IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 May 2024, accepted 11 August 2024, date of publication 14 August 2024, date of current version 28 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3443604

RESEARCH ARTICLE

Ground Handling Process Optimization Model Linked to Flight Delay Prediction Results

ZHEN-TENG XU[1,2,3], YAN-JUN LI[1], HONG-FU ZUO[1], TENG-ZHOU XU[2,3], QI WANG[4], WANG-WANG YU[3,5], HONG-SHENG YAN[2,3], BING LIU[6], TAO CHEN[2,3], AND MAO-HUI ZHOU[1]

[1]College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Jiangning, Nanjing 211100, China
[2]School of Aeronautic Engineering, Nanjing Vocational University of Industry Technology, Qixia, Nanjing 210046, China
[3]Aeronautic Intelligent Manufacturing and Digital Health Management Technology Engineering Research Center of Jiangsu Province, Qixia, Nanjing 210046, China
[4]Shanghai Airport (Group) Company Ltd., Shanghai 201300, China
[5]School of Mechanical Engineering, Nanjing Vocational University of Industry Technology, Qixia, Nanjing 210046, China
[6]Department of Public Foundational Courses, Nanjing Vocational University of Industry Technology, Qixia, Nanjing 210046, China

Corresponding author: Zhen-Teng Xu (xuzt@niit.edu.cn)

This work was supported in part by the Ministry of Science and Technology, China, under Grant 2023YFB4302403.

ABSTRACT In response to the challenge of mitigating flight delays, this study introduces an innovative solution that encompasses the prediction of delay durations for existing flights and the subsequent optimization of ground service processes based on these predictions. The indirect forecasting of flight delays is achieved through the construction of a random forest model, exhibiting a remarkable 100% accuracy when considering a 15-minute standard for flight delays. In light of the delay prediction outcomes, distinct delay coefficients are assigned to individual flights, facilitating the development of a ground service optimization model through the application of a genetic algorithm. Within the genetic algorithm optimization framework, significant enhancements have been implemented in the gene encoding of the initial population, incorporating a segmented encoding approach. Employing this refined model to optimize the service sequence and duration of ground service vehicles for all flights culminates in the notable accomplishment of achieving zero delays for the entire set of flights.

INDEX TERMS Flight delay prediction, ground service optimization, random forest model, genetic algorithm model.

I. INTRODUCTION

The aviation transportation industry has consistently faced numerous challenges, with flight delay being a prominent issue. Flight delays not only result in significant economic losses for airlines and airport operators but also inconvenience passengers. Despite various measures implemented by airlines to mitigate delays, multiple factors such as weather conditions, mechanical failures, and air traffic control persist, contributing to the occurrence of flight delays. According to information provided by the online platform FLIGHTSTATS, based on global flight data, the average on-time performance of the top 10 airports with the highest punctuality rates in August 2023 was 83.15%. However, Asia Pacific Airlines achieved an on-time performance of only 79.62%, while North America Airlines had the lowest punctuality rate at 71.82% [1].

In recent years, scholars have focused their research on two main aspects to further reduce the impact of flight delays. Firstly, there is an emphasis on establishing effective flight delay prediction models to provide accurate forecasts of the duration of flight delays, offering valuable references for airlines. Secondly, optimization of ground service processes is explored, aiming to identify more rational service pathways and thereby reduce the overall duration of flight delays.

Most of the existing papers on flight delay prediction have focused on both influence factor extraction and prediction models [2]. Wu et al. combed the advantages of Dense Net and SEN et al, can not only enhance the transmission of deep information, avoid the problem of vanishing

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose.

gradients, but also achieve feature recalibration by the feature extraction process [3]. Esmaeilzadeh et al. employed a support vector machine (SVM) model to explore the non-linear relationship between flight delay outcomes [4]. Ye et al. used four popular supervised learning methods: multiple linear regression, a support vector machine, extremely randomized trees and LightGBM are investigated to improve the predictability and accuracy of the departure delays prediction model [5]. Bisandu et al. proposed a novel alternative method for flight delay prediction, namely social ski driver conditional autoregressive-based (SSDCA-based) deep learning [6]. Such as [5], Wang et al. selected three machine learning algorithms to predict the distribution of flight delays: multilayer perceptron (MLP), LightGBM and random forest (RF) [7]. Unlike [3], [4], [5], [6], [7], Wang et al. focused on the use of macro and micro factors for predicting flight delay [8]. They proposed two novel explanatory variables from a microscopic perspective: departure pressure and cruise pressure. Shi. constructed a departure flight delay prediction model based on a deep fully connected neural network [9].

The aircraft Ground Handling (GH) operations represents the airside activities at airports in charge of processing passengers, cargo, facilities and supplies at and around parked aircraft. Most of these operations are performed by different service providers, using a wide panoply of specialized vehicles and equipment known as Ground Support Equipment (GSE) whose management is core to GH [10]. The emerging technologies lead the airport industry towards Smart airports. Smart Airport concept is the future of Airport operation and it may dramatically change the industry towards modern technology adaptation [11]. Airport Collaborative Decision Making (ACDM) is one of the products of smart Airport. The primary objective of ACDM is to enhance the efficient utilization of airport resources, reduce flight delays, improve flight punctuality, and enhance the overall operational efficiency of the airport. In [10], Tabares et al. developed a multi-time scale airport GH management structure with on-line solutions for the GSE fleet operations compatible with increased GH automation (both for information and GSE hardware) in the context of ACDM operations. Ip et al. proposed a genetic algorithm with a hybrid encoding scheme [12]. In this paper, the encoding scheme describes the bit of a chromosome as appositive real number. Its integer part stands for the assigned crew of a flight and its decimal part for the position in job sequence by percent. For accurately estimating the ground service time of the transit flight and realizing the accuracy of the flight push time control, Xing et al. proposed a time estimation method based on Markov Monte Carlo (MCMC) for the ground service of the transit flight [13]. In addition to the above research, some scholars have also proposed new algorithm models from the fields of service vehicle operating costs [14] and crew scheduling process [15].

At present, there are mainly the following problems in research on the above two fields.

(i) The data used in the flight delay prediction model is not comprehensive enough. Agogino et al. [16] failed to consider the airport's delays to optimize take-off and landing intervals. Bansal et al. [17] does not utilize features such as unique air carriers, tail number of aircraft, and origin/destination airports with trigonometric transform function in the model training to enhance predictions' accuracy. Yu et al. [18] the method did not use an optimization algorithm even without access to the required data on air traffic control.

(ii) The definitions involved in ground service optimization problems are not clear enough, such as the turnaround time (TRT). In [10], TRT of an aircraft is the time duration this aircraft is stopped at its parking position (from on-blocks until off-blocks, when chocks are removed from aircraft landing gear) and subject to GH activities. Fricke et al. [19] estimated TRT as the duration of the critical path of the different GH handling activities taking place with that aircraft, plus eventually some buffer time.

(iii) When using a genetic algorithm for ground service optimization, the problem model and encoding method require further refinement. Ground service optimization is a complex temporal and path sorting problem, and the model varies with different objectives. Ip et al. [12] and Tang [20] both modeled optimization problems under specific conditions and proposed encoding methods for the matching model. However, these approaches have certain limitations and poor portability.

(iv) There is no specific treatment measure for delayed flights. While flight delay prediction models can provide forecast results, they do not offer measures to avoid delays for the identified delayed flights. Similarly, ground service optimization focuses on general improvements for the entire airport and lacks effective linkage with delayed flights.

To address the aforementioned issues, this paper proposes a ground service optimization model associated with flight delay prediction results. This model utilizes ACDM data as the prediction basis, ensuring selected data comprehensively considers various factors influencing flight delays. Based on the predicted delay levels, the model optimizes the ground service processes for relevant flights, ensuring they do not experience further delays. To implement this model, an RF-GA (Random Forest and Genetic Algorithm) fusion algorithm is developed. The core principle involves incorporating RF prediction results into the problem description of GA. The multi-objective optimization algorithm NSGA-II focuses on optimizing the ground service processes for delayed flights, ensuring normal flights remain unaffected while planning reasonable ground service processes for delayed flights. The ultimate goal is to normalize the operations of delayed flights.

II. ACDM DATA PROCESSING AND ANALYSIS

This section primarily involves preliminary processing and analysis of the data downloaded from the ACDM system, defining some basic parameters. Data processing aims to modify the data format for readability and ensure the accuracy

TABLE 1. Raw data sample.

CLA	ITY	STA	ALDT	LTOT	· · ·	In Block	Trailer In Place	Off Block	Takeoff Instructions
H/Z	B77L/E	2023-5-31 11:35:00	2023-5-31 10:55:00	2023-6-1 0:35:00	· · ·	2023-5-31 10:59:00	2023-5-31 22:56:00	2023-5-31 23:54:00	2023-6-1 0:15:00
W/Z	B77W/E	2023-5-31 22:20:00	2023-5-31 22:06:00	2023-6-1 0:35:00	· · ·	2023-5-31 22:14:00	2023-5-31 23:47:00	2023-6-1 0:32:00	2023-6-1 0:52:00
· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·

TABLE 2. Acronyms in this paper.

Serial Number	Term	Acronyms	Paraphrase
1	Actual Landing Time	ALDT	The actual landing time of the aircraft is the time when the aircraft arrives at the destination airport and lands successfully.
2	Actual Takeoff Time	ATOT	The actual departure time of the flight.
3	Common Logistics Assessment	CLA	Used to assess and report logistics situations to ensure effective air transport and logistics management.
4	Gate Open Time	GOT	Boarding gate opening time.
5	Identity	ITY	Refers to aircraft identification information, usually including the aircraft model, registration number or airline logo.
6	Last Takeoff Time	LTOT	Within a certain period of time, the last time a flight is allowed to take off.
7	Cabin Door Open Time	CDOT	Cabin door opening time, indicating that passengers can get on and off normally.
8	Service Time	ST	Flight service time
9	Scheduled Time of Arrival	STA	The estimated time of arrival of an aircraft at its destination, usually specified in the flight plan.
10	Scheduled Time of Departure	STD	The scheduled departure time of the aircraft according to the flight plan
11	Stand Update	STUA	The ground status of the aircraft, generally refers to the arrival status.
12	Stand Update	STUD	The ground status of the aircraft, generally refers to the departure status
13	Terminal Aerodrome Reference Code	TAR	An international standard for describing the classification and size classes of airport runways and parking bays.

of data used for prediction. Data analysis involves initial calculations and analysis of flight operation data to identify basic trends and patterns, providing a basis for further research.

A. DATA SOURCES

The ACDM aims to establish a highly collaborative airport operational environment, ensuring efficient utilization of airport resources, increased capacity, reduced delays, enhanced passenger experience, cost reduction in operations, and promotion of sustainability across the aviation industry. This system finds widespread application globally, providing support to airports of various sizes and types. The ACDM has introduced a milestones approach for the whole turnaround process of each aircraft at airports [10]. In the realm of ACDM, numerous milestones abound, encompassing elements such as the Scheduled Time of Arrival (STA), Latest Time of Turnaround (LTOT), Scheduled Time of Departure (STD), Gate Open Time (GOT), and many others. A sample of raw data exported from ACDM is shown in Table 1.

Since most of the data names in ACDM are given in the form of abbreviations, which are difficult to read, all abbreviations involved in this article are listed in Table 2 to facilitate comparison.

Since the beginning of 2020, COVID-19 has swept the world and had a catastrophic impact on the aviation industry, causing flight volumes to plummet. According to IATA statistics, the airline sector produced an economic loss of USD 175 billion in 2020 (10 times larger than the average annual value destruction pre-pandemic) and USD 104 billion in 2021, resulting in economic profit margins of -46% and -21% respectively [21]. In another report [22], there were more new airlines born, and fewer airlines being closed in 2021 compared to 2019. The statistical results can be found in Fig 1. In the long-term forecast, the global passenger travel to return to the 2019 level of activity in 2024 and to expand substantially over the next two decades.

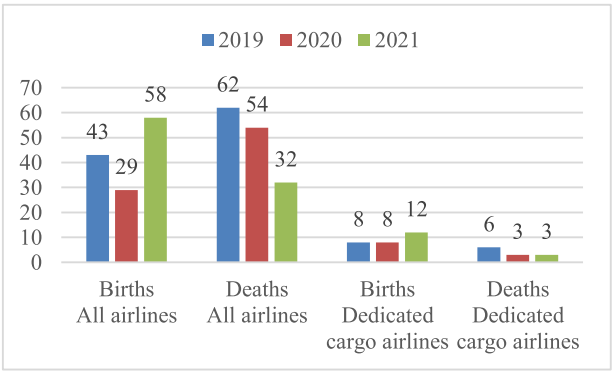

FIGURE 1. Number of airline births and, deaths.

