their effectiveness, and provide insights into their applicability; and 3) develop a fuel consumption prediction model for the considered tractor model so as to evaluate its energy efficiency in practice. To our best knowledge, it is the first attempt to examine the comprehensive fuel consumption on different tractor models in practical scenarios. The main contributions of this study include: 1) This study systematically constructs a large-scale dataset of fuel consumption for three tractor models, addressing a gap in the existing literature regarding fuel consumption under varying operational conditions and in different regions. This reliable dataset serves as a foundational resource for future research. 2) The development of the fuel consumption prediction model in this study simplifies the influencing factors, condensing them into four key parameters. This theoretical framework not only advances the application of machine learning in the field of agricultural machinery but also provides new insights for model construction in related fields. 3) Through rigorous experimental analysis, the study validates the effectiveness and transfer capability of the developed models under varying conditions. This theoretical contribution offers a basis for the application of machine learning models in other types of agricultural equipment or diverse operational environments. 4) The fuel consumption data and prediction model provided by this study enable tractor operators to better understand and anticipate fuel consumption, allowing them to optimize fuel usage in actual operations, reduce operational costs, and enhance economic efficiency. In summary, this study holds significant contributions in both theoretical and practical domains, advancing research in the field of agricultural machinery and its real-world applications.

2 Materials and methods

2.1 Data

The experiment in this study was conducted on fuel consumption data, which were constructed using tractors from the same manufacturer (Jiangsu World Agricultural Machinery Co., Ltd) but chosen to cover two powers (66.2 kW and 180 kW). In general, there are two steps in the dataset construction.

Data collection: In the Precision Agriculture Application Project Data Service Platform, it was found that some tractor models' data records contained engine parameters (torque percentage, engine speed) and instantaneous fuel consumption fields. When the platform detected missing fields, the system would automatically fill them with default values. Through statistical analysis of abnormal situations, we found that the proportion of abnormal situations in the data records of three tractor models (WD904, WD904G, and WH1804) was relatively low, which could provide a more reliable analysis basis. These three tractor models with different horse powers were selected, where both WD904 and WD904G use the same type of engine (low-powered, 66.2 kW), and WH1804 is high-powered (180 kW). For each tractor model, 30 tractors were randomly selected. Then, for each tractor, its recordings collected between May 2023 and June 2023 were retrieved from the Precision Agriculture Application Project Data Service Platform. Tractors, the primary power source for various implements, are usually intensively used between May and June (a wheat harvesting season in China), and the platform receives the recordings of agricultural machinery in real time from multiple agricultural machinery manufacturers in China[21].

As listed in Table 1, for each record (i.e., a point or a sample), there are eight recorded parameters: timestamp (YYYYMMDDhhmmss), longitude (WGS84), latitude (WGS84), speed (m/s), direction (°), torque percentage (instantaneous torque/nominal torque, %), engine speed (r/min), and fuel flow rate (L/h). The first five parameters are provided by GNSS (Global Navigation Satellite System) devices with a collection accuracy of 5 meters (CEP), and the latter three parameters are collected from ECU devices. Both devices have been installed on tractors, and their update frequency is 5 s.

Data pre-processing: As shown in Figure 1, to provide high-quality data, two kinds of data pre-processing were applied: data cleaning and field-road trajectory segmentation. Firstly, in order to provide clean input data, existing data cleaning pre-processing was applied, which includes smoothing noise points and removing duplicated points in recordings[22,23]. Then, to distinguish the locations (or operations) of each tractor, the existing field-road trajectory segmentation algorithm was applied to recordings[22]. As illustrated in Figure 2, the segmentation algorithm splits each recording into daily trajectories and then segments each daily trajectory into field segments and road segments, where a daily trajectory refers to a piece of the recording collected on the same day. After field-road trajectory segmentation, each point in the recordings has a label, either "field" (denoted as "1") or "road" (denoted as "0"), as shown in the "Location" column in Table 1.

