

Research on Flight Safety Indicators Based on Confidence and Principal Component Analysis

Shanshan Yu^{1, #, *}, Yizheng Liu^{2, #}, Chuyang Kang^{2, #}, Zhutao Hu^{2, #}

¹ Department of Basic Disciplines, Space Engineering University, Beijing, China, 101416

² School of Aerospace Information, Space Engineering University, Beijing, China, 101416

* Corresponding Author Email: yuss_seu@163.com

Abstract. Modern technology protects flight safety, especially in the specific research and application of flight quality monitoring, and similar studies are currently focusing on large-scale reading of flight data, and storing and analysing it. In this paper, the data from switching on the autopilot is extracted and the reliability is firstly investigated using the Cronbach's alpha reliability coefficient method. Considering the various types of data as the questions of the questionnaire, and the result of Min-Max normalisation of the data as the score of the questionnaire, the size of the reliability coefficients of the data for the eight flight segments was finally derived, and the reliability of the data was quantitatively described based on this. In order to filter the importance of factors reflecting flight safety, seven indicators related to flight safety were then extracted. Then, through principal component analysis, the importance of the indicators was analysed by the size of the coefficients, and the degrees of importance, from the largest to the smallest, were as follows: airspeed, ground speed, descent rate, altitude, and attitude.

Keywords: Autopilot; Reliability study; Principal component analysis.

1. Introduction

For the civil aviation industry, flight safety is the basis for survival and development. Research on flight safety is becoming more and more important with the rapid development of China's civil aviation industry. If a serious flight data occurs, it will bring huge economic loss to the airline, and more often, it will cause serious life threat to the passengers. Therefore, we need to focus on flight safety issues, deepen the research on aviation safety, use the existing data to strengthen scientific management, improve the quality of relevant personnel through a variety of targeted, systematic means, monitoring and early warning of risks, so as to reduce the probability of flight accidents[1-2].

At present, the main research in China's civil aviation industry is divided into two aspects: one is the research, analysis and application of overrun events; the other is the statistical analysis and application of non-overrun data. Since the overrun is not all caused by human factors, so just overrun analysis want to identify difficult, QAR overrun is used for aviation safety management and flight training data support. However, these data are mixed, this paper for some possible errors in the QAR data, data preprocessing, so that the impact of erroneous data on the research and analysis to reduce. Then, some key data items related to flight safety are extracted and analysed for their importance[3-4].

2. Modelling and solving

2.1. Data reliability study

2.1.1 Data pre-processing

The data is from <http://mathorcup.org/>. The data when the autopilot was turned on was selected because we assumed that there was no more human manipulation after the autopilot was turned on, so the error data was only caused by equipment or instrument factors, and human factor interference was excluded. Because we want to analyse the reliability of the data, so we only select this part of the data can facilitate us to analyse the reliability of the data [5].

2.1.2 Modelling

In order to describe the reliability of the data, the idea of confidence quantification is utilised. Reliability, also known as reliability, refers to the degree of consistency of the results obtained from repeated measurements of the same object using the same method, i.e., the degree of reliability of the measurement data[6]. The reliability index is mostly expressed by the correlation coefficient, that is, the correlation coefficient of the two sets of data obtained from the same test sample is used as an index of measurement consistency, as the reliability coefficient. Using the Cronbach alpha reliability coefficient method, the principle is:

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K S_i^2}{S_x^2} \right) \quad (1)$$

K is the number of questions in a particular questionnaire, S_x^2 is the variance of the total sample, That is, the variance of the total score of each respondent's rating on each question of a particular questionnaire, S_i^2 is the variance of the current observation sample, i.e., the variance of each respondent's rating on a particular topic.

The method was applied with the study of the reliability of the questionnaire, which we applied after processing the data.

2.1.3 Solution of the model

The method was applied to analyse the data for reliability by taking data that were relatively stable during the mid-flight period[6-7], considering the various types of data as questionnaire questions, normalising the various types of data, taking the recorded values of each parameter in each second as a questionnaire score, and solving for the variance to obtain S_i^2 ; The values recorded for each type of data in each second were added together as the total score for the questionnaire and the variance was found to yield the S_x^2 . And the results are obtained by solving using MATLAB according to the formulae (table 1).