To further determine the selected years and months for experimental data, an analysis of flight operation data at Shanghai Pudong Airport from 2019 to 2023 was conducted, and the results are depicted in Figure 2. It is evident that the total number of flights in 2023 still exhibits a significant

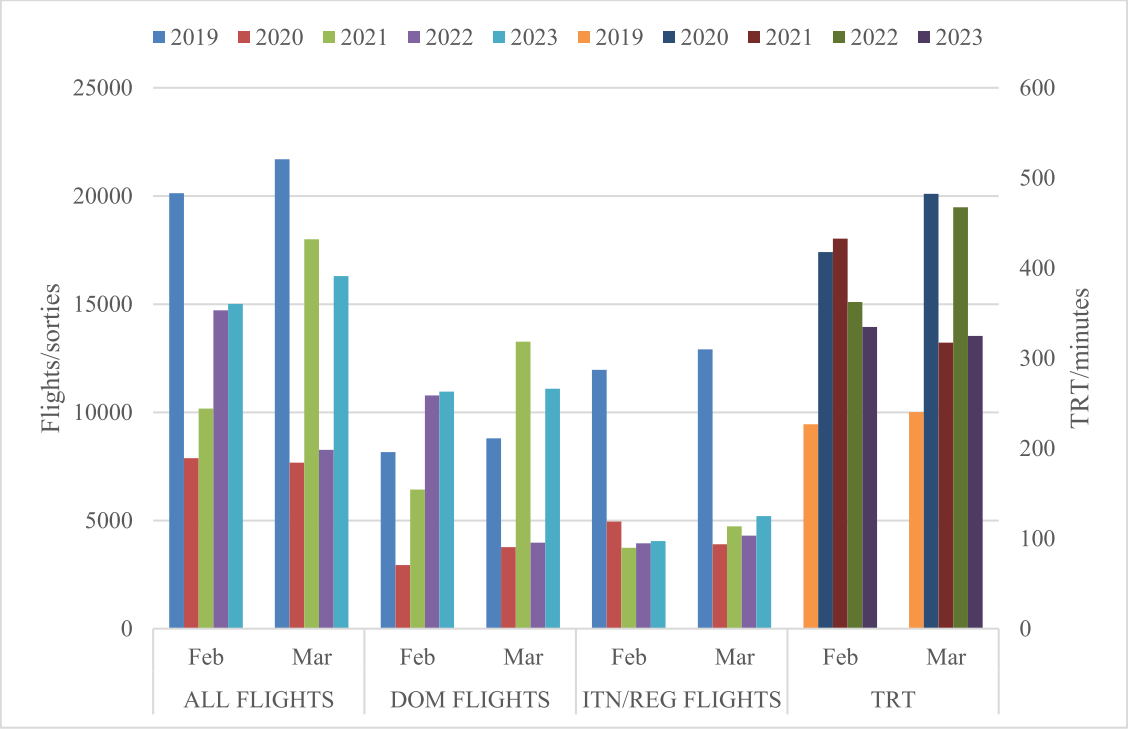

FIGURE 2. Number of flights and TRT from 2019 to 2023.

gap compared to 2019. However, the domestic flight volume has surpassed the levels observed during the pandemic, with notable differences observed in the international/regional flight volume. Although the quantity of international/regional flights has not yet reached pre-pandemic levels, it shows a consistent annual increase, suggesting that over time, it may surpass the pre-pandemic flight volumes. Due to strict requirements for epidemic prevention and control at transportation airports during the pandemic [23], the turnaround time (TRT) of an aircraft has notably increased. In Fig 2, the TRT in 2023 has significantly decreased compared to 2020 but still remains longer than that in 2019.

There are two possible reasons why the Turnaround Time (TRT) in 2023 might be longer than that in 2019: firstly, the current TRT may still be in a declining phase and has not yet returned to pre-pandemic levels; secondly, new operational processes were introduced during the pandemic to increase flight turnover, and these processes have been retained even after the pandemic, leading to an extended TRT. To further ascertain the patterns of TRT variations, an analysis of flight data from February to June 2023 was conducted, and the results are presented in Fig 3.

In Fig 3, the overall trend of flight quantity exhibits an upward trajectory, reaching its peak in June, while the Turnaround Time (TRT) demonstrates an overall declining trend, reaching its lowest point in June. Combining the analysis presented earlier, both flight quantity and TRT in 2023 indicate a recovery period post-pandemic, expected to reach pre-pandemic levels in 2024. The data for June provides

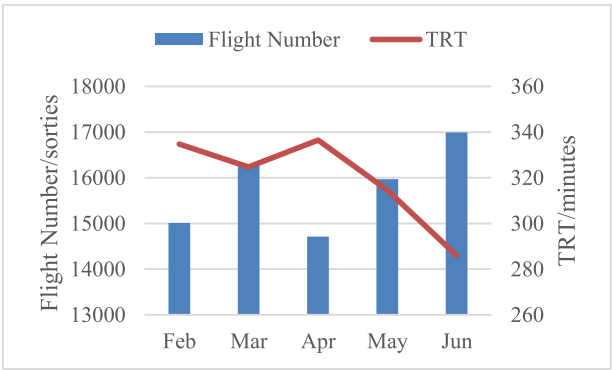

FIGURE 3. Number of flights and TRT from February to March 2023.

the closest approximation to pre-pandemic levels currently available. Consequently, the operational data for Shanghai Pudong Airport in June 2023 is selected for all model-related data in this paper.

B. DATA PREPROCESSING

The data exported from ACDM contains a plethora of information, and if not appropriately handled, certain information may adversely impact the analysis. Therefore, data preprocessing is necessary. The data preprocessing in this paper involves a total of eight steps, as outlined in Table 3.

Most temporal data in ACDM is in the form of dates and times. To facilitate subsequent processing, it is essential to establish a starting point and convert other time points

TABLE 3. Data preprocessing steps.

Step	Name	Details
Step one	Time conversion *	Convert node time to duration (minutes)
Step two	Delete non-arrival flights **	Based on STUA and STUD, select the flights that actually arrive at the airport.
Step three	Confirm flight time **	Select flights between 07:00-18:00 based on ALDT
Step four	Determine parking bays **	Select the flight on apron 1 based on TAR, TAR=Nr.1-12.14-32
Step five	Extract important parameters **	Important parameters: ITY/TAR/ALDT/STD/ In Block/ Open Cabin Door/Gate Open/Off Block/ATOT/LTOT/ Delay Time
Step six	Data cleaning *	Delete empty data, Delete or correct erroneous data

In the aforementioned steps, *-marked steps involve calculations, while **-marked steps require data filtering based on conditions.

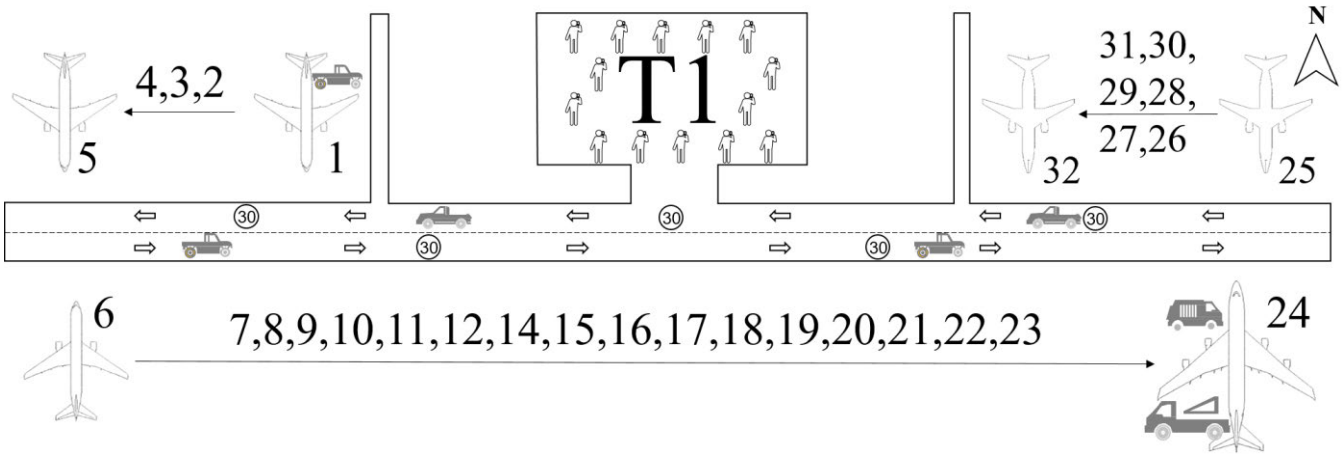

FIGURE 4. Parking space distribution map.

into time differences measured in minutes. Besides known temporal nodes, the times used for predictive analysis are also presented in the form of time differences. While many scholars prioritize the deletion of erroneous data or filling missing data based on known information during data cleaning, this paper introduces a novel method for correcting erroneous data. This approach involves determining the causes of erroneous data based on the numerical characteristics of related data and subsequently correcting the data accordingly.

As the flight operation data encompasses all flights, including canceled and diverted flights, which significantly compromise the accuracy of delay prediction, it is crucial to filter the data using STUA and STUD. During the filtering process, only data with STUA and STUD set to ''Arrival'' is retained, and other data is deleted. This paper specifically addresses the delay issues of normally transiting flights, excluding delays of overnight parked flights, and filters the data based on ALDT time. One advantage of this approach is that time exhibits a linear relationship, allowing for direct encoding. If delays within a 24-hour period are considered, trigonometric conversion of time would be necessary, as detailed in [24].

The planning of flight ground services is closely related to the location of parking positions. Parking positions on the same apron are often closer, making it easier for a ground service vehicle to move between different flights with shorter transfer times. This arrangement facilitates the efficient servicing of multiple flights by a single ground service vehicle, thereby saving costs for the airline. Using the example of the No. 1 apron at Shanghai Pudong Airport, the layout of parking positions on this apron is illustrated in Fig 4.

In Fig 4, the No. 1 apron has a total of 31 parking positions distributed across three areas. Positions Nr.1-5 are on the west side of Terminal 1 (T1), positions Nr.6-12 and Nr.14-24 are located on the south side of the terminal, and positions Nr.25-32 are on the east side of the terminal. Due to different aircraft types parked at each position, the distances between them vary, directly affecting the transfer time of ground service vehicles. Specific information about parking positions is provided in Table 4.

Flight area standards II, wingspan, and aircraft types are referenced from [25]. In Table 3, the aircraft parked at positions Nr.25-32 have the smallest wingspan, belonging to Class C flight areas; those at positions Nr.1-5 have slightly larger wingspans, falling under Class D aircraft; and positions on the south side of the terminal, accommodating aircraft with the largest wingspans, belong to Class E and F flight areas. Referring to Table 1, information about airport flight areas is displayed in ITY, positioned after the aircraft types. In the absence of knowledge about specific aircraft types, the correctness of parking positions can be directly verified based on flight area information.

TABLE 4. Parking space information.

Flight area standards II	Wingspan (m)	Aircraft Types	Parking space
A	<15	None	None
B	15~24(Not included)	None	None
C	24~36(Not included)	B737、B738	N r.25-32
D	36~52(Not included)	A20N、A319、A320、A321	Nr.1-5
E	52~65(Not included)	A332、A333、A339、A359、B772、B77W、B788、B789	Nr.6-12.14.15.16.18.20.22.23
F	65~80(Not included)	None	Nr.17*.19*.21*.24

*-marked aircraft wingspan is 65-68.5.

The process of extracting important parameters involves creating a new data table. This step must be performed before data cleaning to avoid unintentionally deleting necessary data due to irrelevant errors. The data extracted in this step is determined based on subsequent requirements, with reference to Table 2 for the data needed in this paper.

C. DATA ANALYSIS

This section primarily conducts statistical analyses on two crucial parameters: average service time and departure waiting time. The average service time measures the total time a flight spends receiving ground services at the airport, including waiting time for services. Its calculation formula is

$$T_{sev} = \frac{\sum_1^n (t_{obi} - t_{ibi})}{n} (i = 1, 2 \cdots n). \tag{1}$$

In the formula: T_{sev} is average service time, t_{obi} is off block time, t_{ibi} is in block time, n is the number of flights with same identity. With (1), we can calculate the average service time for different identities. Statistical analysis was performed on the average service time, calculating the Median Time, Maximum Time, Minimum Time, and Average Time. To facilitate the observation of data characteristics, the above results were visualized using Candlestick Charts.