Table 1 An example of data recordings

Timestamp	Longitude	Latitude	Speed/m·s^{-1}	Direction/(°)	Torque percentage/%	Engine speed/m·s^{-1}	Fuel flow rate/L·h^{-1}	Location
20230525085016	116.374 26	29.951 972	5	300	38	2709	19.10	0
20230525085021	116.374 19	29.951 945	5	273	41	2704	19.55	0
…	…	…	…	…	…	…	…	…
20230614054437	117.782 57	37.469 909	6	191	23	1788	6.25	1

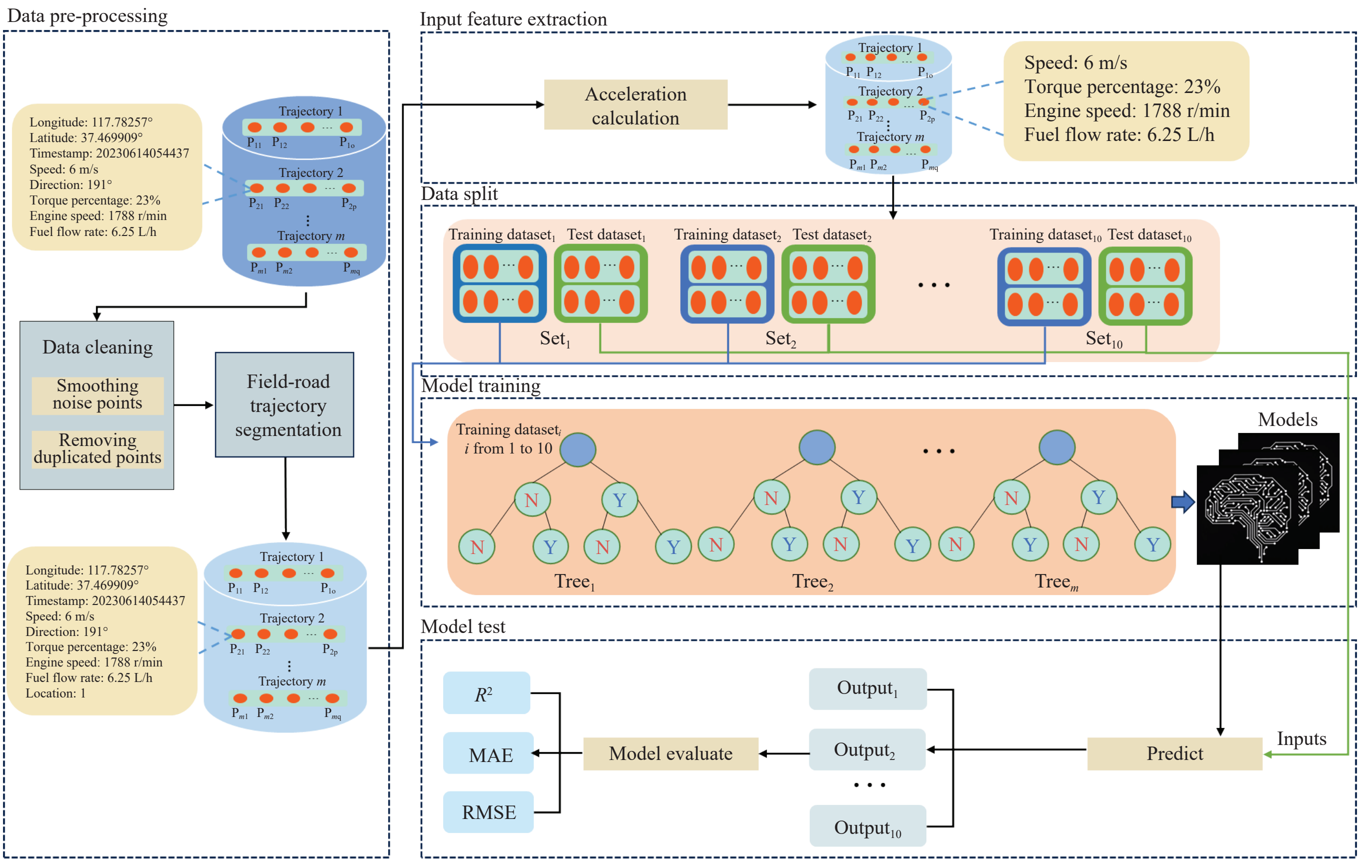

Figure 1 The workflow of our fuel consumption prediction method

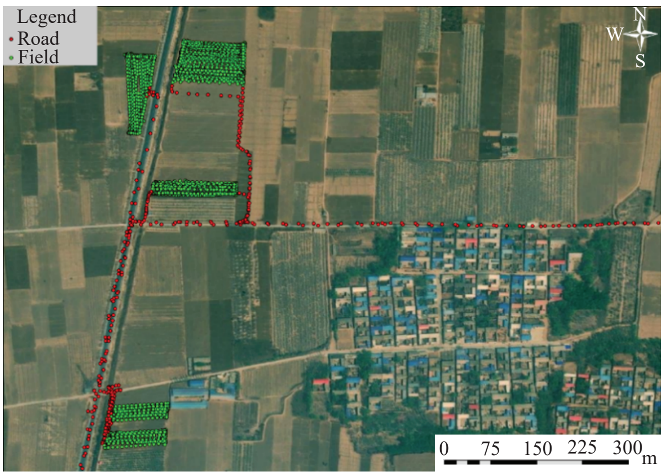

Figure 2 An example of the field-road trajectory segmentation

To demonstrate the effectiveness of our data, its statistics and regional distribution were examined. As listed in Table 2, for each tractor model, there are billions of points in its corresponding dataset, representing that the data is large-scale enough for model training and testing. Meanwhile, as listed in Table 3, for each tractor model, its daily trajectories are distributed across at least nine provinces, and most of the provinces are main grain-producing areas in China, representing various fuel consumption scenarios.

Table 2 Dataset statistics for three tractor models

Dataset statistics	WD904 (66.2 kW)	WD904G (66.2 kW)	WH1804 (180 kW)
Field point (×10^4)/percentage	97/90.7%	128/90.3%	127/91.2%
Road point (×10^4)/percentage	10/9.3%	14/9.7%	12/8.8%
Total points (×10^4)	107	142	139
Field segment (×10^2)/percentage	37/62.1%	52/64.2%	46/63.6%
Road segment (×10^2)/percentage	23/37.9%	29/35.8%	27/36.4%
Total segment (×10^2)	60	82	72

Furthermore, as listed in Table 2, for each tractor model, although most of its points (more than 90%) are located in the field, there are still about 9% of points that are located on the road, allowing two kinds of fuel consumption scenarios: in-field and on-road. Moreover, for each tractor model, its data cover more than 3700 fields with different field sizes and shapes, including various in-field fuel consumption scenarios[16].

Table 3 Regional distribution of the fuel consumption dataset

Province	WD904	WD904G	WH1804
Tianjin	0	0	9
Hebei	8	7	11
Shanxi	3	2	6