Table. 1 The Magnitude of Cronbach α Reliability Coefficient Corresponding to Data from Different Flight Segments

Segment Data Number	Reliability Coefficient α
201404070532	0.8570
201404071917	0.7782
201404080617	0.8253
201404081034	0.8371
201404090110	0.7650
201404091701	0.8333
201404100843	0.7615
201404101159	0.8367

Five of the flights had reliability coefficients greater than 0.8, which is very good; the remaining three flights had reliability coefficients between 0.7 and 0.8, which is good. Overall, the reliability of the data is high.

2.2. Analysis of key data importance

2.2.1 Data pre-processing

Before building the model, the data were screened for the data that did not have the autopilot switched on in the front and back segments. For the 0 data in the slope data and the data with a great span in the elevation data and other obvious errors, we thought that they were caused by the equipment

failure during data collection. As a result of our analyses, we identified eight data items within each data item that are relevant to flight safety[8-9]:

Altitude: the flight altitude should be a certain height to ensure safety.

Descent rate: Too fast a descent rate will cause big impact and damage to the aircraft, affecting the performance of the aircraft.

Radio altitude: related to the take-off and landing of the aircraft, it has a certain impact on flight safety.

Landing G-value: affects the falling speed of the aircraft, which has a certain impact on flight safety.

Attitude: affects the movement of the aircraft, i.e. affects the flight height and direction.

Slope: ensures that the aircraft will not rub the wingtips during take-off and landing.

Airspeed: excessive airspeed during take-off and descent will affect flight safety.

Ground speed: Fluctuations in weather conditions can be a threat to flight safety.

We analyse the above eight data items and find that the correlation between altitude and radio altitude is very strong, and the radio altitude will have a large sudden change when the airspace standard of different airports is used after consulting the information, so we do not analyse the data item of radio altitude and choose seven characteristic values related to flight safety such as flight altitude, descent rate, and so on, and we choose the data which has a smaller error, i.e., the data which has less than three standard deviations from the expectation is chosen. data with a difference of less than three standard deviations for normalisation.

2.2.2 Model building and solution

For the key data selection, we downsize all kinds of data by principal component analysis. By analysing and calculating the internal structural relationships of the correlation matrix of the original data, some new variables that are not related to each other are generated. The n-dimensional original features are mapped onto ($k < n$) dimensions using principal component analysis, then these k-dimensional features are principal components. Principal component analysis is the reconstruction of k-dimensional orthogonal features, and the newly generated data contains as much information as possible from the original n-dimensional data[10].

We preprocess the data and list the preprocessed data as a dataset of m samples, n eigenvalues, downscaled to dimension k. Denote the sample dataset as a matrix X .

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} \quad (2)$$

In which each row represents a sample, each column represents a feature, and the column number indicates that it is an n-dimensional feature.

It is first normalised to compute the covariance matrix of the normalised sample cha, with the equation:

$$R = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (3)$$

Afterwards the eigenvalues and eigenvalue vectors of R are computed: (R is a semi-positive definite matrix and $tr(R) = \sum_{k=1}^p \lambda_k = p$)

Eigenvector:

$$\alpha_1 = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \dots \\ \alpha_{p1} \end{pmatrix}, \alpha_2 = \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \\ \dots \\ \alpha_{p2} \end{pmatrix}, \dots, \alpha_p = \begin{pmatrix} \alpha_{1p} \\ \alpha_{2p} \\ \dots \\ \alpha_{pp} \end{pmatrix} \quad (4)$$

The contribution rate and cumulative contribution rate of principal components are then derived:
Contribution rate:

$$\rho_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k}, (i=1, 2, \dots, p) \quad (5)$$

Cumulative contribution rate:

$$\rho = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k}, (i=1, 2, \dots, p) \quad (6)$$

Finally, we write the principal components: usually we take the first, second, ..., mth ($m \leq p$) principal components corresponding to the eigenvalues with a cumulative contribution of more than eighty per cent. The ith principal component:

$$F_i = \alpha_{1i}X_1 + \alpha_{2i}X_2 + \dots + \alpha_{pi}X_p \quad (7)$$

Principal component analysis was applied, and the results were obtained through MATLAB software. Cumulative % is used to illustrate the contribution of the factors, the higher cumulative % indicates that these factors explain more to the overall. Generally cumulative % above 80% indicates more satisfactory. The absolute value of the principal component coefficients indicates the influence of different data items on the principal components, i.e. the degree of importance to flight safety, the higher the absolute value of the principal component coefficients indicates the greater the degree of importance to flight safety.