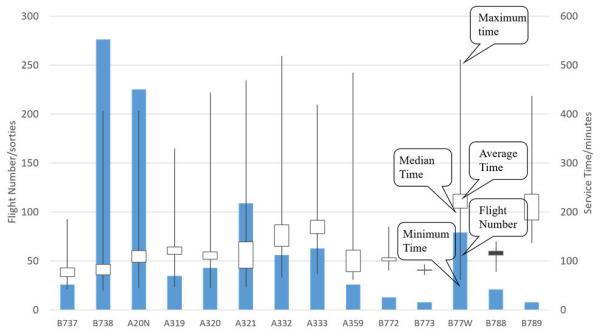

FIGURE 5. Average service time candlestick charts.

In Fig 5, the minimum times for different aircraft types are very close, rendering this data less significant as a reference. The maximum time, on the other hand, varies significantly due to differences in aircraft types and can serve as a reference for extreme conditions. With the exception of B773 and B788, which display negative trends, all other aircraft types show positive trends, indicating that the average time is greater than the median time. This suggests that the majority of flights have service times less than the average.

Takeoff waiting time is used to measure the time a flight needs to wait for departure after completing all ground service tasks. Takeoff waiting time directly reflects the current traffic conditions at the airport, with longer waiting times indicating busier traffic. Its calculation formula is

$$T_{tow} = \frac{\sum_1^n (t_{LTOT\,i} - t_{obi})}{n} (i = 1, 2 \cdots n). \tag{2}$$

In the formula: T_{tow} is takeoff waiting time, $t_{LTOT\,i}$ is last takeoff time, t_{obi} is off block time, n is the number of flights with same identity. Perform statistical calculations of takeoff waiting times and plot them on Candlestick Charts.

In Fig 6, the minimum times for all flights are generally short, with some flights approaching 0, rendering them without significant reference value. There is considerable variability in the maximum times for different flights, showing a substantial degree of randomness. Consequently, these values are not considered in the subsequent analysis. In Figure 6, there are 8 flights represented by negative trends, while 6 flights exhibit positive trends. Moreover, the physical areas are relatively small. Consequently, there is not a substantial difference between the average time and the median time. Therefore, this paper adopts the average time as the takeoff waiting time for each flight.

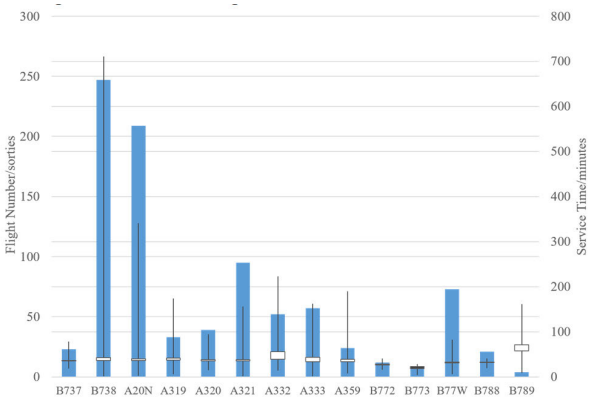

FIGURE 6. Takeoff waiting time candlestick charts.

III. FLIGHT DELAY PREDICTION MODEL

In this section, we will analyze different algorithm models and provide reasons for choosing the Random Forest algorithm for model development. Subsequently, in conjunction with the flight delay prediction problem, we will use the Random Forest algorithm for modeling and evaluate the predictive results.

A. COMPARATIVE ANALYSIS OF DIFFERENT ALGORITHMS

In the field of predictive modeling, there are currently two main types of popular algorithms: machine learning algorithms and deep learning algorithms. These two algorithms have slightly different application areas, with machine learning algorithms primarily focusing on predictive modeling for large datasets, while deep learning is more inclined towards handling textual and graphical data. Therefore, when predicting overstay models, the majority of choices are machine learning algorithms. Commonly used machine learning algorithms include Random Forest, Generalized Linear Model (GLM), Gradient Boosting Model (GBM), K-means, and Prophet.

Each of the mentioned algorithms has its own applicable domain. The advantages of the Generalized Linear Model include its fast-training speed, applicability to response variables with any form of exponential distribution, clear understanding of how each predictor variable affects the outcome, and resistance to overfitting. However, it requires relatively large datasets and is susceptible to the influence of outliers.

The prominent feature of the Gradient Boosting Model is that it builds one tree at a time. The GBM method provides more expressive data, and benchmark results indicate its preference for overall data integrity. However, it also requires more time as it sequentially builds each tree.

K-means is a very popular high-speed algorithm that involves placing unlabeled data points in separate groups based on similarity, commonly used in clustering models. K-means is highly effective for implementing personalized plans on large datasets, especially widely applied in predictive analytics in the healthcare sector.

The Prophet algorithm is particularly useful in capacity planning, such as resource allocation and setting sales targets. It is flexible enough to incorporate heuristic methods and useful assumptions. The algorithm's speed, reliability, and robustness make it a popular alternative algorithm choice for time series and predictive analysis models, especially when dealing with messy data.

The predictive model chosen for this paper is the Random Forest model, primarily considering the following advantages:

(i) The model operates using multiple trees, effectively reducing the error of individual trees.

(ii) The Random Forest model can resist overfitting effectively and simultaneously handle a larger amount of data.

(iii) For multi-variable predictions, it can estimate the importance of variables and maintain prediction accuracy in the event of data loss.

B. CONSTRUCTION OF THE PREDICTION MODEL

The prediction target of this model is the flight delay time, leading to two prediction approaches: direct prediction and indirect prediction. Direct prediction involves forecasting the flight delay duration based on known feature variables. Indirect prediction requires first identifying the calculation basis for flight delay time, determining parameters related to its calculation, predicting these parameters through the model, and subsequently calculating the flight delay duration. Both direct and indirect predictions have their advantages and disadvantages. Firstly, direct prediction demands higher requirements for feature variables since it predicts a single result, necessitating the identification of feature variables that can accurately predict the outcome, which can be challenging. Secondly, indirect prediction, due to having multiple prediction results, may lead to the accumulation of prediction errors. Through extensive experiments, it has been found that direct prediction struggles to achieve satisfactory predictive performance, while indirect prediction exhibits higher accuracy. Therefore, the final method adopted is the indirect prediction approach. The Random Forest prediction model is illustrated in Fig 7.

In Fig7, the random forest prediction process mainly includes three parts: dataset splitting, training and Testing. In dataset splitting, the dataset was divided into two sub-datasets. As one of the sub-datasets, training dataset accounts for 50% of the total dataset, and another 50% of the total dataset are testing dataset. Random Forests regression (RFR) is a class of decision tree-based machine learning algorithms [26]. During the process of training the model using the training dataset, the majority of the data will be directly incorporated and fitted through various decision trees. Additionally, 10% of the data is reserved for assessing whether the model exhibits overfitting.

The criterion chosen for evaluating the regression quality of decision trees is MSE (Mean Squared Error). The calculation formula for MSE is

$$\mathrm{MSE} = \frac{1}{m}\sum_{1}^{m}\left(y_{f\,i} - y_{ci}\right)^2 (i = 1, 2 \cdots n)\,. \qquad (3)$$

In the formula: m is the number of nodes in one decision tree, y_{fi} is the value of father node, y_{ci} is the value of child node. In the selection process of each node in the decision tree, nodes with smaller MSE are considered to have higher regression quality.

Each tree is constructed by randomly choosing a fixed number of feature subsets from three features. The trees grow maximally without any pruning process. Ultimately, each tree obtains its corresponding prediction results, and based on the proportion of the prediction results, the most reasonable result is obtained and used as the final prediction. Throughout this process, parameters are continuously adjusted according to the learning curve to achieve the optimal results.

The primary purpose of the Testing Dataset is to validate the model's effectiveness. The decision trees used in the

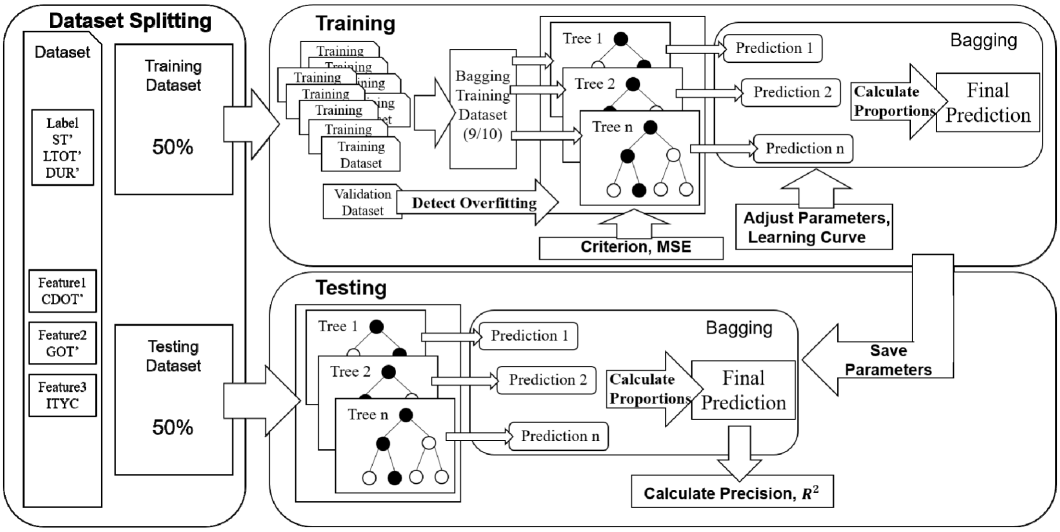

FIGURE 7. Random forest prediction process.

testing phase do not need further training and directly use the trees determined during the training phase. Similarly, the parameters used in bagging are those saved during the training process. Once the final predictions for the testing dataset are obtained, the accuracy of the model can be calculated. This paper employs the coefficient of determination (R-squared) to evaluate the model. The formula for calculating R-squared is

$$R^2 = 1 - \frac{\sum_1^n \left(\hat{y}_i - y_i\right)^2}{\sum_1^n \left(\overline{y_i} - y_i\right)^2} (i = 1, 2, \cdots, n). \tag{4}$$

In the formula: n is the number of datasets, y_i is raw data, $\hat{y}_i$ is forecast data, $\overline{y_i}$ is the mean of raw data. The range of R-squared values is [0, 1]. If the result is 0, it indicates a poor fit of the model, while a result of 1 signifies a model without errors.

The training and testing processes are mutually correlated and validated. The correlation lies in the fact that the most suitable parameters obtained during the training process are directly applied to testing. Therefore, the testing process can directly verify the training effect. To further improve the adaptability of the model, the training set and the testing set are non-overlapping, ensuring that the model can adapt to new datasets. It is precisely because of the difference in datasets that when evaluating the training effect of the model using R-squared, the results of the testing set are always slightly lower than those of the training set. Therefore, when evaluating the quality of the model, attention should be paid to the two R-squared results, but emphasis should be placed on the R-squared of the training set.

C. CALCULATION PROCESS

The features in ACDM are either categorical, time-related, or numerical. Zoutendijk et al. proposed three encoding methods for the above features: target encode, trigonometrically encode, and numerically encode [27]. This paper employs feature encoding methods, including target encode and numerically encode, and introduces a time transformation encoding method for time-related features. The core principle of the time transformation encoding method is to establish a time baseline and convert calendar time into duration. The time baseline used in this paper is the in-block time t_{ib}.

Based on the data provided by ACDM, the calculation formula for flight delay time (FDT) is

$$\mathrm{FDT}' = t_{\mathrm{ATOT}} - t_{LTOT}. \tag{5}$$

The calculation formula for converting Last Takeoff Time (LTOT) into duration is

$$LTOT' = t_{LTOT} - t_{ib}. \tag{6}$$

The calculation formula for the duration time flight stay in airport is

$$DUR' = t_{ATOT} - t_{ib}. \tag{7}$$

The calculation formula for flight service time is

$$ST' = t_{ob} - t_{ib}. \tag{8}$$

The calculation formula to convert cabin door open Time (CDOT) into duration is

$$CDOT' = t_{CDOT} - t_{ib}. \tag{9}$$

The calculation formula for converting gate open time (GOT) into duration is

$$GOT' = t_{GOT} - t_{ib}. \tag{10}$$

In formula (5)-formula (10), $(\cdot)'$ is the duration, and the unit is minutes, $t_{(\cdot)}$ is calendar time and can be obtained directly in ACDM. Based on formulas (5), (6), and (7), it can be derived that

$$\mathrm{FDT}' = DUR' - LTOT'. \tag{11}$$

According to (8) and (11), the labels for the prediction model are determined as ST', DUR', and $LTOT'$.The selection of feature variables is achieved through extensive experiments, leading to the determination of three feature variables: DUR', $CDOT'$, and ITYC. The former two can be calculated using (7) and (9), while ITYC (Identity Code) requires target encoding.