Jiangsu	43	122	113
Zhejiang	1	10	0
Anhui	90	112	88
Shandong	8	29	73
Henan	58	43	40
Hubei	37	11	17
Hunan	0	0	2
Guangdong	4	4	0
Shaanxi	4	5	0
Total	256	345	359

2.2 Methods

As shown in Figure 1, our fuel consumption prediction method generally consists of two steps: input feature extraction and model training and testing. The input feature extraction represents each point by a vector using information related to fuel consumption. Model training and testing develop fuel consumption prediction models and then evaluate their performances.

2.2.1 Input feature extraction

According to previous studies, the fuel consumption of tractors is influenced by many factors, including tractor configuration[8], implement specification[24], operation behavior[25], and field environment[26]. For tractor configuration, the study of Grisso et al.[7] has pointed out that instant engine working condition plays a crucial role in fuel consumption, so two engine-based parameters, torque and engine speed, were selected in this study. Moreover, driving speed and driving acceleration are fundamental motion-based parameters reflecting operation behaviors[27], where “driving speed” is a recorded parameter (“Speed” in Table 1) and “driving acceleration” is a parameter derived from the parameter “speed” with the calculation used in Poteko et al.[28] Overall, each point is transformed into a four-dimensional vector, including torque, engine speed, driving speed, and driving acceleration.

In addition, although there are various factors about implementation specification and field environments[29], most of them can be indirectly reflected through the combination of the aforementioned engine-based and motion-based parameters. For example, for a tractor operating using a large-width implement or in a wet-soiled field, its torque should be large in order to keep a normal driving speed. In order to balance the cost of data collection and the effectiveness of fuel consumption models, the four parameters (torque, engine speed, driving speed, and driving acceleration) were chosen in our study.