The ascending stage is:

Table. 2 Roots of Characteristics and Explanation of Variance for Ascending Stage

Component	Initial Eigenvalues			Extract Eigenvalue		
	Total	% of Varianc	Cumulative %	Total	% of Varianc	Cumulative %
1	0.44527	89.34	89.34	0.44527	89.34	89.34
2	0.02082	4.18	93.52	-	-	-
3	0.01975	3.96	97.48	-	-	-
4	0.00767	1.54	99.02	-	-	-
5	0.00322	0.65	99.67	-	-	-
6	0.00149	0.30	99.97	-	-	-
7	0.00017	0.03	100.0	-	-	-

As can be seen from Table 2 above: since the variance explained by the first principal component reaches more than 89%, we believe that the construction of a single principal component can be a good description of the flight safety indicator.

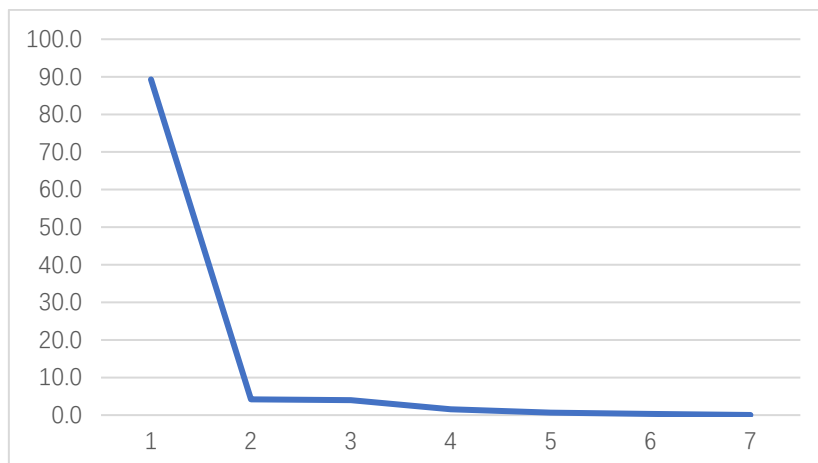


Figure. 1 Scree Plot of The Rising Phase

From the above Figure 1, it can be seen that: the absolute values of the principal component coefficients of airspeed and ground speed are larger, both exceeding 0.5, and thus have a greater influence on the principal components, i.e., a greater degree of importance to flight safety; the absolute values of the landing G-value and gradient (left-negative and right-positive) are smaller, both less than 0.1, and thus have a smaller influence on the principal components, i.e., a lesser degree of importance to flight safety.

The descent phase was shown in table 3:

Table. 3 Table of Eigenroots and Variance Explanations in The Decline Phase

Component	Initial Eigenvalues			Extract Eigenvalue		
	Total	% of Varianc	Cumulative %	Total	% of Varianc	Cumulative %
1.000	0.31620	86.96	86.96	0.31620	86.96	86.96
2.000	0.01705	4.69	91.65	-	-	-
3.000	0.01553	4.27	95.92	-	-	-
4.000	0.00655	1.80	97.72	-	-	-
5.000	0.00446	1.23	98.95	-	-	-
6.000	0.00337	0.93	99.88	-	-	-
7.000	0.00046	0.13	100.0	-	-	-

As can be seen from Table 3 above: since the variance explained by the first principal component reaches more than 86%, we believe that the construction of a single principal component can be a good description of the flight safety indicator.