ITYC is the encoding for ITY in ACDM. Referring to Table 1, 1TY corresponds to the aircraft type code. In the data considered in this paper, the closest relationship with the aircraft type is the service time. Therefore, the average service time for aircraft types in Fig 5 is taken as the original data for target encoding, and the final determined ITYC is presented in Table 5.

TABLE 5. ITYC comparison table.

ITY	B737	B738	A20N	A319	A320	A321	A332
ITYC	0.86	0.9342	1.214	1.2883	1.1835	1.3938	1.742
ITY	A333	A359	B772	B773	B77W	B788	B789
ITYC	1.8341	1.2227	1.0654	0.815	2.3623	1.1229	2.3638

D. PREDICTION RESULTS

The random forest model construction was implemented using the sklearn toolkit in Python, followed by parameter settings. After extensive experiments and tuning, the final parameters are shown in Table 6.

TABLE 6. Random forest model parameters.

Term	Value
test_size	0.5
random_state	1
max_depth	20
n_estimators	1000
n_jobs	-1

The total data set used in this paper is the operational data of Shanghai Pudong Airport in June 2023 (excluding June 20). The original dataset consists of 1092 entries, and after processing, a total of 974 usable data entries remains. In the model training process, the test set accounts for 50% of the dataset. Using the parameters from Table 6, the model's performance is presented in Table 7.

TABLE 7. Random forest model performance.

Features	Label	R^2_train	R^2_test
DUR', $CDOT'$, ITYC	ST'	0.997	0.985
	DUR'	0.979	0.911
	$LTOT'$	0.996	0.974

Based on the results in Table 7, the R-squared for the training set is close to 1, indicating minimal model error. For the test set, except for DUR', the R-squared values for the other two labels are also close to 1. Generally, for a well-trained model, an R-squared value exceeding 0.8 is considered good training performance. Therefore, the model's prediction for DUR' is significantly higher than the standard.

Next, this model is used to predict the delay status of 20 valid flights on June 20th.

The selected 20 flights in this paper have arrival times ranging from 7:35 to 18:40, covering all transit flights on that day except for parked flights. In addition, from 15:28 to 15:55, there are five flights arriving at the airport, indicating a relatively high flight density. Therefore, considering both the time span and flight density, the selected data is representative.

Using the trained model to predict the delay for the above flights, the results are shown in Fig 8.

In Fig 8, the line chart indicates that the predicted values closely align with the actual values. The bar chart shows that the errors are within ±9 minutes. The maximum positive error in this prediction is approximately 6.7 minutes, the minimum positive error is around 1 minute, the maximum negative error is approximately -8.2 minutes, and the minimum negative error is around -0.3 minutes. The error span (positive error minus negative error) is maintained between 1.3 minutes and 14.9 minutes. According to the Civil Aviation Administration of China's definition of a flight delay of 15 minutes [28], the accuracy of this prediction result is 100%. Even when limiting the time to within ±8 minutes, the accuracy of this prediction model is as high as 95%. It should be noted that the accuracy rates presented in this paper are based on the premise of "defined delay time," and thus are higher than the true accuracy rates of the random forest algorithm. The reason for not using true accuracy rates in this paper is to maintain consistency throughout the text and facilitate the subsequent use of this data in the program. The delay standards adopted in this paper are all in accordance with the 15-minute rule specified in reference [28], wherein if the actual delay time of a flight is within 15 minutes, it is considered a normal flight; however, if the actual delay time exceeds 15 minutes, then the flight is defined as delayed.

IV. GROUND HANDLING PROCESS OPTIMIZATION MODEL

According to the random forest model mentioned earlier, accurate predictions of flight delays can be made. However, prediction is not the ultimate goal; preventing flight delays is what airlines most anticipate. In [19], six main categories of the reason for flight delays were allocated: Rotation, ATFM/ATC, Airport Authorities, Handling, Technical, and Weather. These major categories cover up to 85% of potential flight delays. However, most of these reasons are beyond control, except for handling. Between 5% and 10% of the primary delay sources can be estimated as GH (Ground Handling) related, which is not negligible and may contribute significantly to airport performance [29]. Therefore, this paper, in conjunction with the delay prediction results, optimized the ground handling processes for key flights.

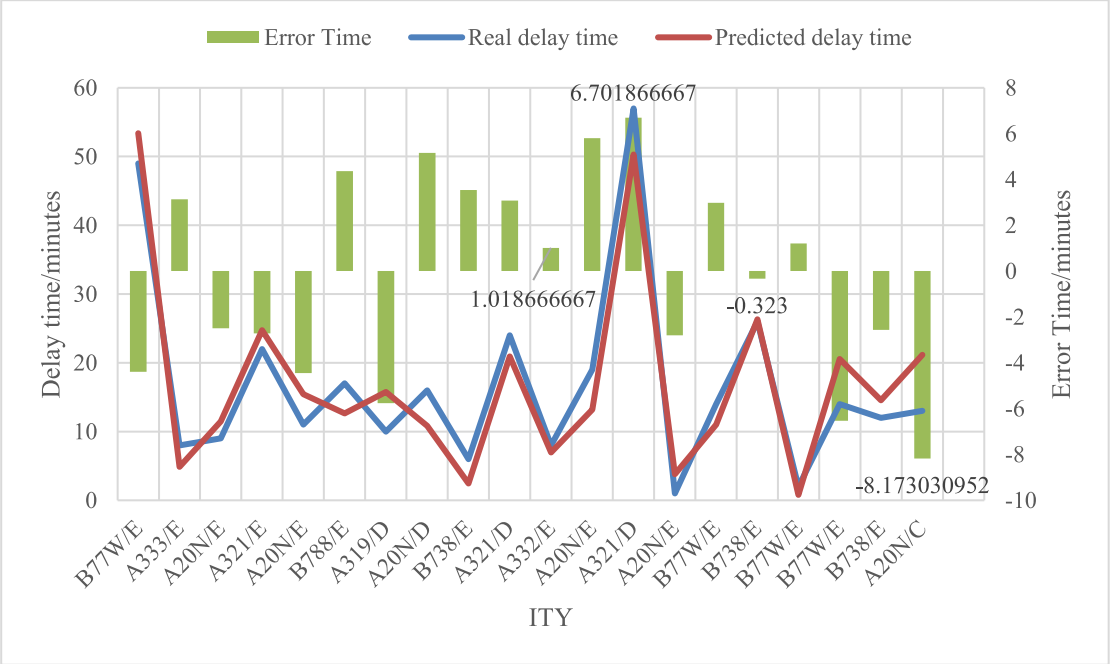

FIGURE 8. Comparison of flight delay prediction results.

A. GROUND HANDLING PROCESSES

Analyzing the flight operations at Shanghai Pudong Airport, the entire process of a flight operation can be divided into four dimensions: Aircraft, Passengers, Baggage, and Cargo. For general passenger flights, the focus is primarily on the dimensions of Aircraft, Passengers, and Baggage. The basic flow of flight operations is illustrated in Fig 9.

Through Fig 9, it can be observed that for general flights, the main ground handling processes occur between "In Block" and "Off Block." In the arrival process, "In Block," "Covered Bridge Docking," and "Open Cabin Door" are arranged sequentially over time, while "Cleaning," "Potable Water," and "Waste Water" are parallel processes without a defined order. In the departure process, "Close Cabin Door," "Off Block," and "Line Up" are arranged sequentially over time, while "Catering," "Refueling," "Maintain," and "Boarding" are parallel processes without a defined order. For ground handling vehicles, most of their services are arranged in parallel, so they are not significantly constrained by the order of time and can be treated as a whole for single-service analysis.

In addition to the mentioned service processes, the ground handling process also needs to consider another aspect: the movement of ground handling vehicles [20]. This process refers to the movement of ground handling vehicles between the garage and parking positions or between parking positions. Therefore, optimizing this process is equivalent to optimizing vehicle routing, resembling the Vehicle Routing Problem (VRP). The schematic diagram of parking positions involved in this paper is referenced from Fig 4, where the spatial relationships can determine the distances between different parking positions.

B. GENETIC ALGORITHM

Regarding algorithmic solutions to path optimization problems, the mainstream approaches include Genetic Algorithms (GA), Simulated Annealing, and Particle Swarm Optimization (PSO). Among them, Particle Swarm Optimization is an evolutionary computation technique that seeks optimal solutions through collaboration and information sharing among individuals in a swarm. PSO's main drawbacks are slow convergence speed and susceptibility to local optima. Simulated Annealing is a stochastic search method suitable for solving large-scale combinatorial optimization problems and can yield global optimal solutions. Its main drawbacks include long solution times and challenges in determining the initial values and step sizes for temperature T. Genetic Algorithm is a method that simulates the natural evolution process to search for optimal solutions. The algorithm's individual selection is stochastic, and it can avoid local optima through mutation mechanisms. Its main drawback lies in the complexity associated with gene encoding and decoding.

The ground handling process optimized in this paper is prone to entering local optima because each flight is independent. Once one or several flights are optimized, the overall flight operations are defined as "smooth," leading to a few flights unable to undergo optimization. This issue arose when using PSO, hence this algorithm was ultimately not selected. As mentioned above, although Simulated Annealing can avoid local optimization problems, it is mainly oriented towards large-scale combinatorial optimization and has slow

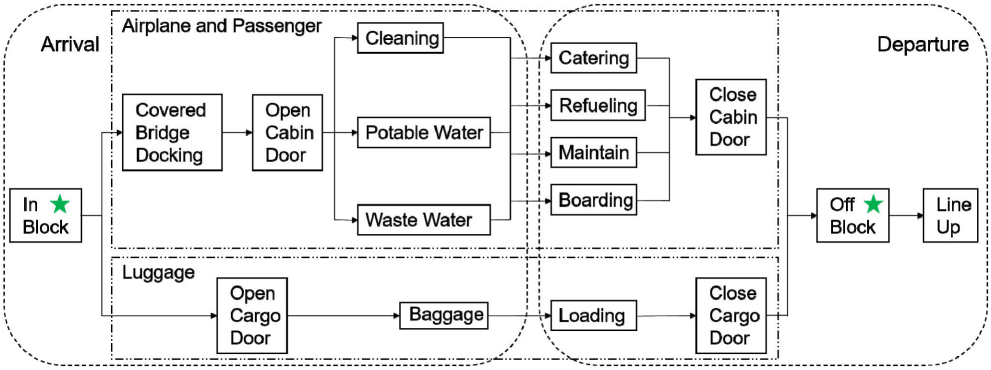

FIGURE 9. Flight operation process.

computational speed. Ground handling optimization involves combinations of flights and ground service vehicles, which do not fall under large-scale combinations, hence Simulated Annealing is not considered. Upon comparing these three methods, it is evident that, given the ability to address the complexity of genetic algorithm encoding, this algorithm's advantages in solving path optimization problems are clear. Therefore, aligning with the choices of many researchers, this paper also employs Genetic Algorithm to solve the ground handling optimization problem [10], [12], [20]. Additionally, many researchers have combined these algorithms with other computing technologies, deriving some more comprehensive algorithms, but these lack universality.

1) OVERVIEW OF GENETIC ALGORITHM

Genetic Algorithms [30] are a class of the artificial intelligence methods which are based on a Darwinian notion of natural selection and the survival of the fittest, and Mendel's principles of genetic transmission (the passage of alleles from one generation to the next) [31]. Genetic algorithms can be seen as a virtualization of reality and transplanting ideas from the natural to the artificial. The solutions of an optimization problem are treated as individuals of a biological system and the value of the objective function of a solution is the quality measure (called fitness in the GA literature) of an individual [32]. The basic flow of the genetic algorithm is shown in Fig 10.

According to Fig 10, the implementation process of genetic algorithms includes generating an initial population, evaluating and selecting, genetic operations, and stopping criteria. The following is a brief description of the implementation methods in the aforementioned processes.

(i) Generate initial population. In this process, the key issues are population size and encoding methods. The population size is determined based on requirements, where larger sizes yield better computational results but increase computation time. The encoding method is crucial, determining the form of population individuals. Common encoding methods include binary encoding, integer encoding, and real-number encoding.