2.2.2 Model training and testing

Learning method: Tractor operations involve various factors such as tractor model, task type, terrain, weather conditions, and soil type. The relationship between fuel consumption and influencing factors is often non-linear, as parameters like engine speed and load have complex effects influenced by multiple variables. Meanwhile, fuel consumption prediction needs to be done in real time to optimize operations and resource allocation, which places high demands on computational efficiency. Machine learning algorithms can effectively capture complex non-linear relationships, process multi-dimensional feature data, and are suitable for modeling the relationship between fuel consumption and various influencing factors. Machine learning models can continuously update and optimize based on new data, adapt to dynamic changes in agricultural environments, and provide more accurate predictions. Machine learning algorithms are particularly adept at handling large-scale data. As fuel consumption prediction is a complex problem, machine learning methods were used to develop prediction models[30]. In large-scale data, Random Forests are usually superior to Decision Tree (DT, Breiman et al.[31,32]), Support Vector Regression (SVR, Awad et al.[33,34]), Light Gradient Boosting Machine (LightGBM, Ke et al.[35,36]), and Feedforward Neural Network (FNN, Al-Sager et al.[12]) due to their strong resistance to overfitting, high robustness, ability to handle high-dimensional data, automatic feature selection, parallel processing capabilities, fewer hyperparameter tuning requirements, and good interpretability. In order to efficiently train a model with large-scale data (e.g., billions of points in Table 1), Random Forest was chosen in this study.

Random Forest is an ensemble learning method that combines multiple weak learners (decision trees) into a strong model, allowing it to excel in tasks requiring high accuracy and robustness[37,38]. In order to improve the model’s generalization and reduce the model’s overfitting, instead of relying on a single decision tree, Random Forest builds a forest of many trees, as shown in the model training part of Figure 1, where each tree is trained on a different subset of training data. During model testing, the results from all trees are aggregated to make the final prediction, as shown in the model prediction part of Figure 1. In this study, Random Forest was implemented with Scikit-learn (a Python-based machine learning library, Pedregosa et al.[39]).

Model processing: For each tractor model, a fuel consumption prediction model was developed and tested using Random Forest. As shown in Figure 1, there are generally three steps in our model processing: data split, model training, and model test.

The goal of data splitting is to prepare training and test data. To restrict the inclusion of data from the same tractor to be either in the training or testing sets, the data split was designed as follows. For each tractor model, its 30 tractors are divided, where 27 tractors are for model training (e.g., training dataset 1 in Figure 1) and three tractors for model test (e.g., test dataset 1 in Figure 1). For training (or test) tractors, all recordings of these tractors serve as training (test) data. The data is split at the source level, and thus prevents fuel consumption prediction models from registering tractor-specific patterns. The presence of new tractors in the test data involves an entirely new environment, requiring a higher degree of generalization from the prediction model.

Given the training data (e.g., training dataset 1 in Figure 1), a model is trained using Random Forest. As Random Forest is a supervised machine learning method, besides inputting the four parameters, the training samples use their ground-truth fuel consumption values to guide the learning of model parameters.

Finally, the test data (e.g., test dataset 1 in Figure 1) are fed to the trained model to obtain predicted values, and the results are evaluated through a comparison between the predicted values and ground-truth values. As fuel consumption prediction is a regression problem, similar to the previous work[40], three performance metrics, R^2 (determination coefficient), *RMSE* (root mean squared error), and *MAE* (mean absolute error), are used to evaluate model performance[41]. For each performance metric, an average performance is calculated, which is the average of the performances of all points in the test data.

Moreover, as shown in Figure 1, in order to avoid the impact of the data split, the aforementioned model processing is repeated 10 times to generate 10 datasets, e.g., (training dataset 1, test dataset 1),

…, (training dataset 10, test dataset 10) in Figure 1, and 10 corresponding prediction models. Then, the average of performances from the prediction model test serves as the performance of our fuel consumption prediction method. The repeated model processing makes the evaluation of our proposed method statistically credible.

3 Results and discussion

Intensive experiments have been conducted to examine the effectiveness of the proposed fuel consumption prediction method and the transfer capability of the developed fuel consumption prediction models.

