

Article

Improved LS-SVM Method for Flight Data Fitting of Civil Aircraft Flying at High Plateau

Nongtian Chen ^{1,2}, Youchao Sun ^{1,*} , Zongpeng Wang ¹ and Chong Peng ¹

¹ College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China; chennongtian@nuaa.edu.cn (N.C.); wangzp@nuaa.edu.cn (Z.W.); chong_peng@nuaa.edu.cn (C.P.)

² College of Aviation Engineering, Civil Aviation Flight University of China, Guanghan 618307, China

* Correspondence: sunyc@nuaa.edu.cn

Abstract: High-plateau flight safety is an important research hotspot in the field of civil aviation transportation safety science. Complete and accurate high-plateau flight data are beneficial for effectively assessing and improving the flight status of civil aviation aircrafts, and can play an important role in carrying out high-plateau operation safety risk analysis. Due to various reasons, such as low temperature and low pressure in the harsh environment of high-plateau flights, the abnormality or loss of the quick access recorder (QAR) data affects the flight data processing and analysis results to a certain extent. In order to effectively solve this problem, an improved least squares support vector machines method is proposed. Firstly, the entropy weight method is used to obtain the index weights. Secondly, the principal component analysis method is used for dimensionality reduction. Finally, the data are fitted and repaired by selecting appropriate eigenvalues through multiple tests based on the LS-SVM. In order to verify the effectiveness of this method, the QAR data related to multiple real plateau flights are used for testing and comparing with the improved method for verification. The fitting results show that the error measurement index mean absolute error of the average error accuracy is more than 90%, and the error index value equal coefficient reaches a high fit degree of 0.99, which proves that the improved least squares support vector machines machine learning model can fit and supplement the missing QAR data in the plateau area through historical flight data to effectively meet application needs.

Keywords: least squares method; support vector machines; principal component analysis; quick access recorder; mean absolute error; high-plateau flight



Citation: Chen, N.; Sun, Y.; Wang, Z.; Peng, C. Improved LS-SVM Method for Flight Data Fitting of Civil Aircraft Flying at High Plateau. *Electronics* **2022**, *11*, 1558. <https://doi.org/10.3390/electronics11101558>

Academic Editor: Gyu Myoung Lee

Received: 24 March 2022

Accepted: 10 May 2022

Published: 13 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

High-plateau flights represent an important safety issue for civil aviation, especially for China's civil aviation transportation. High-plateau airports are mainly distributed in China, Nepal, Peru, Bolivia, Ecuador, and other countries. Among the 42 high-plateau airports in the world, 16 are located in China, so their operation safety problems have a profound impact on China's civil aviation [1]. On 14 May 2018, the flight mission of Chinese Sichuan Airlines flight 3U8633 from Chongqing to Lhasa plateau was an example of the typical unsafe event; the front windshield of the cockpit burst and fell off during the flight in high-plateau airspace, and the crew made an emergency descent. Compared with ordinary flight, high-plateau flight has low air density and atmospheric pressure, complex terrain, solar radiation, uneven heating of the terrain facing the sun, and many other environmental characteristics which result in stricter takeoff and landing conditions for aircrafts on high plateaus. The technical requirements of the personnel are more stringent and certain factors such as modification on the basis of ordinary civil aircrafts will cause the flight parameters of high-plateau civil airliners to change from those of civil airliners on general routes. During the entire flight phase, the quick access recorder (QAR) data may be abnormal or lost due to the influence of the high plateau's harsh environment,

detection equipment, transmission equipment, or other unknown conditions. QAR is an important data warehouse for post-flight flight technical analysis, engine health analysis, flight safety incident investigation, flight quality analysis, operational quality analysis, and aircraft health management. The abnormality of these data will bring inconvenience and hidden hazards for monitoring and analyzing the safety status of high-plateau flights for theoretical research.

Many scholars have carried out fruitful research on flight data analysis and application, mainly focusing on flight data processing, flight data application, and other application research. Flight data have many applications in aviation operation safety research [2–6]. Some scholars have applied flight data to turbine fault diagnosis, general aviation anomaly detection, aviation safety key landing index prediction [7–11], tower flight data manager man–machine system integration design processes, and new methods for nonlinear aerodynamic modeling of flight data [12–14]. Some scholars also analyze the flight characteristics of QAR data for landing at high-altitude airports, and use it for airline flight data monitoring machine learning methods, generating new operational safety knowledge from existing data, safety science insights gained from black-box-to-flight data monitoring, composite fault diagnosis using optimized MCKD and sparse representation of rolling bearings, rolling elements based on VMD, and sensitivity MCKD fault diagnosis, etc. [15–19]. Some scholars have carried out research on the impact of leveling operation on landing safety based on variance analysis of real flight data, civil aircraft hazard identification and prediction based on deep learning [20,21], unsteady aerodynamic modeling of unstable dynamic processes [22], and small-sample inspection data-driven diagnosis of critical deviation sources in aircraft structural assembly [23].

In the research of flight data processing methods and technologies, many scholars have also carried out a series of studies [24–26]. Some scholars have proposed improved binary gray wolf optimizer and support vector machine methods, arithmetic optimization algorithms, particle swarm optimization, average impact value-support vector machine algorithms, etc., for in-flight data processing and optimization [27–29]. Some scholars combined multiple classifiers to quantitatively sort the impact of anomalies in flight data based on frequency domain specification and improved particle swarm optimization algorithms, as well as enhanced fast non-dominated solution sorting genetic algorithms for multi-objective problems research [30–32].