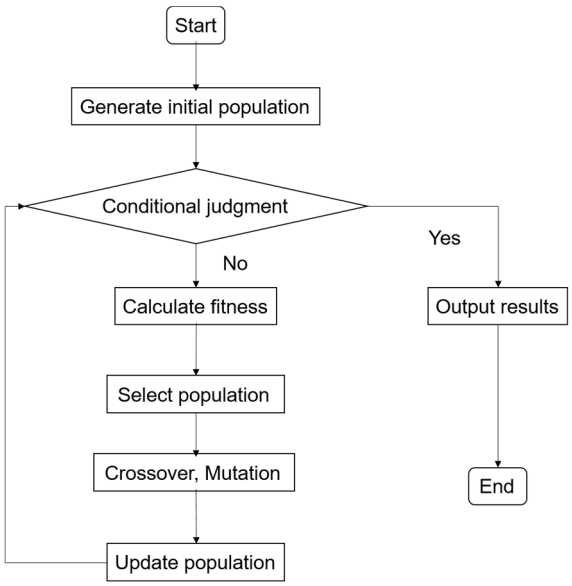

FIGURE 10. Flow chart of genetic algorithm.

(ii) Evaluation and Selection. The evaluation process primarily involves calculating the fitness of individuals in the population, and selection is based on fitness to identify individuals that meet the criteria. The fitness function is related to the problem's objective function, where higher fitness indicates stronger adaptability, leading to a higher probability of being selected.

(iii) Genetic Operations. Genetic operations mainly simulate gene changes in the biological evolution process, involving two operators: crossover and mutation. After crossover or mutation, parents are ''merged'' (mated) and produce offspring. If the offspring is better than the worst population member and is not identical to another population member, it is added to the population and the worst population member is removed. Otherwise, the population is unchanged.

(iv) Stopping Criteria. For single-objective optimization projects, stopping criteria are generally defined by the target value and the maximum number of generations. In the case of multi-objective optimization projects, as the directions

of different objectives are usually challenging to unify, the maximum number of generations is commonly used as the stopping criterion.

2) CHROMOSOME ENCODING

Chromosome encoding and decoding are critical issues in the application of genetic algorithms, where a rational and effective encoding contributes to solving optimization problems. However, encoding and decoding are complex and multifaceted problems. Gen and Cheng [33] had introduced various encoding schemes for different application problems. Detailed descriptions of scheduling problem solutions can be found in [34]. Drawing upon the descriptions in the mentioned literature and incorporating the encoding scheme proposed in [20], this paper introduces an improved chromosome real-number encoding scheme, outlined as follows.

(i) The chromosome consists of two parts: the first half is used to generate flight service information, and the second half is used to generate service start times. Flight service information includes two components: service vehicle number and service order, represented by $x_{i,j}$ Service start time is represented by $st_{i,j}$.

(ii) The length of the chromosome is 2N, where N is the number of flights.

(iii) The gene positions for $x_{i,j}$ and $st_{i,j}$ are segmented and correspond to the arrival flight index i.

(iv) The integer part of $x_{i,j}$ represents the ground service vehicle number j, and the decimal part represents the service order of service vehicle j. The larger the number, the later the service order.

According to the above encoding scheme, the final chromosome is formulated as:

$$\mathrm{Y} = [\mathrm{X}, \mathrm{ST}] = [x_1, x_2, \cdots .x_N, st_1, st_2, \cdots, st_N] \tag{12}$$

The information contained in chromosome (12) indicates that four flights are serviced by two ground service vehicles, with Vehicle A responsible for flights 1 and 3, and Vehicle B responsible for flights 2 and 4. The initial service times for flights 1 and 3 are 8 and 35, and for flights 2 and 4 are 15 and 60.

3) GROUND SERVICE OPTIMIZATION MODEL

To address the ground service optimization problem using genetic algorithms, a corresponding mathematical model needs to be established. The ultimate goal of this optimization problem is to prevent flight delays by efficiently arranging existing ground service vehicles and planning the service flights and service order. The decision variables and their symbolic meanings involved in the model-building process are summarized in Table 8.

The mathematical model established in this paper is as follows:

$$\min Z_1 = \sum_{i=1}^{N} d_i \tag{13}$$

$$\min Z_2 = \sqrt{\sum_{i=1}^{N} \left(d_i - \bar{d}\right)^2 / \sum_{i=1}^{N} sign\left(d_i\right)} \tag{14}$$

$$s.t \sum_{j=1}^{M} g_{i,j} = 1 \quad \forall i \tag{15}$$

$$st_{i,j} \geq a_i \quad \forall i \tag{16}$$

$$t_i + tr_{k,k'} \leq st_{i'} + 100000 \left(1 - g_{i,j} \cdot g_{i',j} \cdot \operatorname{sign}\left(f_{i',j} - f_{i,j}\right)\right) \forall i, i', j, k, k' \tag{17}$$

$$t_i = \left(st_{i,j} + p_{i,j}\right) \cdot e_i \quad \forall i, j \tag{18}$$

$$d_i = \max\left\{t_i - l_i, 0\right\} \quad \forall i \tag{19}$$

$$\bar{d} = \sum_{i=1}^{N} d_i / \sum_{i=1}^{N} sign\left(d_i\right). \tag{20}$$

Among them, $sign(\cdot)$ used in the model is a binary function, and its expression is:

$$sign(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0. \end{cases} \tag{21}$$

In the above model, (13) and (14) represent the objective functions, making this model a multi-objective function model. Objective function (13) aims to minimize the total flight delay time, while objective function (14) aims to minimize the standard deviation of flight delay time. Equations (15) to (20) represent the constraint conditions. Constraint (15) ensures that each flight is serviced by one and only one ground-support vehicle. Constraint (16) ensures that the service start time is later than the arrival time of the flight. Constraint (17) ensures that for two flights serviced by the same ground-support vehicle, the service start time of the current flight is later than the end time of the previous flight plus the transfer time. Constraint (18) defines the service completion time, taking into account the impact of the delay coefficient to strengthen the optimization for flights predicted to be delayed. Constraint (19) defines the delay time, which is 0 if the flight is not delayed. Constraint (20) represents the mean delay time.

C. OPTIMIZATION PROCESS

The optimization process of the model based on the genetic algorithm is illustrated in Fig 11.

In Fig 11, the optimization process of the model mainly involves three stages: Generate Initial Population, Iterative and Analysis, and Evaluation. In the Generate Initial Population stage, the Optimization Object is first determined as flights and ground-support vehicles, and three decision variables are extracted: $g_{i,j}$, $st_{i,j}$, and $f_{i,j}$. The Iterative stage involves substituting decision variables into the mathematical model, calculating the objective function values, continuously applying crossover and mutation to decision variables, and iteratively calculating until the specified number of generations is reached. The Analysis and Evaluation stage analyzes the Pareto front to find the optimal solution and evaluates the comprehensive performance of the

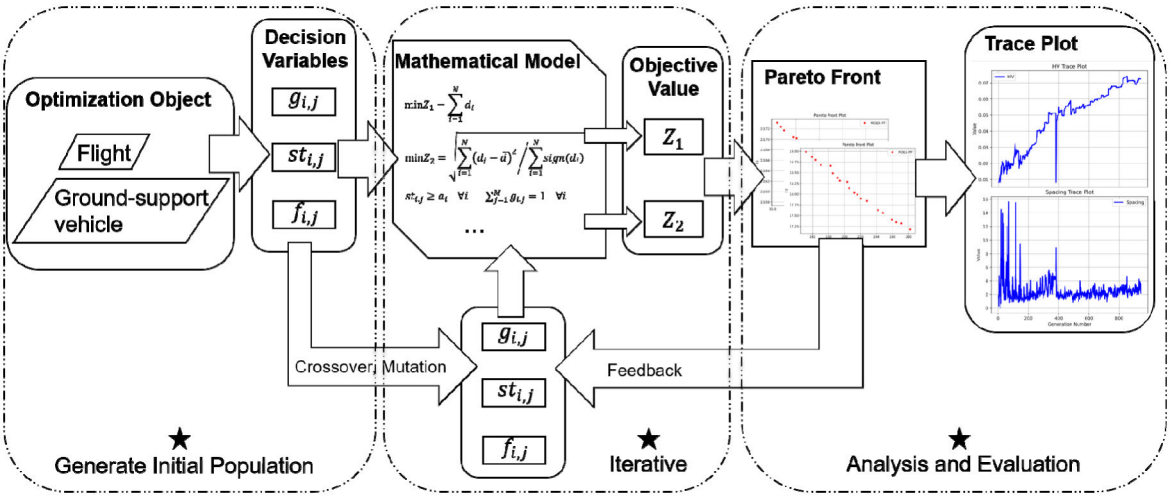

FIGURE 11. Optimization process based on genetic algorithm.

TABLE 8. Decision variables and symbols.

Symbol	Definition
N	Total number of flights
M	Total number of service vehicles
R	Number of aircraft types
i	Flight number arranged according to arrival time, $\forall i = 1,2,3,\cdots$, N
j	Ground service vehicle number, $\forall j = 1,2,3,\cdots$, M
k_i	The parking space number of flight i. The value of k refers to the parking space number in Table 4.
r_i	The aircraft type serial number of flight i, generally an integer starting from 1
e_i	Delay coefficient of flight i,$\forall e_i = 0.96,0.98,0.99,1$
a_i	Arrival time of flight i
l_i	Scheduled departure time of flight i
$p_{i,j}$	Time taken by service vehicle j to serve flight i
$tr_{k,k'}$	The time it takes for service vehicles to transfer between parking bays k and k'
$g_{i,j}$*	If service vehicle j serves flight i, then $g_{i,j} = 1$, otherwise $g_{i,j} = 0$
$st_{i,j}$*	The start time of service vehicle j for flight i
$f_{i,j}$*	The order in which service vehicle j serves flight i
t_i*	Service completion time for flight i
d_i*	Delay time of flight i, non-delayed flight $d_i = 0$
$\bar{d}$*	Average delay time for all flights

*-marked decision variables

genetic algorithm using two metrics: Hyper volume (HV) and Spacing.

The computation of the optimization model relies on three main sources of data: random population, ACDM data, and random forest prediction data. The random population generates decision variables, and their ranges are determined based on the number of ground-support vehicles and flight arrival/departure times. Random forest prediction data is crucial for obtaining the delay coefficient e_i, which is related to the predicted delay time.

Based on the results from Chapter III (Flight Delay Prediction Model), the random forest model predicts the delay time for specific flights. To tailor ground services optimization for delayed flights, a delay coefficient is introduced. The relationship between the delay coefficient and flight delay time is presented in Table 9.

TABLE 9. Delay coefficient comparison.

Delay time (minutes)	Delay coefficient (e_i)
0-15	1
16-30	0.99
31-45	0.98
above 46	0.96

The delay coefficients are experimentally determined based on the final optimization results and are not involved in the calculation process.

Apart from the mentioned data, all other required data for the optimization model are obtained from ACDM. Now, combining the earlier prediction results, the optimization for 20 selected flights on June 20th is conducted to validate the outcomes of this study.

V. APPLICATION EFFECT EVALUATION

The 20 selected flights in this study experienced delays from a time perspective, ranging from a minimum delay of 1 minute to a maximum delay of 57 minutes. From the perspective of the delay definition, flights with delays less than 15 minutes are not considered delayed. Therefore, among the selected flights, 8 experienced delays, representing a relatively high proportion. To gain a clearer understanding of the flight operations, a flight process chart has been created, as depicted in Fig 12.