3.1 Model performance

In order to examine factors influencing the performance of fuel consumption prediction models, the following four sets of experiments were designed from the perspective of the machine learning method, input parameters, data split, and training data scale, respectively.

3.1.1 Machine learning method impact

A comparison was made between our proposed method and four other fuel consumption prediction methods based on different machine learning methods: DT, SVR, LightGBM, and FNN. Except for the choice of the machine learning method, the model processing of these baselines is the same as our approach (Section 2.2.2). The average performances of different fuel consumption prediction methods are listed in Table 4.

Table 4 Average performances of different fuel consumption prediction methods, where the used input parameters are the ones from our input feature extraction

Method	WD904			WD904G			WH1804		
	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*
FNN	0.38	3.73	3.13	0.57	3.16	2.34	0.25	4.45	3.71
SVR	0.63	2.86	2.44	0.59	3.30	2.88	0.57	3.57	2.86
DT	0.91	1.32	0.87	0.78	1.99	1.63	0.86	1.85	1.28
LightGBM	0.92	1.38	0.99	0.88	1.80	1.29	0.84	2.11	1.61
RF	**0.92**	**1.31**	**0.90**	**0.88**	**1.75**	**1.21**	**0.89**	**1.75**	**1.23**

In Table 4, our method consistently performs better than all baseline methods on the three tractor models, yielding at least an R^2 value of 0.88. This confirms the effectiveness of our proposed method. In addition, compared to LightGBM, another machine learning method based on decision trees, RF demonstrates more robustness. For example, on WD904 and WD904G, fuel consumption prediction methods based on LightGBM and RF achieve the same R^2 values, but the method based on RF obtains lower RMSE and MAE values; on WH1804, compared to LightGBM, the method based on RF achieves an improvement of 0.05 in R^2 values. The experimental results indicate the high accuracy of the RF-based fuel consumption prediction models.

3.1.2 Input parameter impact

There are four input parameters used in our method: torque, engine speed (EngSpd), driving speed (DrivSpd), and driving acceleration (DrivAcc). In order to examine the importance of each parameter, several parameter combinations were manually selected. For each input parameter combination, its model processing was applied in the same way as our approach (Section 2.2.2), and their average performances are listed in Table 5.

In Table 5, the engine-based parameters (torque and engine speed) have been shown to have their importance in determining the tractor's fuel consumption, achieving at least a high R^2 value of 0.84. This is consistent with the observations of previous studies[20,23]. When the motion-based parameters are incorporated, the model performances further increase by 0.05 on WH1804. This indicates our motion-based parameters are marginally helpful for the fuel consumption prediction in some cases.

Table 5 Average performances (R^2 value) of our method using different input parameter combinations

Parameters	WD904	WD904G	WH1804
Torque	0.73	0.67	0.59
EngSpd	0.60	0.62	0.67
DrivSpd	0.15	0.16	0.15
DrivAcc	−0.07	−0.02	−0.01
Torque+EngSpd	0.92	0.87	0.84
Torque+EngSpd+DrivSpd	0.92	0.88	0.87
Torque+EngSpd+DrivSpd+DrivAcc	**0.92**	**0.88**	**0.89**

3.1.3 Data split impact

Besides the input parameters and machine learning method, the performance of a fuel consumption prediction model is also influenced by its training and test data, which is set by the data split in our model processing. In order to examine the impact of the data split, as mentioned in Section 2.2.2, the model processing has been repeated to generate 10 fuel consumption prediction models for each tractor model, and the performance of each prediction model is given in Table 6.

Table 6 Performances (R^2 value) of each developed model using our proposed method

Model index	0	1	2	3	4	5	6	7	8	9	Average
WD904	0.94	0.82	0.85	0.91	0.92	0.97	0.95	0.97	0.93	0.96	0.92
WD904G	0.79	0.85	0.84	0.81	0.92	0.96	0.90	0.95	0.87	0.95	0.88
WH1804	0.80	0.83	0.88	0.84	0.85	0.96	0.93	0.93	0.94	0.94	0.89

In Table 6, the 10 fuel consumption prediction models developed for a tractor model have different performances, e.g., R^2 values range from 0.82 to 0.97 on WD904, from 0.79 to 0.96 on WD904G, and from 0.80 to 0.96 on WH1804. These performance variations are mainly dependent on the quality and distribution of training and test data used in model development and evaluation. The performance analysis confirms the necessity of repeating model processing for comprehensively evaluating a fuel consumption prediction method.