In short, many scholars have carried out a series of researches on flight data collection and analysis, as well as application methods and technologies, and have also achieved many valuable results. However, research on high-altitude flight data is rare, especially research on the filling and simulation of flight data loss due to high altitude, low temperature, low pressure, and other elements of the special operating environments. To effectively solve the problem of high-plateau QAR flight data padding, an improved least squares support vector machines method is proposed. The entropy weight method is used to obtain the index weights, and the principal component analysis method is used for dimensionality reduction. The flight data are fitted and repaired by selecting appropriate eigenvalues through multiple tests based on LS-SVM. The data are fitted and repaired by selecting appropriate eigenvalues through multiple tests based on LS-SVM. In order to verify the effectiveness of this method, the QAR data related to multiple real plateau flights are used for testing and are compared with the improved method for verification.

2. Principle of Data Restoration Method

2.1. LS-SVM Principle

The support vector machine is a generalized linear classifier proposed to perform binary classification of data in a supervised learning manner. Its decision boundary is the maximum margin hyperplane for the learning sample solution. The basic principle is shown in Figure 1.

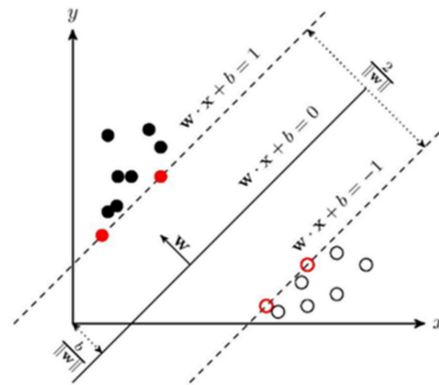


Figure 1. Support vector machine hyperplane conceptual model.

It is a machine learning method that is based on a complete statistical learning theory and has excellent learning capabilities. It has strict mathematical theory support, strong interpretability, and does not rely on statistical methods, thus simplifying the usual problems of classification and regression. It can also find key samples (support vectors) that are critical to the task. After adopting nuclear techniques, it can handle non-linear classification–regression tasks. The final decision function is determined by only a small number of support vectors and the complexity of the calculation depends on the number of support vectors, not the dimensionality of the sample space.

The LS-SVM demonstrates an improvement in the standard support vector machine, a new type of support vector machine method proposed by Suykens and Vandewalbe. Compared with the standard SVM, it replaces the inequality constraints in SVM with equality constraints, which increases the convergence speed, improves classification progress in problems with desired goals, and achieves good results [33].

Supposing the data training set of a given LS-SVM is expressed as (1)

$$(x_1, y_1), \dots, (x_n, y_n), x \in R_n, y \in \{-1, +1\} \tag{1}$$

$x_i \in R_n$ is the n -dimensional system input vector, $y_i \in R_n$ is the system output and $f(x) = \omega^T \varphi(x) + b$ is the unknown function to be estimated. Making a nonlinear mapping $\gamma: R_n \rightarrow H$, where Φ is called the feature map and H is the feature space, the unknown function is estimated to use the function of the form (2).

$$f(x) = \omega^T \varphi(x) + b \tag{2}$$

Among them, ω is the weight vector in R_n space, and $b \in R$ is the bias. The SVM algorithm uses the kernel function of the original space to replace the dot product operation in the high-dimensional feature space, avoids complex operations, and uses structural risk to minimize as a learning rule, which is mathematically described as $\omega^T \omega \leq \text{constant}$. The standard SVM algorithm takes the insensitive loss function as the structural risk minimization estimation problem. The meaning of the ϵ -insensitive loss function is as follows: when the difference between the observed value y of the x point and the predicted value $f(x)$ does not exceed the predetermined ϵ , it is considered that the predicted value $f(x)$ at this point is lossless, although the predicted value $f(x)$ and the observed value y may not be equal. On the other hand, LS-VSM chooses the second norm e^i of ζ_i as the loss function to make the equation true. Therefore, the optimization equation is established as (3) and (4).

$$\min_{\omega, b, e} (J\omega e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^N e^2, \gamma > \tag{3}$$

$$y_i = \omega^T \varphi(x_i) + b + e_i^2, i = 1, 2, \dots, N \tag{4}$$

Here, γ is a real constant which determines the relative size of $\frac{1}{2}\omega^T\omega$ and $\frac{1}{2}\sum_{i=1}^N e_i^2$, which can be between the training error and the compromised model complexity so that the function can seek better generalization ability. The LS-SVM algorithm defines a loss function that is different from the standard SVM algorithm and changes its inequality constraints to equality constraints, which can obtain ω in the dual space. The Lagrange Function (5) is as follows:

$$L(\omega, b, e, a) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N a_i\omega^T\varphi(x_i) + b + e_i - y_i \tag{5}$$

where $\alpha^i \in R, \alpha^i > 0$ is the Lagrange multiplier so the optimal solution condition is as follows (6):

$$\begin{aligned} \frac{\delta L}{\delta \omega} &= 0, \omega = \sum_{i=1}^N a_i\varphi(x_i) \\ \frac{\delta L}{\delta b} &= 0, \sum_{i=1}^N a_i = 0 \\ \frac{\delta L}{\delta e_i} &= 0, a_i = \gamma e_i \\ \frac{\delta L}{\delta a_i} &= 0, y_i = \omega^T\varphi(x_i) + b + e_i, i = 1, \dots, N \end{aligned} \tag{6}$$