As per the analysis of decision variables in Chapter IV (Ground Handling Process Optimization Model), the value of decision variable $g_{i,j}$ is determined by the total number of ground service vehicles. Through experimental testing, the minimum number of service vehicles that this model can solve is found to be 6, which is less than the actual number of service vehicles. Therefore, based on the chromosome coding

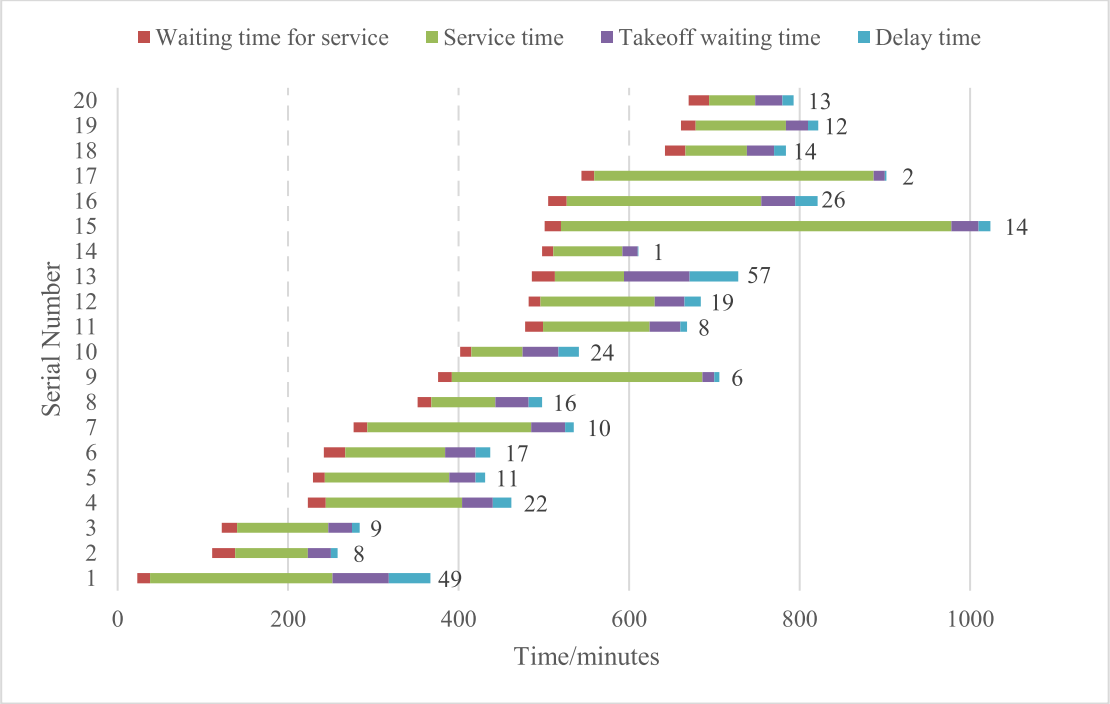

FIGURE 12. Flight ground service progress chart.

discussed in Section IV (B (2)), the lower and upper bounds of the chromosome for this project are:

$$
\begin{aligned}
Y_{low} = [&1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\\
&1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\\
&23\ \ 111\ \ 122\ \ 223\ \ 229\ \ 242\ \ 277\ \ 352\ \ 376\ \ 402\\
&478\ \ 482\ \ 486\ \ 498\ \ 501\ \ 505\ \ 544\ \ 642\ \ 661\ \ 670]\\
Y_{up} = [&7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\\
&7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\ \ 7\\
&236\ \ 198\ \ 226\ \ 380\ \ 369\ \ 370\ \ 473\ \ 426\ \ 663\ \ 455\\
&596\ \ 606\ \ 576\ \ 569\ \ 963\ \ 726\ \ 869\ \ 723\ \ 755\ \ 727].
\end{aligned}
$$

Note that the ST field content in Y_{up}corresponds not to the departure time of the flight but rather the departure time minus the average waiting time (refer to Fig 6). Using this time has the advantage of incorporating airport traffic conditions, providing more stringent conditions for service planning. The upper and lower bounds of the chromosome are both open intervals.

The 20 flights involved in this study belong to 8 different aircraft types, numbered as shown in Table 10.

TABLE 10. Identity serial number.

ITY	B738	A20N	A319	A321	A332	A333	B77W	B788
Serial number	1	2	3	4	5	6	7	8
Amount	3	6	1	3	1	1	4	1

Based on the known data, the flight basic information matrix D_f is constructed to facilitate the use of flight-related information in calculations. The information contained in the flight basic information matrix is:

$$
D_f = \begin{bmatrix} r_i \\ k_i \\ a_i \\ l_i - T_{tow} \end{bmatrix}. \tag{22}
$$

According to (22), the flight basic information matrix is determined as (23), shown at the bottom of the next page.

Since ground service vehicles need to move from the completion of service for one flight to the parking position of another flight, there is a transfer time involved. The calculation of transfer time includes the driving time within the parking position and the driving time on the route.

The driving time within the parking position refers to the time taken by the ground service vehicle to leave the old parking position area after completing the service and enter the new parking position area. As the size of the parking position has a minor impact on this time, it is considered a fixed value. The calculation is as follows:

$$
t_{TAR} = \frac{s_1}{v_1} = \frac{60 * 60}{1000 * 5}(min) = 0.72(min). \tag{24}
$$

In (24), the total distance covered by the ground service vehicle when leaving the old parking position and entering the new parking position is 60 m, and the driving speed within the parking position is 5 km/h.

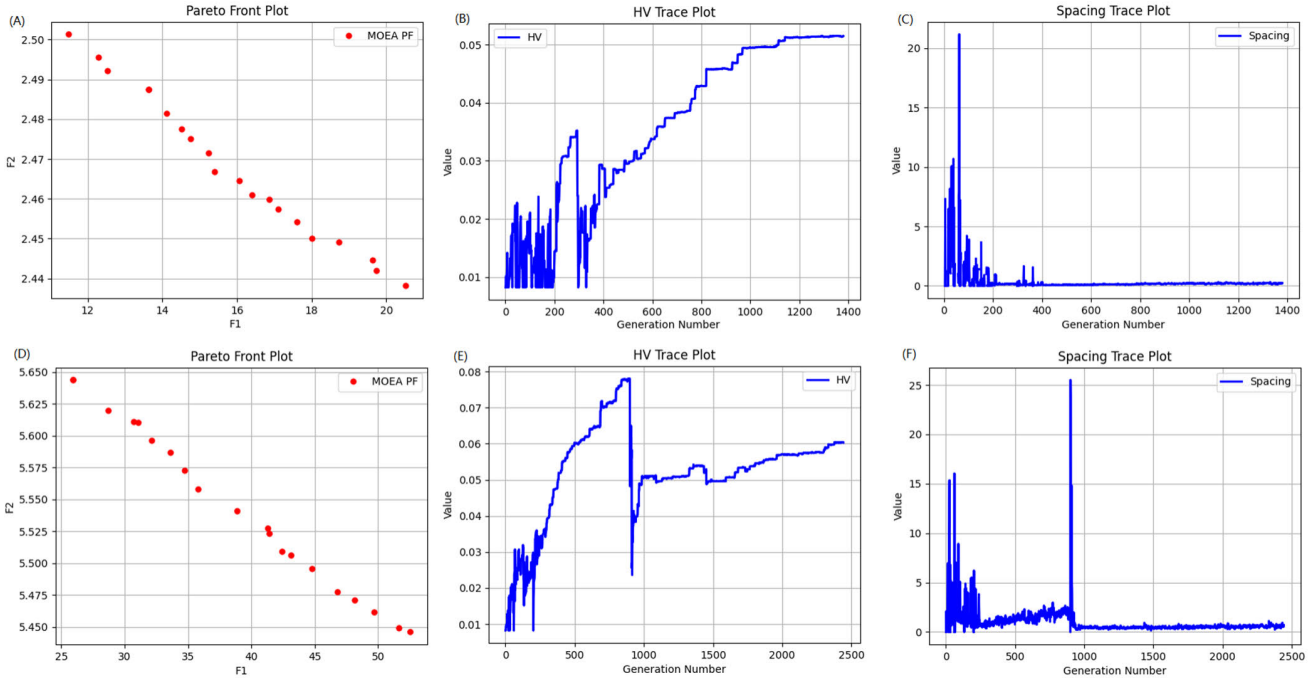

FIGURE 13. Comparison of optimization model results.(A) Pareto front plot with 1500 generations; (B) HV trace plot with 1500 generations; (C) Spacing trace plot with 1500 generations; (D) Pareto front plot with 2500 generations; (E) HV trace plot with 2500 generations; (F) Spacing trace plot with 2500 generations.

The driving time on the route refers to the time spent by the ground service vehicle traveling between two parking positions. The formula for calculating this time is:

$$t_{IT} = \frac{s_2}{v_2} = \frac{s_2 * 60}{30 * 1000}(min) = \frac{s_2}{500}(min). \tag{25}$$

The transfer time for the ground service vehicle is calculated as:

$$t_{TRVS} = t_{TAR} + t_{IT}. \tag{26}$$

Calculating the transfer time of ground service vehicles based on (24), (25), and (26) yields the transfer time matrix (27), as shown at the bottom of the next page.

The service time for ground service vehicles is mainly related to the aircraft type. In the preceding text, the service times for all flights in June were statistically analyzed, with results referenced from Fig 5. The service times used in this paper are the Maximum Time and Average Time from Fig 5. Average Time, being relatively shorter, best reflects the true service time for the majority of flights. However, if unexpected situations occur, Average Time may be insufficient. Therefore, Maximum Time is introduced to ensure that flights have enough time to complete the service work under any unforeseen circumstances.

TABLE 11. Identity serial number.

ITY serial number	Special service vehicle service time(minutes)	Ordinary service vehicle service time(minutes)
1	93.42	40
2	121.4	45
3	128.83	46
4	139.38	46
5	174.2	67
6	183.41	73
7	236.23	62
8	112.29	78

Special service vehicles 1#, 3#, and 5# use Maximum Time, designated as special service vehicles, while ordinary service vehicles 2#, 4#, and 6# use Average Time, designated as ordinary service vehicles. The allocation of service times for ground service vehicles is shown in Table 11.

Plugging in the above parameters into the optimization model built using NSGA-II in Python, with a population size of 20, the run data for maximum genetic generations of 500, 1000, 1500, and 2500 are shown in Table 12.

$$\mathrm{D_f} = \begin{bmatrix} 7 & 6 & 2 & 4 & 2 & 8 & 3 & 2 & 1 & 4 & 5 & 2 & 4 & 2 & 7 & 1 & 7 & 7 & 1 & 2 \\ 14 & 8 & 6 & 7 & 5 & 12 & 2 & 3 & 8 & 1 & 10 & 4 & 3 & 5 & 11 & 7 & 13 & 9 & 4 & 15 \\ 23 & 111 & 122 & 223 & 229 & 242 & 277 & 352 & 376 & 402 & 478 & 482 & 486 & 498 & 501 & 505 & 544 & 642 & 661 & 670 \\ 236 & 198 & 226 & 380 & 369 & 370 & 473 & 426 & 663 & 455 & 596 & 606 & 576 & 569 & 963 & 726 & 869 & 723 & 755 & 727 \end{bmatrix} \tag{23}$$

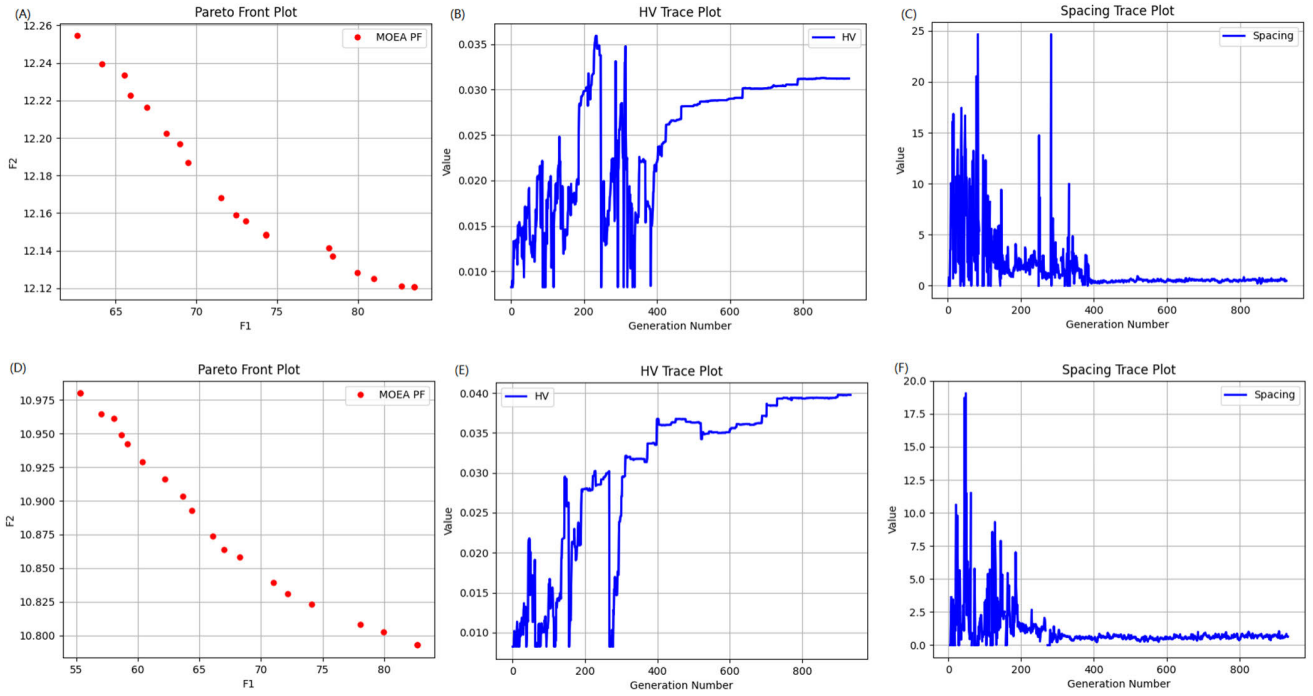

FIGURE 14. Comparison of optimization results for the 1000 generation model. (A) Pareto front plot without delay coefficients; (B) HV trace plot without delay coefficients; (C) Spacing trace plot without delay coefficients; (D) Pareto front plot with delay coefficients; (E) HV trace plot with delay coefficients; (F) Spacing trace plot with delay coefficients.