3.1.4 Data scale impact

To investigate the influence of the scale of training data on fuel consumption prediction, one experiment was run, where the size of the training data for fuel consumption prediction models are different. Specifically, one data split described in Section 2.2.2 is chosen to generate the training and test data. Then, n points are randomly selected from the training data to train a prediction model, and the model is evaluated using the test data. The model performances with varying n are illustrated in Figure 3.

In Figure 3, when the training data size n varies, the model performance changes. As n increases, the performance of each tractor model increases. Particularly, the performance of WH1804 increases the fastest. This is because the more training data are used, the more variation of input parameters is provided, leading to a more robust prediction model. Moreover, when n is greater than 300 000, the performance becomes stable. So, the large scale of training data is necessary for fuel consumption prediction model development.

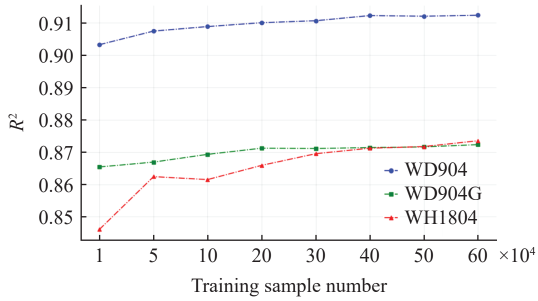

Figure 3 Performances of fuel consumption prediction models when the training sample number n varies

3.2 Model transferring capability

To examine the transferring capability of a developed fuel consumption prediction model, an experiment was designed that evaluates the model's performance with test data from the three tractor models. In other words, for the fuel consumption prediction model, its training and test data may be from different tractor models. For example, as shown in the "WD904" column of Table 7, a model was developed using the training data from WD904, but its test data are from the three tractor models.

Table 7 Performances (R^2 value) of our method, whose training and test data are from different tractor models

Test data	Training data		
	WD904	WD904G	WH1804
WD904	**0.92**	**0.90**	0.85
WD904G	0.86	0.88	0.87
WH1804	0.75	0.78	**0.89**

In Table 7, except for the models based on WD904G, other models perform best when their training and test data are from the same tractor model. This follows the observation of previous studies that tractor configuration is the most crucial factor for fuel consumption[3]. Furthermore, the model based on WD904G can perform well on WD904, because WD904 and WD904G use the same type of engine. So, the performance does not go against the previous observation. Moreover, the transferring capability analysis also provides the conditions to apply a developed fuel consumption prediction model.

4 Conclusions

In this study, we constructed a fuel consumption dataset for 90 tractors operated in various regions of China, and proposed a machine-learning prediction approach that estimates the instant fuel flow rate of tractors in practical scenarios. The fuel consumption prediction approach uses Random Forest and the four parameters to develop fuel consumption prediction models, where torque and engine speed reflect engine conditions, and driving speed and acceleration reflect motion characteristics. The experiment result demonstrates that our fuel consumption prediction method can achieve state-of-the-art performances (R^2≥0.88). Moreover, the in-depth experiment analysis not only shows the contributions of four factors (i.e., Random Forest, the input parameters, data setting, and training data size) to model effectiveness, but also demonstrates the impact of tractor models on the transferring capability of a developed prediction model.

Our future studies will focus on collecting fuel consumption data with higher quality, such as using ECU and GNSS devices with a higher acquisition frequency, to effectively capture factors involved in fuel consumption. Meanwhile, we aim to develop a general fuel consumption prediction model that could handle different tractor models. We believe that such directions can further enhance the evaluation accuracy of tractor fuel consumption.

Acknowledgements

This research work was supported by the National Precision Agriculture Application Project (Grant No. JZNYYY001) and Jiangxi Academy of Agricultural Sciences Special Fund for Basic Research and Talent Cultivation (Grant No. JXSNKYJCRC202632).