After eliminating ω and e_i from Equation (6), this optimization problem is transformed into solving the following equation:

$$\begin{bmatrix} b \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & B + \gamma^{-1} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \gamma \end{bmatrix} \tag{7}$$

Among them, $y = [y_1, y_2, \dots, y_N]^T, a = [a_1, a_2, \dots, a_N]^T, 1 = [1, \dots, 1]^T$, and B represent a square matrix; the element in the i -th column and row j is $B_{ij} = \varphi(x_i)^T\varphi(x_j) = K(x_i, x_j), i, j = 1, \dots, N$; and $K(x_i, x_j)$ is the kernel function. On the basis of Formula (3), ω can be further obtained, so as to obtain the nonlinear approximation of the training data set

$$f(x) = \sum_{i=1}^N a_i K(x_i, x_j) + b \tag{8}$$

2.2. The Choice of Kernel Function

The kernel function is used to prevent the non-linear transformation from mapping its input space to the high-latitude space, causing particularly high-dimensional complex operations. When the support vector machine only needs the inner product operation and looks for a function that represents a low-dimensional input space that is exactly equal to the inner product in the high-dimensional space, the result can be obtained directly to avoid complicated operations. The choice of the kernel function requires Mercer’s theorem to be satisfied, that is, any Gram matrix of the kernel function in the sample space is a semi-positive definite matrix (semi-positive definite) [34]. Currently, the commonly used kernel functions in research and practice are as follows:

(1) Linear kernel function:

$$K(x, x_i) = x \cdot x_i \tag{9}$$

(2) Polynomial kernel function:

$$K(x, x_i) = (x \cdot x_i + 1)^d \tag{10}$$

(d value is the order of the polynomial)

(3) Radial basis kernel function:

$$K(x, x_i) = \exp\left(-\frac{(x - x_i)^2}{2\sigma^2}\right) \tag{11}$$

(4) B-spline kernel function:

$$K(x, x_i) = B_{2n+1}(x - x_i) \tag{12}$$

(5) Perceptual kernel function:

$$K(x, x_i) = \tanh(\beta x_i + b) \tag{13}$$

2.3. LS-SVM Principle

Entropy comes from physical thermodynamics and is one of the parameters that can characterize matter. It was first introduced into information theory by C.E. Shannon and called information entropy. The entropy weight method (EWM) abstracts information and tests its degree of variation through various eigenvalues. In this way, the weight of each feature is calculated and modified to achieve a more reasonable weight index [35]. The specific process is as follows:

(1) Perform data standardization processing on each feature value. Suppose that k feature quantities $Y_{ij} = \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)}$ are given, where $X_i = x_1, x_2, \dots, x_n$, assuming that the standardized value of each feature value is Y_1, Y_2, \dots, Y_K

$$Y_{ij} = \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{14}$$

(2) Find the information entropy of each eigenvalue. According to the definition of information entropy in information theory, the information entropy of a set of data can be written as

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^n Y_{ij}} \tag{15}$$

where $p_{ij} = \frac{Y_{ij}}{\sum_{i=1}^n Y_{ij}}$, if $\lim_{p_{ij} \rightarrow 0} \sum_{i=1}^n P_{ij} \ln P_{ij} = 0$, then define $\lim_{p_{ij} \rightarrow 0} \sum_{i=1}^n P_{ij} \ln P_{ij} = 0$, determine the weight w of each feature quantity:

$$w_i = \frac{1 - E_i}{k - \sum E_i} (i = 1, 2, \dots, k) \tag{16}$$

2.4. Principles of Principal Component Analysis (PCA)

The principal component analysis (PCA) method is currently the most widely used data dimensionality reduction algorithm. It aims to sequentially find a set of mutually orthogonal coordinate axes from the original high-dimensional space to determine its correlation by comparing the variance of the original data under the new coordinate axis; the degree is used to exclude zero-correlation or low-correlation feature quantities to achieve a dimensionality reduction of data features. Because of the efficiency and simplicity of PCA processing high-dimensional data sets, it is widely used in various fields in practice, especially in the field of compressed data [36].

2.5. Verification Method

In order to judge the conformity of the selected number of feature quantities, the coefficient of determination (R^2) is introduced. The coefficient of determination indicates how much the fluctuation of the dependent variable can be described by the fluctuation of the independent variable. Its expression is as follows:

$$R^2 = \left(\frac{\sum_{i=1}^n (y_i - \hat{y}) * (\hat{y}_i - \ddot{y})}{\sqrt{\sum_{i=1}^n (y_i - \hat{y})^2 * \sum_{i=1}^n (\hat{y}_i - \ddot{y})^2}} \right)^2 \tag{17}$$

y and \hat{y} represent the actual value and the predicted value of the simulation result. The closer the R^2 value is to 1, the better the correlation between the two.