TABLE 12. Optimization results without delay coefficient.

MAXGEN	HV	Spacing	minZ_1	minZ_2	Solutions
500	0.03843	1.11303	145	17	20
1000	0.03121	0.46771	62.59	12.12	20
1500	0.05146	0.23217	12.3	2.44	20
2500	0.06031	0.67977	25.9	5.45	20

From Table 12, it can be observed that the model with a maximum genetic generation of 1500 achieves the best optimization performance. The total delay time is minimized to 12.3 minutes, with a standard deviation of 2.44. Prior to this model, with an increase in genetic generations, the optimization performance improved, but afterward, the optimization performance declined. The optimization process for models with 1500 and 2500 generations is compared in Fig 13.

In Fig 13(A and D), elevating the number of generations from 1500 to 2500 resulted in an increase in both output values of the training model's objective

$$TRVS = \begin{bmatrix} 0 & 0.836 & 1.068 & 1.524 & 1.674 & 2.134 & 2.434 & 2.596 & 2.896 & 3.046 & 3.204 & 3.354 & 3.156 & 3.826 & 4.616 \\ 0.836 & 0 & 0.952 & 1.408 & 1.558 & 2.018 & 2.318 & 2.48 & 2.78 & 2.93 & 3.088 & 3.238 & 3.4 & 3.71 & 4.5 \\ 1.068 & 0.952 & 0 & 1.2 & 1.35 & 1.81 & 2.11 & 2.272 & 2.572 & 2.722 & 2.88 & 3.03 & 3.192 & 3.502 & 4.292 \\ 1.524 & 1.408 & 1.2 & 0 & 0.87 & 1.33 & 1.63 & 1.792 & 2.092 & 2.242 & 2.4 & 2.55 & 2.712 & 3.022 & 3.812 \\ 1.674 & 1.558 & 1.35 & 0.87 & 0 & 1.18 & 1.48 & 1.642 & 1.942 & 2.092 & 2.25 & 2.4 & 2.562 & 2.872 & 3.662 \\ 2.134 & 2.018 & 1.81 & 1.33 & 1.18 & 0 & 1.02 & 1.182 & 1.482 & 1.632 & 1.79 & 1.94 & 2.102 & 2.412 & 3.202 \\ 2.434 & 2.318 & 2.11 & 1.63 & 1.48 & 1.02 & 0 & 0.882 & 1.182 & 1.332 & 1.49 & 1.64 & 1.802 & 2.112 & 2.902 \\ 2.596 & 2.48 & 2.272 & 1.792 & 1.642 & 1.182 & 0.882 & 0 & 1.02 & 1.17 & 1.328 & 1.478 & 1.64 & 1.95 & 2.74 \\ 2.896 & 2.78 & 2.572 & 2.092 & 1.942 & 1.482 & 1.182 & 1.02 & 0 & 0.87 & 1.028 & 1.178 & 1.34 & 1.65 & 2.44 \\ 3.046 & 2.93 & 2.722 & 2.242 & 2.092 & 1.632 & 1.332 & 1.17 & 0.87 & 0 & 0.882 & 1.032 & 1.194 & 1.504 & 2.294 \\ 3.204 & 3.088 & 2.88 & 2.4 & 2.25 & 1.79 & 1.49 & 1.328 & 1.028 & 0.882 & 0 & 0.87 & 1.032 & 1.342 & 2.132 \\ 3.354 & 3.238 & 3.03 & 2.55 & 2.4 & 1.94 & 1.64 & 1.478 & 1.178 & 1.032 & 0.87 & 0 & 0.87 & 1.18 & 1.97 \\ 3.516 & 3.4 & 3.192 & 2.712 & 2.562 & 2.102 & 1.802 & 1.64 & 1.34 & 1.194 & 1.032 & 0.87 & 0 & 1.032 & 1.822 \\ 3.826 & 3.71 & 3.502 & 3.022 & 2.872 & 2.412 & 2.112 & 1.95 & 1.65 & 1.504 & 1.342 & 1.18 & 1.032 & 0 & 1.542 \\ 4.616 & 4.5 & 4.292 & 3.812 & 3.662 & 3.202 & 2.902 & 2.74 & 2.44 & 2.294 & 2.132 & 1.97 & 1.822 & 1.542 & 0 \end{bmatrix} \quad (27)$$

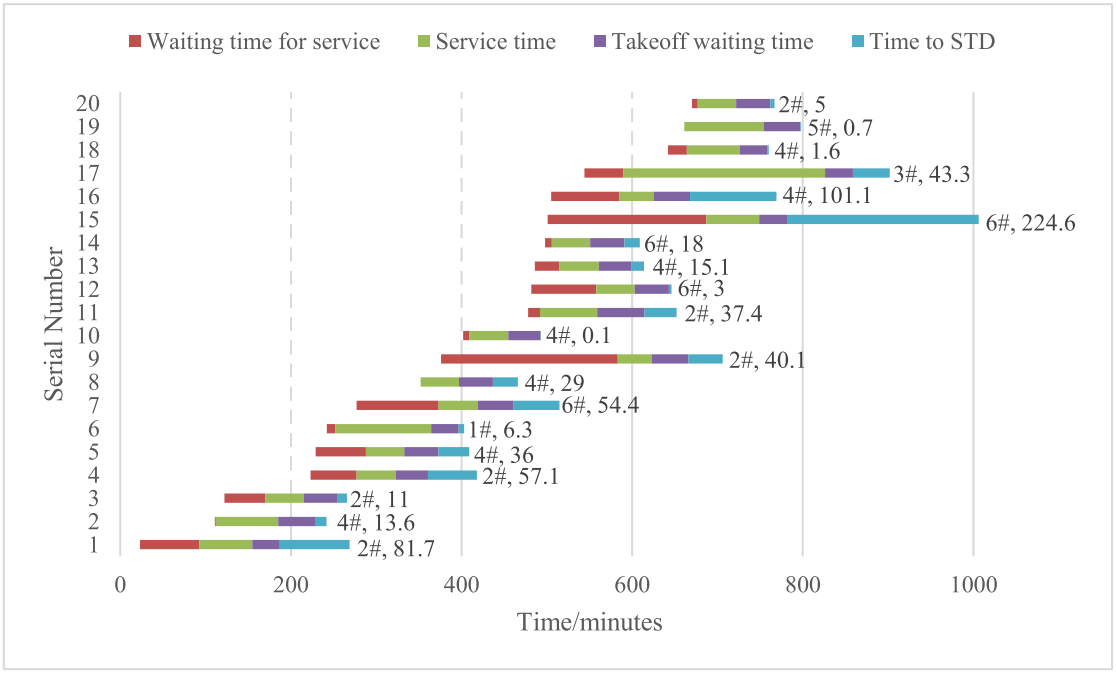

FIGURE 15. Optimized flight ground service progress chart.

functions, indicating a deterioration in the training performance. In Fig 13(B, C, E and F), the optimization model for 2500 generations encounters an error at around 900 generations, leading to an increase in optimization results instead of a decrease. Adjusting the model between 1500 and 2500 generations did not yield results superior to the 1500 generations. Therefore, without using correction factors, this model failed to reduce delays to a lower level, e_i, as shown at the bottom of the page.

Next, the correction factors (delay coefficients) proposed in this paper are applied to adjust the model. Based on the prediction results from Chapter III (refer to Fig 8) and Table 9, the delay coefficients for flights are determined as follows:

These coefficients are then plugged into (18) to complete the adjustment of the mathematical model. Subsequently, NSGA-II is used again to optimize the model with unchanged population size and a maximum genetic generation set to 1000. The results are compared with the model run for 1000 generations without delay coefficients, as shown in Fig 14.

In Fig 14, the results in the lower part correspond to the model output with delay coefficients. both objective function values are smaller with delay coefficients (comparing A and D), and the Hyper Volume (HV) is larger (comparing B and E). Moreover, it requires fewer generations to achieve stable output (comparing C and F). Therefore, the model with delay coefficients surpasses the model without delay coefficients. Nevertheless, under the condition of 1000 generations, the results in Fig 14 represent the optimal solution obtained from multiple runs. The minimum total delay time is 55.28 minutes, with a standard deviation of 10.79.

Increasing the maximum genetic generation to 1500 and continuing to run the model with delay coefficients yielded excellent optimization results. Under this condition, the minimum total delay time and standard deviation were both reduced to 0, and there were 20 non-dominated solutions. One set of optimal solutions is chosen as follows:

$$\begin{aligned} X = [&2.11 \quad 4.19 \quad 2.26 \quad 2.80 \quad 4.35 \quad 1.26 \quad 6.56 \quad 4.38 \quad 2.91 \quad 4.47 \\ &2.85 \quad 6.84 \quad 4.53 \quad 6.61 \quad 6.99 \quad 4.73 \quad 3.11 \quad 4.86 \quad 5.86 \quad 2.98] \\ ST = [&92.8 \quad 112 \quad 170 \quad 277 \quad 288 \quad 252 \quad 373 \quad 352 \quad 583 \quad 409 \\ &492 \quad 558 \quad 515 \quad 506 \quad 687 \quad 585 \quad 590 \quad 660 \quad 661 \quad 677]. \end{aligned}$$

Decoding X provides information on the flights and their sequence serviced by the ground service vehicles. The service start time can be obtained based on ST. Combined with the previously discussed process data, the optimized flight process can be determined. The optimized flight process is shown in Fig 15.

In Fig 15, the information on the right side is displayed in the format "Service vehicle number #, Time to STD". It is evident from the figure that service vehicle 4# serves the most flights, a total of 7 flights. Additionally, due to the longer service time of special service vehicles

$$e_i = \begin{bmatrix} 0.96 & 1 & 1 & 0.99 & 1 & 1 & 1 & 1 & 1 & 0.99 & 1 & 1 & 0.96 & 1 & 1 & 0.99 & 1 & 0.99 & 1 & 0.99 \end{bmatrix}$$

(1#, 3#, 5#), the number of flights assigned for service is relatively fewer, with each serving only 1 flight. A comparison with Figure 12 reveals a significant improvement in the ground service process after optimization, with no delays occurring for any of the flights. Notably, flights 1 (delay of 49 minutes) and 13 (delay of 57 minutes), which experienced substantial actual delays, both took off more than 15 minutes ahead of schedule after optimization. The comparison results indicate that the predictive model incorporating the correlation flight delay coefficient can effectively optimize the flight ground service process, mitigate flight delays, and enhance the on-time performance of airport flights.

VI. CONCLUSION

This paper has successfully achieved all the planned objectives. The following summarizes the main work and achievements of this paper.

(i) For the first time, this paper proposes integrating flight delay prediction results with a flight ground service optimization model. It designs the flight delay coefficient e_i, which serves as a link in optimizing flights predicted to have delays, demonstrating effective results.

(ii) A flight delay prediction model is established based on the random forest algorithm. The model utilizes an indirect prediction approach, employing three feature variables to predict two labels, and subsequently calculates the flight delay time. This model achieves a prediction accuracy of 100% according to the 15-minute delay standard.

(iii) The study investigates the real-number encoding method for decision variables and designs a novel chromosome $Y = [X, ST]$. This chromosome contains information on the time and sequence of ground service vehicle services for corresponding flights, establishing rules for generating a random population.

(iv) A mathematical model expressing the optimization problem of ground service vehicles is established. This model uses total delay time and delay time standard deviation as objective functions, formulates corresponding constraints, and introduces the delay coefficient e_i to correct flight service times.

(v) A ground service optimization model is developed based on NSGA-II, addressing the dual-objective optimization problem proposed in this paper.

(vi) Using flight operation data from Shanghai Pudong Airport in June as an example, the entire model is tested, and the results are compared with actual operational data. The study proves that the proposed method can accurately predict flight delays and prevent them through optimization of ground service processes.