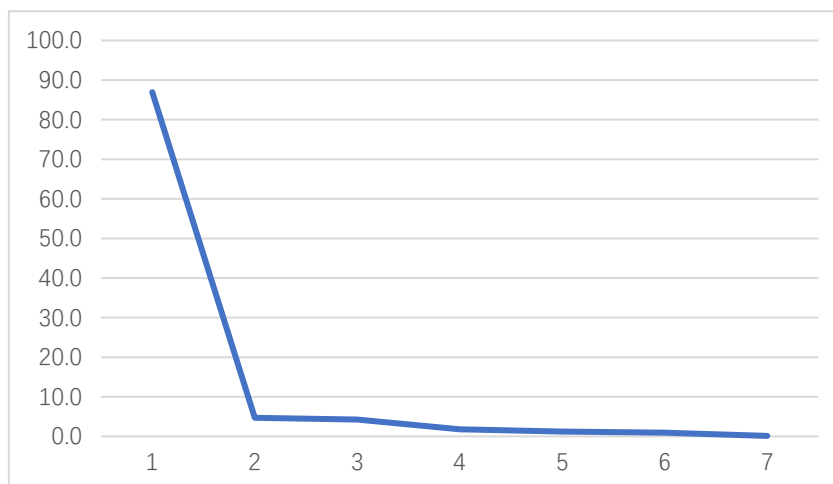


Figure. 2 Scree Plot of The Descent Phase

Table. 4 Table of Principal Component Coefficients for Different Data Items

Designation	Principal Component Coefficient μ
ALTITUDE (1013)	0.291
Inertial Vertical Speed	0.338
COMPUTED AIR SPD	0.644
GROUNDSPEED	0.594
COG NORM ACCEL	0.007
PITCH ATT	0.186
ROLL ATT	-0.002

From the above Figure. 2 and Table 4, we can see that: the absolute values of the principal component coefficients of airspeed and ground speed are larger, both reaching about 0.6, so they have a greater influence on the principal components, i.e., a greater degree of importance to flight safety; the absolute values of the landing G-value and gradient (left-negative and right-positive) are smaller, both less than 0.01, so they have a smaller influence on the principal components, i.e., a lesser degree of importance to flight safety.

After principal component analysis of each data item in the ascent and descent phases, we concluded that airspeed and ground speed are relatively more important to aircraft safety, and landing G-value and slope (left-negative and right-positive) are less important to aircraft safety. For the combined analysis of the data, out of the 7 data items, we concluded that landing G-value and slope (left-negative and right-positive) were not flight-related critical data items. For the 5 critical data items we selected, we ranked them in descending order of importance to flight safety: Airspeed > Ground Speed > Descent Rate > Altitude > Attitude

3. Conclusions

In this study, reliability and principal component analysis are used as the methods to carry out an in-depth study on flight safety indicators. Firstly, the reliability of the flight data with the autopilot switched on is studied by the Cronbach's alpha reliability coefficient method. Through the Min-Max normalisation of the data, the reliability coefficients of the data for the eight flight segments were derived, and the reliability of the data was verified. Next, the researcher extracted seven indicators related to flight safety and applied principal component analysis to successfully downscale these seven indicators to one principal component, explaining 89.3% of the information. By analysing the principal component coefficients, five key data items such as airspeed, ground speed, descent rate, altitude and attitude were identified, reflecting their importance in flight safety. Ultimately, the study provides a reliable methodology and important data references for the assessment of flight safety indicators. The methodology and results of this study are of great significance to flight safety management and flight training, which can help reduce the probability of flight accidents and improve flight safety.

References

- [1] Yuan Ding, Luo Xufeng. Study on flight path reproduction of QAR data based on digital earth [J]. Comprehensive Transport, 2023, 45 (10): 133-138.
- [2] Cui Hao, Zhang Shenli, Ren Haijun et al. Research and application of unstable approach based on convolutional neural network [J]. Aviation Computing Technology, 2023, 53 (05): 20-23+28.
- [3] Wang Xiangzhang, Wang He, Xu Bohao. Risk prediction of civil aircraft heavy landing based on GBDT-GS method [J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 1-12[2023-11-30].
- [4] ZHANG Zhaoyang, SUN Hong, WANG Qi et al. Wind shear risk control based on AHP and QAR data[J]. Project Management Technology, 2023, 21 (09): 115-120.
- [5] PENG Hongbo, MA Pengyu, SHI Lei et al. Research on engine VBV regulation law based on BP neural network [J]. Computer Simulation, 2023, 40 (08): 66-71.

- [6] Feng Yunwen, Liu Jiaqi, Pan Weihuang et al. Progress of civil aircraft operational reliability research [J]. Journal of National University of Defence Technology, 2023, 45 (04): 66-93.
- [7] Niu Weilong, Zhang Baosheng. Research on RB211 engine speed monitoring based on QAR data [J]. Aviation Maintenance and Engineering, 2023, (07): 70-72.
- [8] F.E. Luo, Q. Zhao, X. Zhang et al. Dynamic assessment of flight operation risk based on overrun events [J]. Modern Computer, 2023, 29 (12): 22-26.
- [9] QIAO Kui-Yu, YU Jiang. Influence factors of aircraft levelling down segment distance based on QAR data [J]. Automation Application, 2023, 64 (12): 201-204.
- [10] LIU Jiaxue, CHEN Tianjunhong, GUO Runxia et al. Flight environment wind modelling method based on QAR data [J]. Journal of Weapons and Equipment Engineering, 2023, 44 (05): 200-206.