[References]

[1] Wang Y, Wang L, Wang S M, Zhao J S, Zhang Y J, Wen C K, et al. Loading method for tractor rotary tillage load spectrum based on extreme load retention resampling. Int J Agric & Biol Eng, 2025; 18(4): 128–138.

[2] Lee J W, Kim J S, Kim K U. Computer simulations to maximise fuel efficiency and work performance of agricultural tractors in rotovating and ploughing operations. Biosystems Engineering, 2016; 142: 1–11.

[3] Zhang W P, Guo H Z, Zhao B, Zhou L M, Wang F Z, Wang D Y, et al. Full-condition monitoring and intelligent yield prediction and decision-making technology for wheat combine harvesters. Int J Agric & Biol Eng, 2025; 18(6): 202–211.

[4] Romanelli T L, Milan M. Machinery management as an environmental tool-material embodiment in agriculture. Agricultural Engineering International: CIGR Journal, 2009; 14(1): 63–73.

[5] Tang Q J, Ren B Y, Wu J P, Hu J C, Fu J Q, Zhang D Q. Experimental study on diesel engine performance of tractor under transient conditions. Thermal Science and Engineering Progress, 2025; 61: 103509.

[6] Pan L W. Research on fuel consumption prediction for agricultural machinery enhanced by motion features. Master's thesis. China Agricultural University, 2025; 5p. (in Chinese)

[7] Grisso R D, Kocher M F, Vaughan D H. Predicting tractor fuel consumption. Applied Engineering in Agriculture, 2004; 20(5): 553–561.

[8] Battiato A, Diserens E. Influence of soil on the traction performance of a 65 kW MFWD tractor. Journal of Agricultural Science, 2024; 11(17): 11–27.

[9] Damanauskas V, Velykis A, Satkus A. Efficiency of disc harrow adjustment for stubble tillage quality and fuel consumption. Soil and Tillage Research, 2019; 194: 104311.

[10] Ekemube R A, Atta A T, Ndirika V I O. Optimization of fuel consumption for tractor-tilled land area during harrowing operation using full factorial experimental design. Covenant Journal of Engineering Technology, 2023; 7(2): 21–30.

[11] Md-Tahir H, Zhang J, Xia J, Zhou Y, Zhou H, Du J, et al. Experimental investigation of traction power transfer indices of farm-tractors for efficient energy utilization in soil tillage and cultivation operations. Agronomy, 2021; 11(1): 168.

[12] Al-Sager S M, Almady S S, Marey S A, Al-Hamed S A, Aboukarima A M. Prediction of specific fuel consumption of a tractor during the tillage process using an artificial neural network method. Agronomy-Basel, 2024; 14(3): 492.

[13] Siddique M A A, Baek S Y, Baek S M, Jeon H H, Lee J H, Son M A, et al. The selection of an energy-saving engine mode based on the power delivery and fuel consumption of a 95 kW tractor during rotary tillage. Agriculture, 2023; 13(7): 1376.

[14] Moinfar A, Shahgholi G, Gilandeh Y A, Gundoshmian T M. The effect of the tractor driving system on its performance and fuel consumption. Energy, 2020; 202: 117803.

[15] Tihanov G, Ivanov N. Fuel consumption of a machine-tractor unit in direct sowing of wheat. Agricultural Science and Technology, 2021. DOI:10.15547/AST.2021.01.007.

[16] Janulevičius A, Šarauskis E, Čiplienė A, Juostas A. Estimation of farm tractor performance as a function of time efficiency during ploughing in fields of different sizes. Biosystems Engineering, 2019; 179: 80–93.

[17] Bbla N, Mileusni Z, Dragievi A, Milanovi M, Rajkovi A, Miodragovi R, et al. Implementation of XGBoost models for predicting CO_2 emission and specific tractor fuel consumption. Agriculture, 2025; 15(11): 1209.

[18] Breiman L. Random forests. Machine Learning, 2001; 45(1): 5–32.

[19] Lin B H, Wu C C, Song W, Jaafar H A. Detection of the farm road positioning scene classification for unmanned driving based on GNSS. Int J Agric & Biol Eng, 2026; 19(2): 226–234.