For the evaluation of the complementation results, four commonly used indicators for data repair are introduced for analysis purposes: mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and equal coefficient (EC). The calculation is as follows:

$$\begin{aligned} \text{MSE} &= \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \\ \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \\ \text{MAE} &= \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \\ \text{EC} &= 1 - \frac{\sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^N y_i^2} - \sqrt{\sum_{i=1}^N \hat{y}_i^2}} \end{aligned} \quad (18)$$

y and \hat{y} still represent the actual value and predict the value of the simulation result, and N represents the number of samples in the training set. The smaller value of MSE, the higher the accuracy of the machine learning simulation results describing the experimental data. EC indicates the degree of fit between the output value and the true value. Generally, any value above 0.9 indicates a good fit.

3. Compensation Model and Simulation of High-Plateau Missing Data

The pseudo-compensation of missing data by other QAR data is essentially based on the existence of a certain functional relationship between QAR parameters. The value of the parameter can be derived from other parameter values. Therefore, the purpose of simulation is to determine this functional relationship. To be more specific, high-plateau flight data padding is essentially a function approximation problem.

This paper takes some flight parameters of QAR flight data as the assumed missing data in order to show the feasibility of this method. According to the actual meaning of the QAR data, the loss parameter $N(\tau) = N_{real}$ is set as the missing QAR parameter, where τ is the current moment of the missing data and other intact QAR parameters are used as the known vector set ω_τ^T according to the previous setting. Finding a functional relationship between the two or its first approximation such that $N(\tau) = N_{real}$, the relationship model can be written as $N(\tau) = \omega_\tau^T \varphi(x, t) + b$, where the parameter requirements are (8) the same, so LS-SVM can be used to complement the QAR loss parameters.

3.1. Data Selection

In order to verify the feasibility of the high-plateau QAR data patching, this paper collects ten flight data of a certain airline's civil transport aircraft in the same time period and the same origin and destination for simulation analysis. In order to reduce irrelevant external factors, interference data selection controls possible related variables, such as changing in crew members, and determines whether it is pre-flight or post-flight to ensure that the accuracy of the simulation is improved. After selection, nine groups were randomly selected as the model training group and the last group was used as the comparison group to test the accuracy of the experimental results.

3.2. Algorithm Improvement

Based on the support vector machine algorithm, an improved method is proposed for the shortcomings of difficulty in training and analyzing large-scale samples. The eigenvalue range definition plays a very important role in training. The input and output are put into a small range and then predicted by the support vector machine model. On the one hand, it can avoid overfitting caused by large-value data dominating small-value data. On the other hand, scaling the data to a small range can avoid the "dimension disaster" and reduce the computational load. The principal component analysis method, as a commonly used

dimensionality reduction algorithm, can easily simplify and refine complex data, process the data through the entropy method, and complete the algorithm optimization to achieve concise and accurate data under the premise of ensuring the robustness of the data.

3.3. Algorithm Flow

Before the simulation starts, it is necessary to determine the key parameters γ and the core width σ^2 in advance and then use the above algorithm to perform simulation training to fill in the missing data; the specific details and steps are shown in Figure 2.

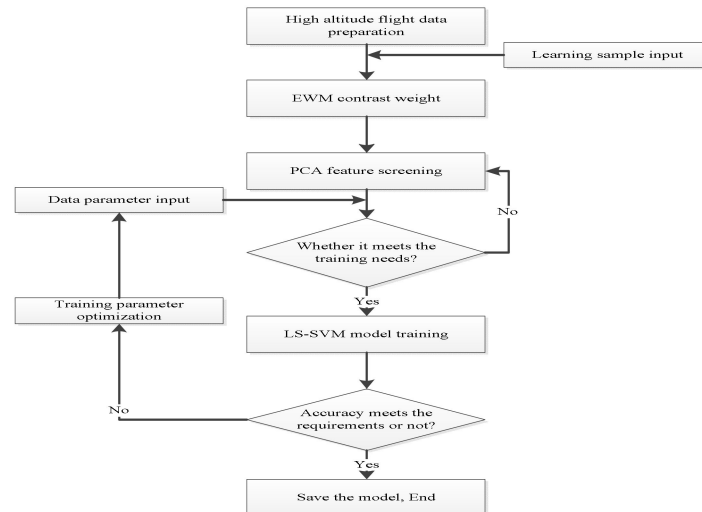


Figure 2. Flow chart of flight data fitting based on improved LS-SVM.

3.4. Simulation Application

QAR’s overall data cannot be analyzed due to the existence of text items and 78 data items remain after all text items are excluded. Python is taken as the expected environment, which measures the weight of each item through the EWM method and divides the interval to select the data items for simulation training. After multiple rounds of testing, the coefficient of determination is compared. It is found that when the number of feature quantities is smaller, the coefficient of fit is larger and the change tends to stably increase; thus, few features are prone to overfitting. After weighing and selecting the 17 feature items with the largest weight, they have good accuracy and credibility. The relationship between the number of specific features and the accuracy rate, as well as the weight ratio of the feature quantity, are shown in Figure 3.

Compared to the algorithm without the improved method, the improved algorithm not only improves the fitting effect but also greatly reduces the amount of data in the simulation. The fitting coefficient is increased by 0.64% but the amount of data calculation is reduced by 78.21%. The details are shown in Table 1.

Table 1. Performance table of improved method.