This model is intended to be applied to the ground service vehicle dispatch management system at airports. The following discussions elaborate on its deployment and maintenance at airports:

(i) The training data for this model is sourced from the airport's ACDM system, eliminating the need for additional sensor deployment to collect data and significantly reducing the airport's investment costs. As the model is still in the development stage, offline data packages from the ACDM are currently utilized. However, once deployed at the airport, the model can access ACDM data online for real-time prediction and optimization.

(ii) The model requires a dedicated computer for computation, with minimum hardware specifications including an Intel i5 processor or above, 8GB RAM, Intel graphics630 or above, and a 1TB hard drive.

(iii) The final optimization results are currently outputted in offline format. However, after completing the airport deployment, the model can be directly integrated into the dispatch management system to facilitate real-time scheduling information updates.

(iv) The maintenance cost of this system is relatively low, primarily focusing on software maintenance. Operational personnel are required to possess certain software debugging capabilities for maintenance purposes.

These are the current research achievements. In the next phase of the research, the focus will be on the following three aspects:

(i) As the aviation industry exhibits significant fluctuations throughout the year, this study has concentrated on flight operations for one month. The next step will involve adjusting the model to enable accurate prediction and optimization over an entire year.

(ii) The classification of ground service vehicles was not considered in this paper. Different types of ground service vehicles were simply merged for analysis and treatment. However, for airport scheduling, there is greater guidance significance if ground service vehicles can be subdivided according to their service types. Therefore, in the subsequent research plan, a more in-depth investigation into this aspect will also be conducted.

(iii) Integrating the model into the airport operations and maintenance system involves testing the model's performance, adjusting and improving the model based on the test results. To accomplish this research task, certain hardware design is also required to achieve optimal performance of the model.

REFERENCES

[1] CIRIUM. (Sep. 15, 2023). *The On-Time Performance Monthly Report-Airlines*. [Online]. Available: https://resources.cirium.com/monthly-otp-2023

[2] Z. Zhao, S. Feng, M. Song, and Q. Liang, "A delay prediction method for the whole process of transit flight," *Aerospace*, vol. 9, no. 11, p. 645, Oct. 2022, doi: 10.3390/aerospace9110645.

[3] R. B. Wu, T. Zhao, and J. Y. Qu, "Flight delay prediction model based on deep SE-Dense Net," *J. Electron. Inf. Technol.*, vol. 41, no. 6, pp. 1510–1517, Jun. 2019, doi: 10.11999/JEIT180644.

[4] E. Esmaeilzadeh and S. Mokhtarimousavi, "Machine learning approach for flight departure delay prediction and analysis," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 8, pp. 145–159, Jul. 2020, doi: 10.1177/0361198120930014.

[5] B. Ye, B. Liu, Y. Tian, and L. Wan, "A methodology for predicting aggregate flight departure delays in airports based on supervised learning," *Sustainability*, vol. 12, no. 7, p. 2749, Apr. 2020, doi: 10.3390/su12072749.

[6] D. B. Bisandu, I. Moulitsas, and S. Filippone, "Social ski driver conditional autoregressive-based deep learning classifier for flight delay prediction," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8777–8802, Jan. 2022, doi: 10.1007/s00521-022-06898-y.

[7] Z. Wang, C. Liao, X. Hang, L. Li, D. Delahaye, and M. Hansen, "Distribution prediction of strategic flight delays via machine learning methods," *Sustainability*, vol. 14, no. 22, p. 15180, Nov. 2022, doi: 10.3390/su142215180.

[8] X. Wang, Z. Wang, L. Wan, and Y. Tian, "Prediction of flight delays at Beijing capital international airport based on ensemble methods," *Appl. Sci.*, vol. 12, no. 20, p. 10621, Oct. 2022, doi: 10.3390/app122010621.

[9] J. C. Shi, "Research on deep-learning based departing flight delay prediction and low-cost prevention and control strategy," M.S. thesis, College Civil Aviation Saf. Eng., Civil Aviation Flight Univ. China, Guanghan, China, 2022.

[10] D. A. Tabares, F. Mora-Camino, and A. Drouin, "A multi-time scale management structure for airport ground handling automation," *J. Air Transp. Manage.*, vol. 90, Jan. 2021, Art. no. 101959, doi: 10.1016/j.jairtraman.2020.101959.

[11] A. Rajapaksha and D. N. Jayasuriya, "Smart airport: A review on future of the airport operation," *Global J. Manage. Bus. Res.*, vol. 20, no. 3, pp. 25–34, 2022. [Online]. Available: https://www.researchgate.net/profile/Aruna-Rajapaksha/publication/339800592_Smart_Airport_A_Review_on_Future_of_the_Airport_Operation/links/5e708375299bf14570f291cb/Smart-Airport-A-Review-on-Future-of-the-Airport-Operation.pdf

[12] W. H. Ip, D. Wang, and V. Cho, "Aircraft ground service scheduling problems and their genetic algorithm with hybrid assignment and sequence encoding scheme," *IEEE Syst. J.*, vol. 7, no. 4, pp. 649–657, Dec. 2013, doi: 10.1109/JSYST.2012.2196229.

[13] Z. Xing, B. Li, and Q. Luo, "Estimation of ground service time for transit flight," in *Proc. Asia–Pacific Conf. Intell. Med. Int. Conf. Transp. Traffic Eng.*, Sichuan, China, Dec. 2018, pp. 223–228.

[14] F. Tang, R. Zhang, and S. Liu, "Airport ground service oriented multi-service coordination scheduling algorithm," *Control Eng. China*, vol. 27, no. 10, pp. 1686–1692, Oct. 2020, doi: 10.14107/j.cnki.kzgc.20180647.

[15] G. Zhang, "Research on algorithms of airline crew scheduling," M.S. thesis, College Air Traffic Manag., Civil Aviation Flight Univ. China, Guanghan, China, 2021.

[16] A. K. Agogino and K. Tumer, "A multiagent approach to managing air traffic flow," *Auto. Agents Multi-Agent Syst.*, vol. 24, no. 1, pp. 1–25, Jan. 2012, doi: 10.1007/s10458-010-9142-5.

[17] J. C. Bansal, H. Sharma, S. S. Jadon, and M. Clerc, "Spider monkey optimization algorithm for numerical optimization," *Memetic Comput.*, vol. 6, no. 1, pp. 31–47, Jan. 2014, doi: 10.1007/s12293-013-0128-0.

[18] B. Yu, Z. Guo, S. Asian, H. Wang, and G. Chen, "Flight delay prediction for commercial air transport: A deep learning approach," *Transp. Res. E, Logistics Transp. Rev.*, vol. 125, pp. 203–221, May 2019, doi: 10.1016/j.tre.2019.03.013.

[19] H. Fricke and M. Schultz, "Delay impacts onto turnaround performance," presented at ATM, Napa, CA, USA, Jul. 2009.

[20] F. Tang, "Special vehicle scheduling and its optimization algorithm in airport ground services," Ph.D. dissertation, College Inf. Sci. Eng., Northeastern Univ., Shenyang, China, 2019.

[21] IATA. (Dec. 2022). *Understanding the Pandemic's Impact on the Aviation Value Cain*. [Online]. Available: https://www.iata.org/en/iata-repository/publications/economic-reports/understanding-the-pandemics-impact-on-the-aviation-value-chain/

[22] IATA. (Dec. 2022). *Global Outlook for Air Transport Sustained Recovery Amidst Strong Headwinds*. [Online]. Available: https://www.iata.org/en/iata-repository/publications/economic-reports/global-outlook-for-air-transport—december-2022/

[23] CAAC. (Sep. 24, 2021). *Technical Guidelines for Epidemic Prevention and Control at Transport Airports (Eighth Edition)*. [Online]. Available: https://www.caac.gov.cn/XWZX/MHYW/202109/t20210924_209338.html

[24] Y. Horiguchi, Y. Baba, and H. Kashima, "Predicting fuel consumption and flight delays for low-cost airlines," in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 4–9.

[25] CAAC. (Sep. 18, 2021).*Aerodrome Technical Standards*. [Online]. Available: https://www.caac.gov.cn/XXGK/XXGK/BZGF/HYBZ/202112/P020220922567165385570.pdf

[26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/a:1010933404324.

[27] M. Zoutendijk and M. Mitici, "Probabilistic flight delay predictions using machine learning and applications to the Flight-to-Gate assignment problem," *Aerospace*, vol. 8, no. 6, p. 152, May 2021, doi: 10.3390/aerospace8060152.

[28] CAAC. (May 20, 2016). *Regulations on Normal Flight Management*. [Online]. Available: https://www.caac.gov.cn/XXGK/XXGK/MHGZ/201706/t20170621_44917.html

[29] A. T. Diego, "A contribution to automation of airport ground operations," Ph.D. dissertation, Dept. Comput. Sci. Telecommun., Univ. Toulouse III, Toulouse, France, 2018.

[30] J.-H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: Univ. Michigan, 1975, pp. 123–135.

[31] S. Freeman, M. Harrington, and J. Sharp, *Biological Science*, 2nd ed., Toronto, London, U.K.: Pearson, 2012, pp. 510–532.

[32] Z. Drezner and T. D. Drezner, "Biologically inspired parent selection in genetic algorithms," *Ann. Oper. Res.*, vol. 287, no. 1, pp. 161–183, Jul. 2019, doi: 10.1007/s10479-019-03343-7.

[33] R. Cheng, M. Gen, and Y. Tsujimura, "A tutorial survey of job-shop scheduling problems using genetic algorithms—I. representation," *Comput. Ind. Eng.*, vol. 30, no. 4, pp. 983–997, Sep. 1996, doi: 10.1016/0360-8352(96)00047-2.

[34] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Design*. New York, NY, USA: Wiley, 1996, pp. 234–261.

ZHEN-TENG XU received the M.S. and Ph.D. degrees from the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2016 and 2022, respectively. He is currently a Lecturer with the School of Aeronautic Engineering, Nanjing Vocational University of Industry Technology, Nanjing. His research interests include aircraft fault diagnosis, algorithm application, and digital geometry processing.

YAN-JUN LI received the B.S. and Ph.D. degrees from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1991 and 2005, respectively. He is currently a Professor and a Doctoral Supervisor with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics. His research interests include aircraft fault diagnosis, airworthiness safety certification, and health monitoring.

HONG-FU ZUO received the M.S. degree in mechanics from China University of Mining and Technology, Beijing, China, in 1985, and the Ph.D. degree in mechanical engineering from China University of Mining and Technology, Jiangsu, China, in 1989. Since 1996, he has been a Professor with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics. His research interests include mechanical condition monitoring, fault diagnosis, and lubricating oil wear debris monitoring.

TENG-ZHOU XU received the Ph.D. degree from the College of Materials Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2018. He is currently an Associate Professor with the School of Aeronautic Engineering, Nanjing Vocational University of Industry Technology, Nanjing. His research interests include aerospace composite materials and vacuum insulation materials.

QI WANG received the B.S. degree from the Nanhang Jincheng College, Nanjing, China, in 2013, and the M.S. degree from Nanjing University of Aeronautics and Astronautics, Nanjing, in 2016. He is currently an Engineer with Shanghai Airport (Group) Company Ltd., Shanghai, China. His research interests include runway safety and airport operation support.

WANG-WANG YU received the M.S. and Ph.D. degrees from Nanjing Forestry University, Nanjing, China, in 2011 and 2018, respectively. He is currently an Associate Professor with Nanjing Vocational University of Industry Technology, Nanjing. His research interests include development of high performance 3D printing materials, recycling, and reuse of waste plastics.

HONG-SHENG YAN received the B.S. degree in traffic engineering from Yangzhou University, Yangzhou, China, in 2014, and the M.S. and Ph.D. degrees in vehicle utilization engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2016 and 2022, respectively. He is currently a Lecturer with the School of Aeronautic Engineering, Nanjing Vocational University of Industry Technology, Nanjing. His research interests include condition monitoring, diagnosis, and health management of the aircraft based on the flight data.

BING LIU was born in Shandong, China. He received the M.S. degree from Henan University of Technology, Zhengzhou. He is currently an Associate Professor with the Public Foundational Courses Department, Nanjing Vocational University of Industry Technology. His research interests include statistics and data analysis.

TAO CHEN received the Ph.D. degree from the College of Materials Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2021. He is currently a Lecturer with the School of Aeronautic Engineering, Nanjing Vocational University of Industry Technology, Nanjing. His research interests include aerospace composite materials and advanced connection technology.

MAO-HUI ZHOU received the B.S. and M.S. degrees from the College of Mechanical Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, in 2019 and 2022, respectively. He is currently pursuing the Ph.D. degree in transportation engineering with Nanjing University of Aeronautics and Astronautics. His research interests include data-driven fault diagnosis and health monitoring.