[20] Wu C F, Deng J S, Wang K, Ma L G, Tahmassebi A R S. Object-based classification approach for greenhouse mapping using Landsat-8 imagery. Int J Agric & Biol Eng, 2016; 9(1): 79–88.

[21] Wu C C, Li D, Zhang X Q, Pan J W, Quan L, Yang L L, et al. China's agricultural machinery operation big data system. Computers and Electronics in Agriculture, 2023; 205: 107594.

[22] Chen Y, Zhang X Q, Wu C C, Li G. Field-road trajectory segmentation for agricultural machinery based on direction distribution. Computers and Electronics in Agriculture, 2021; 186: 106180.

[23] Zhang X Q, Chen Y. Field-road classification for agricultural vehicles in China based on pre-trained visual model. Peer J Computer Science, 2024. DOI: 10.7717/peerj-cs.2359.

[24] Kuboń M, Cupia M, Szelg-Sikora A, Kobuszewski M. The impact of purchasing new agricultural machinery on fuel consumption on farms. Sustainability, 2024; 16(1): 18.

[25] Yang L L, Tian W Z, Zhai W X, Wang X X, Chen Z B, Wen L, et al. Behavior recognition and fuel consumption prediction of tractor sowing operations using a smartphone. Int J Agric & Biol Eng, 2022; 15(4): 154–162.

[26] Kolator B A. Modeling of tractor fuel consumption. Energies, 2021; 14(8): 2300.

[27] Huang J, Liang S, Wu C C, Kou Z, Chen Y. Evaluation system for agricultural machinery operation based on smartphone sensors. Applied Engineering in Agriculture, 2022; 38(2): 227–242.

[28] Poteko J, Eder D, Noack P O. Identifying operation modes of agricultural vehicles based on GNSS measurements. Computers and Electronics in Agriculture, 2021; 185: 106105.

[29] Bietresato M, Calcante A, Mazzetto F. A neural network approach for indirectly estimating farm tractors engine performances. Fuel, 2015; 143: 144–154.

[30] Rahimi-Ajdadi F, Abbaspour-Gilandeh Y. Artificial neural network and stepwise multiple range regression methods for prediction of tractor fuel consumption. Measurement, 2011; 44(10): 2104–2111.

[31] Breiman L, Friedman J H, Olshen R A, Stone C J. Classification and regression trees. Encyclopedia of Ecology, 2015; 57(3): 582–588.

[32] Lian Y, Chen J, Guan Z H, Song J. Development of a monitoring system for grain loss of paddy rice based on a decision tree algorithm. Int J Agric & Biol Eng, 2021; 14(1): 224–229.

[33] Khanna R, Awad M. Support vector regression. in: Efficient learning machines: Theories, concepts, and applications for engineers and system designers, 2015; pp.67–80. DOI: 10.1007/978-1-4302-5990-9

[34] Fu Q, Shen W Z, Wei X L, Yin Y L, Zheng P, Zhang Y G, et al. Predicting the excretion of feces, urine and nitrogen using support vector regression: a case study with holstein dry cows. Int J Agric & Biol Eng, 2020; 13(2): 48–56.

[35] Chen J, Liao K, Wan Y, Chen D Z, Wu J. DANets: deep abstract networks for tabular data classification and regression. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2022; 36(4): 3930–3938. DOI: 10.48550/arXiv.2112.02962.

[36] Ke G, Meng Q, Finley T, Wang T F, Chen W, Ma W D, et al. Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 2017; 30.

[37] Belgiu M, Draguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 2016; 114: 24–31.

[38] Hu J C, Szymczak S. A review on longitudinal data analysis with random forest. Briefings in Bioinformatics, 2023; 24(2). DOI: 10.1093/bib/bbad002.

[39] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 2011; 12: 2825–2830.

[40] Sayin C, Ertunc H M, Hosoz M, Kilicaslan I, Canakci M. Performance and exhaust emissions of a gasoline engine using artificial neural network. Applied Thermal Engineering, 2007; 27(1): 46–54.

[41] Zhang J, Hu Y C, Xin Y H, Zhou X Y, Fu C, Tu W, et al. Prediction model and inventory estimation of CO_2 and NOx emissions from tractors based on Beidou trajectory big data. Transactions of the CSAE, 2026; 42(8): 47–56. (in Chinese)