Characteristics of Several	No Improve	Improve	Promotion
R ²	0.991	0.9973	0.64%
The amount of data	778284	169626	78.21%

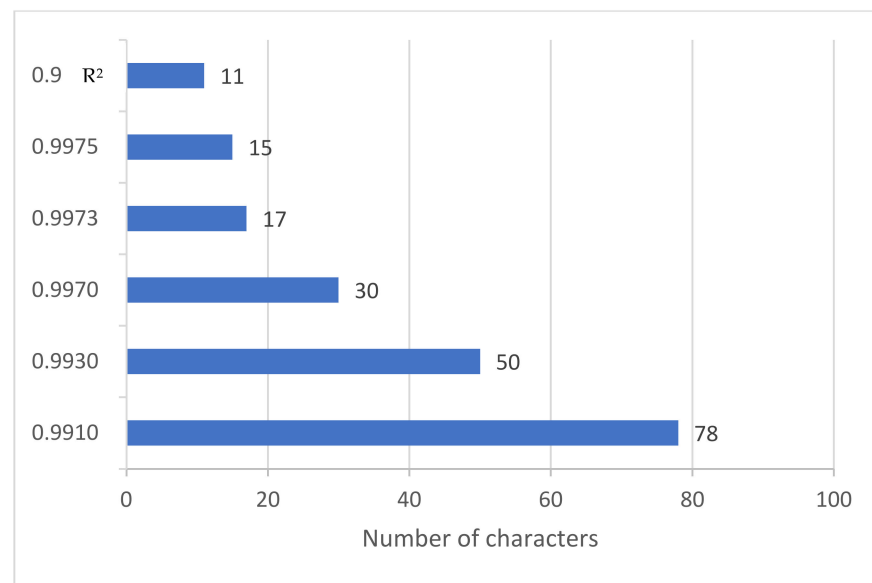


Figure 3. Relationship between the number of features and the correlation coefficient.

Among them, the selected feature quantities and the corresponding weights are shown in Table 2 and Figure 4.

Table 2. Weight-characteristic quantity correspondents.

Serial Number	Abbreviation	Name Connotation	Weight
1	N1_1	Left engine speed	0.041
2	N1_2	Right engine speed	0.042
3	N2_1	Left engine power	0.008
4	N2_2	Right engine power	0.011
5	FLIGHT_PHASE	Flight phase	0.054
6	GS1	True ground speed	0.082
7	GS2	Captain's instrument displays ground speed	0.083
8	GS_FO	The co-pilot's gauge shows ground speed	0.082
9	CAS	Calibrated air speed	0.079
10	DRIFT	Drift angle	0.041
11	TAS	True airspeed	0.099
12	PITCH11	The captain's instrument displays the pitch angle on the left side	0.064
13	PITCH12	The captain's instrument displays the pitch angle on the inner left side	0.064
14	PITCH21	The captain's instrument displays the pitch angle to the outer right	0.064
15	PITCH22	The captain's instrument displays the pitch angle on the inner right side	0.064
16	PITCH_DISP_FO1	The assistant captain's gauge shows the outer left side of the pitch angle	0.061
17	PITCH_DISP_FO2	The assistant captain's instrument displays the pitch angle on the inner left side	0.061

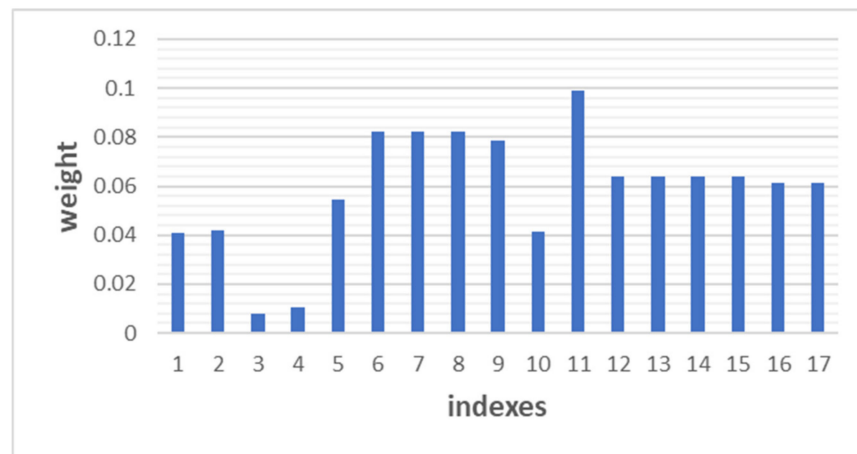


Figure 4. Weight-characteristic quantity correspondents.

Among them, the feature that has the greatest impact on the prediction is the true flight speed (TAS), and the feature that has the least impact is the right engine speed (N2_1). After determining the selection of the feature quantity, due to the large amplitude of the QAR data, in order to reduce the modeling error, the input data and the expected data were normalized on $[-1, 0]$ and $[0, 1]$, respectively. The original interval should be returned to after analysis. In this paper, the kernel function selects the most commonly used radial basis function for data repair:

$$K(x, x_i) = \exp\left(-\frac{(x - x_i)^2}{2\sigma^2}\right) \tag{19}$$

The simulation found that the parameters γ and the kernel width σ^2 have a significant impact on the complementation effect, which needs to be determined according to the specific characteristics of the training data. Generally speaking, a reduction in the kernel width σ^2 can improve the training accuracy but can reduce the generalization ability, and an increase in the parameter γ can also improve the training accuracy. The training shows that when the parameter $\gamma = 3$ and the training model is filled with missing data, the data with core width $\sigma^2 = 0.6$ have the best complementation effect. With the left engine speed (N1, unit: RPM), the aircraft pitch angle (pitch, unit: $^\circ$) and the flap angle (flap angle, unit: $^\circ$), as examples, intercept the data simulation results of the climb, approach, and landing stages to show the degree of flight data padding. In order to facilitate the analysis and observation, the predicted and actual values of the aircraft inclination angle are placed in $(-1,1)$ interval, the predicted value and actual value of the left engine speed are put in the $(1,3)$ interval, and the predicted value and actual value of the flap angle are put in the $(3,5)$ interval, as shown in Figures 5–7.

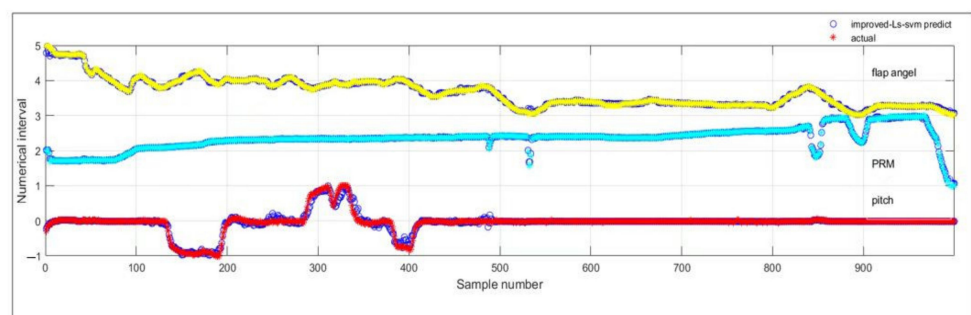


Figure 5. Climbing phase simulation diagram.

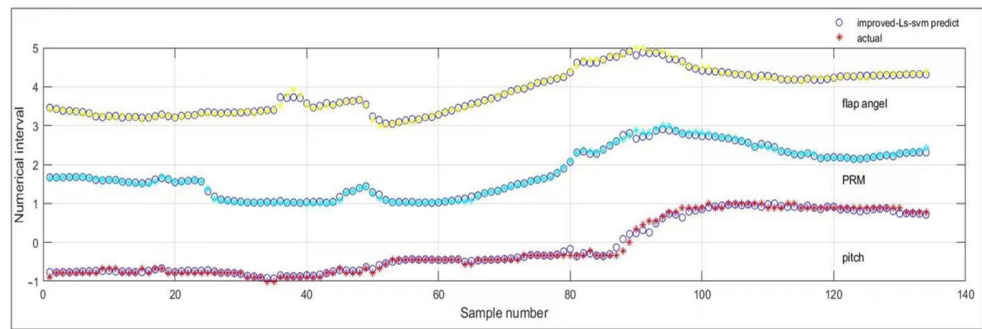


Figure 6. Approach phase simulation diagram.

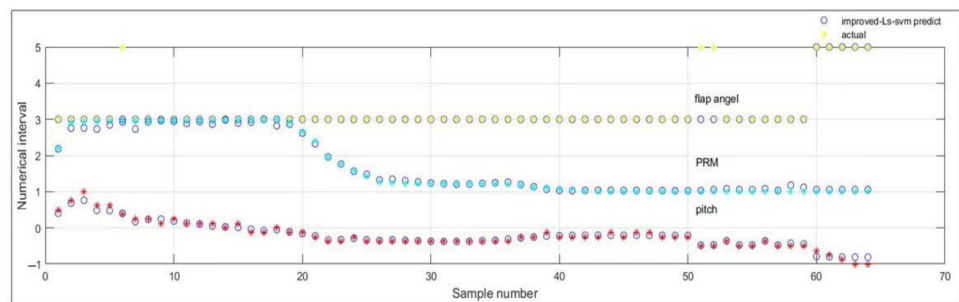


Figure 7. Landing phase simulation diagram.

By observing the image, it is found that the data fitting degree of each factor and each stage is relatively good, so further simulation result analysis can be carried out.

4. Simulation and Discussion

The experimental results are analyzed through simulation methods, and the error indicators of the complement results are shown in Table 3.

Table 3. Error index of missing data completion.

Pitch	MSE	MAE(%)	RMSE	EC
climb	-4.81×10^{-17}	3.59%	5.56×10^{-16}	0.99
approach	-5.15×10^{-16}	7.70%	5.95×10^{-15}	0.99
landing	-1.78×10^{-17}	2.64%	2.06×10^{-16}	0.99
N1	MSE	MAE(%)	RMSE	EC
climb	-5.20×10^{-17}	2.93%	6.02×10^{-16}	0.99
approach	3.89×10^{-17}	7.43%	4.51×10^{-16}	0.99
landing	2.40×10^{-17}	2.61%	2.78×10^{-16}	0.99
Flap angle	MSE	MAE(%)	RMSE	EC
climb	-9.53×10^{-17}	4.07%	1.10×10^{-15}	0.99
approach	-5.15×10^{-16}	9.00%	0.99	0.99
landing	7.66×10^{-17}	2.41%	8.87×10^{-16}	0.99

The error measurement index MAE in the table shows that the lower average error accuracy is more than 90% and the error index value EC in the table has reached a high degree of fit of 0.99. It can be seen that the QAR data item is used as the feature value to assign weights through EWM, and the PCA dimensionality reduction method finally uses the LS-SVM algorithm to fill in the missing data of the QAR to great effect. However, since most of the routes sailed by the aircraft are repeated flights of the same route, when faced with multiple losses or overall losses, the same method can be used to simulate the historical data to restore the lost flight data.

5. Conclusions

The previous data processing experience is based on the QAR itself to detect changes in the body or environment and other actual conditions. Few studies have been conducted on the preservation and restoration of the QAR data itself. This work provides some ideas in this regard. In this paper, the improved LS-SVM method based on the entropy weight method (EWM) and principal component analysis (PCA) is shown to effectively fit the missing QAR data. The parameters are gradually stable during the training process, which ensures that the model can be directly applied for data fitting without retraining, achieving the purpose of fast and simple applicability. This article only considers the case of single item loss, since most of the aircraft sailing on the same route repeats the flight; when faced with multiple losses or overall loss, the same method can be used to simulate historical data to restore this loss of flight data.

Due to the uniqueness of flying at high plateaus, there may be differences when flying on normal routes and the same conclusion may not be applicable for the normal flight. Its practical applicability remains to be further studied.

Author Contributions: Conceptualization, N.C. and Y.S.; data curation, N.C. and Z.W.; methodology, N.C. and C.P.; formal analysis, N.C. and Z.W.; writing—original draft preparation, N.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number: U2033202); the Key R&D Program of the Sichuan Provincial Department of Science and Technology (2022YFG0213); and the Safety Capability Fund Project of the Civil Aviation Administration of China (2022J026).

Data Availability Statement: The data used to support the findings of this study are included within the article.

Acknowledgments: The authors would like to thank the National Natural Science Foundation of China (U2033202), the Key R&D Program of the Sichuan Science and Technology Department (2022YFG0213), the Safety Capability Fund Project of the Civil Aviation Administration of China (2022J026), and the Flight Technology and Flight Safety Research Base Open Fund Project (F2019KF08).

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

References

1. Xu, J.C.; Sun, Y.C. Airworthiness requirement of transportation category aircraft operation on high plateau airports. *Aeronaut. Comput. Tech.* **2018**, *48*, 133–138.
2. Feng, Y.W.; Pan, W.H.; Lu, C. Research on Operation Reliability of Aircraft Power Plant Based on Machine Learning. *Acta Aeronaut. Astronaut. Sin.* **2021**, *42*, 524732. [[CrossRef](#)]
3. Ye, B.J.; Bao, X.; Liu, B. Machine learning for aircraft approach time prediction. *Acta Aeronaut. Astronaut. Sin.* **2020**, *41*, 359–370.
4. Fang, G.C.; Jia, D.P.; Liu, Y.F. Military airplane health assessment technique based on data mining of flight parameters. *Acta Aeronaut. Astronaut. Sin.* **2020**, *41*, 296–306.
5. Liu, J.Y.; Wang, D.Q.; Cui, J.W. Research on classification of screw locking results based on improved kernel LS-SVM algorithm. *Ind. Instrum. Autom.* **2020**, *4*, 12–15.
6. Li, S.; Wang, Y.; Xue, Z.L. Grounding resistance monitoring data regression prediction method based on LS-SVM. *Foreign Electron. Meas. Technol.* **2019**, *8*, 19–22.
7. Wu, H.; Li, B.W.; Zhao, S.F.; Yang, X.; Song, H. Research on initial installed power loss of a certain type of turbo-shaft engine using data mining and statistical approach. *Math. Probl. Eng.* **2018**, *2018*, 9412350. [[CrossRef](#)]
8. Puranik, T.G.; Mavris, D.N. Anomaly detection in general-aviation operations using energy metrics and flight-data records. *J. Aeros. Comp. Inf. Com.* **2018**, *15*, 22–253. [[CrossRef](#)]
9. Puranik, T.G.; Rodriguez, N.; Mavris, D.N. Towards online prediction of safety-critical landing metrics in aviation using supervised machine learning. *Transp. Res. Part C Emerg. Technol.* **2020**, *120*, 102819. [[CrossRef](#)]
10. Yildirim, M.T.; Kurt, B. Aircraft gas turbine engine health monitoring system by real flight data. *Int. J. Aerospace Eng.* **2018**, *2018*, 9570873. [[CrossRef](#)]
11. Yildirim, M.T.; Kurt, B. Confidence interval prediction of ANN estimated LPT parameters. *Aircr. Eng. Aerosp. Technol.* **2019**, *9*, 101–106. [[CrossRef](#)]
12. Martín, F.J.V.; Sequera, J.L.C.; Huerga, M.A.N. Using data mining techniques to discover patterns in an airline's flight hours assignments. *Int. J. Data. Warehous.* **2017**, *13*, 45–62. [[CrossRef](#)]

13. Davison Reynolds, H.J.; Lokhande, K.; Kuffner, M.; Yenson, S. Human–Systems integration design process of the air traffic control tower flight data manager. *J. Cogn. Eng. Decis. Mak.* **2013**, *7*, 273–292. [[CrossRef](#)]
14. Kumar, A.; Ghosh, K. GPR-based novel approach for non-linear aerodynamic modeling from flight data. *Aeronaut. J.* **2019**, *123*, 79–92. [[CrossRef](#)]
15. Lan, C.E.; Wu, K.Y.; Yu, J. Flight characteristics analysis based on QAR data of a jet transport during landing at a high-altitude airport. *Chin. J. Aeronaut.* **2012**, *25*, 13–24. [[CrossRef](#)]
16. Oehling, J.; Barry, D.J. Using machine learning methods in airline flight data monitoring to generate new operational safety knowledge from existing data. *Saf. Sci.* **2019**, *114*, 89–104. [[CrossRef](#)]
17. Walker, G. Redefining the incidents to learn from: Safety science insights acquired on the journey from black boxes to flight data monitoring. *Saf. Sci.* **2017**, *99*, 14–22. [[CrossRef](#)]
18. Deng, W.; Li, Z.; Li, X.; Chen, H.; Zhao, H. Compound fault diagnosis using optimized MCKD and sparse representation for rolling bearings. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–9. [[CrossRef](#)]
19. Cui, H.; Guan, Y.; Chen, H. Rolling element fault diagnosis based on VMD and sensitivity MCKD. *IEEE Access* **2021**, *9*, 120297–120308. [[CrossRef](#)]
20. Wang, L.; Ren, Y.; Wu, C.X. Effects of flare operation on landing safety: A study based on ANOVA of real flight data. *Saf. Sci.* **2018**, *102*, 14–25. [[CrossRef](#)]
21. Zhou, D.; Zhuang, X.; Zuo, H.; Wang, H.; Yan, H. Deep learning-based approach for civil aircraft for civil aircraft hazard identification and prediction. *IEEE Access* **2020**, *8*, 103665–103683. [[CrossRef](#)]
22. Cheng, S.L.; Gao, Z.H.; Zhu, X.Q. Unsteady aerodynamic modelling of unstable dynamic process. *Acta Aeronaut. Astronaut. Sin.* **2020**, *41*, 238–249.
23. Li, M.; Wu, C. A distance model of intuitionistic fuzzy cross entropy to solve preference problem on alternatives. *Math. Probl. Eng.* **2016**, *2016*, 8324124. [[CrossRef](#)]
24. Zhang, X.; Wang, H.; Du, C.; Fan, X.; Cui, L.; Chen, H.; Deng, F.; Tong, Q.; He, M.; Yang, M.; et al. Custom-molded offloading footwear effectively prevents recurrence and amputation, and lowers mortality rates in high-risk diabetic foot patients: A multicenter, prospective observational study. *Diabetes Metab. Syndr. Obes. Targets Ther.* **2022**, *15*, 103–109. [[CrossRef](#)] [[PubMed](#)]
25. Zhu, Y.; Deng, B.; Huo, Z. Key deviation source diagnosis for aircraft structural component assembly driven by small sample inspection data. *China Mech. Eng.* **2019**, *30*, 2725–2733.
26. Gao, X.; Hou, J. An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process. *Neurocomputing* **2016**, *174*, 906–911. [[CrossRef](#)]
27. Safaldin, M.; Otair, M.; Abualigah, L. Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks. *J. Amb. Intel. Hum. Comp.* **2021**, *12*, 1559–1576. [[CrossRef](#)]
28. Abualigah, L.; Diabat, A.; Mirjalili, S.; Elaziz, M.A.; Gandomi, A.H. The arithmetic optimization algorithm. *Comput. Methods Appl. Mech. Eng.* **2021**, *376*, 113609. [[CrossRef](#)]
29. Cai, J.; Bao, H.; Huang, Y.; Zhou, D. Risk identification of civil aviation engine control system based on particle swarm optimization-mean impact value-support vector machine. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* **2022**, *in press*. [[CrossRef](#)]
30. Smart, E.; Brown, D.; Denman, J. Combining multiple classifiers to quantitatively rank the impact of abnormalities in flight data. *Appl. Soft Comput.* **2012**, *12*, 2583–2592. [[CrossRef](#)]
31. Li, G.; Li, Y.; Chen, H.; Deng, W. Fractional-Order Controller for Course-Keeping of Underactuated Surface Vessels Based on Frequency Domain Specification and Improved Particle Swarm Optimization Algorithm. *Appl. Sci.* **2022**, *12*, 3139. [[CrossRef](#)]
32. Deng, W.; Zhang, X.X.; Zhou, Y.Q.; Liu, Y.; Zhou, X.B.; Chen, H.L.; Zhao, H.M. An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. *Inf. Sci.* **2022**, *585*, 441–453. [[CrossRef](#)]
33. Elisa, Q.M.; Lu, S.; Blazquez, C. Use of data imputation tools to reconstruct incomplete air quality datasets: A case-study in Temuco, Chile. *Atmos. Environ.* **2019**, *200*, 40–49.
34. Hadeed, S.J.; O'Rourke, M.K.; Burgess, J.L.; Harris, R.B.; Canales, R.A. Imputation methods for addressing missing data in short-term monitoring of air pollutants. *Sci. Total Environ.* **2020**, *730*, 139140. [[CrossRef](#)]
35. Liu, Z.J.; Wan, J.Q.; Ma, Y.W. Online prediction of effluent COD in the anaerobic wastewater treatment system based on PCA-LS-SVM algorithm. *Environ. Sci. Pollut. Res.* **2019**, *26*, 12828–12841. [[CrossRef](#)]
36. Cheolmin, K.; Klabjan, D. A simple and fast algorithm for L1-norm Kernel PCA. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 1842–1855.