

A Novel Approach for Comprehensive Evaluation of Flight Deck Ergonomic Design: Delphi-Order Relation Analysis (ORA) Method and Improved Radar Chart

Lijing Wang¹(✉), Yanlong Wang¹, Wenjun Dong²,
Dayong Dong², and Xiuli Shu²

¹ Fundamental Science on Ergonomics and
Environment Control Laboratory,

School of Aeronautic Science and Engineering,
Beihang University, Beijing, China

wanglijing@buaa.edu.cn, wyllaf@sina.com

² Commercial Aircraft Corporation of China Ltd., Shanghai, China
{dongwenjun, dongdayong, shuxiuli}@comac.cc

Abstract. A well designed flight deck with full consideration of ergonomic aspect has a significant effect on aircraft safety. Since cockpit is a complicated system, it is necessary to have a comprehensive evaluation of flight deck ergonomic design during the design and certification process in order to grasp the overall ergonomic design quality. The determination of indicator weights and aggregation of indicator evaluation values are key steps of comprehensive evaluation. However, most of existing methods lack a sufficient consideration of uncertainty of subjective judgment and interdependence between indicators. Therefore, Delphi-order relation analysis (ORA) method and improved radar chart were proposed in this paper to address these two problems respectively. A feedback mechanism is introduced in Delphi-ORA method to control the limitation of expert's knowledge structure and experience. The correlation coefficient is incorporated in the improved radar chart to reflect the interdependence between indicators.

Keywords: Flight deck ergonomic design · Comprehensive evaluation · Delphi-ORA · Improved radar chart · Uncertainty · Interdependence

1 Introduction

1.1 The Status and the Meaning of Research About Comprehensive Evaluation on Cockpit Ergonomics Design

According to literature and accident report, about 70 % of all aviation accidents were caused by human factors [1–4]. In order to reduce human error and pilot workload, the flight deck need to be designed carefully with full consideration of ergonomics. Then the aviation safety can be enhanced. In order to ensure the design quality of man-machine interface on flight deck, the ergonomic requirements need to be integrated in design and

the ergonomic evaluation efforts are essential. However, it is difficult to grasp the overall ergonomic design level of cockpit. There is a need for comprehensive evaluation. In the evaluation and comprehensive evaluation aspect of cockpit ergonomic design, plenty of research and application efforts have been done [5–8].

1.2 Limitations of Existing CE Methods

The process consisted of 5 steps, i.e. specify the evaluation purpose, identify the cockpit ergonomic metrics (evaluation indicator system), determine indicators' weights; rate each evaluation indicator, and select/construct comprehensive evaluation model to conduct comprehensive evaluation [9, 10].

Weight determination is a key step in the process of comprehensive evaluation. The rationality of the weight represents a correct description of the relationship of evaluation metrics and object to be evaluated, which determines the validity of the result of CE, apparently. Several methods have been proposed to determine weights. Majority of them can be classified into subjective methods, objective methods and combinations of them depending on the information provided [11]. In general, objective methods assign indicator weights according to the structure or the internal mechanism of evaluation object. However, for the comprehensive evaluation problem of cockpit ergonomic design, the relationship between various indicators is complex, the effect of each individual indicator on the overall ergonomic level is not very clear. Moreover, ergonomic evaluation is greatly influenced by evaluator's subjective factors. These reasons make it difficult to determine indicator weights using objective methods. Thus, in general case, the subjective method, which is based on expert judgment, is commonly used to acquire indicator weights. However, the subjective weighting method has the following limitations:

- Strong subjective, the weighting results depend on experts' work experience, knowledge structure and their preference,
- Expert judgment is equivalent to the black box operation and the transparency of the evaluation process is poor,
- And, since the uncertainty of subjective judgment, the repeatability of results is poor.

Therefore, the same decision maker can give different weight to the same indicators under different situations. In order to achieve group consensus, Mukherjee (2014) advised the Brian Storming method [12], and Dempster–Shafer theory was used in Ju and Wang's (2012) research [13].

Comprehensive evaluation is a method using a mathematical model ($y = f(\mathbf{w}, \mathbf{x})$) to aggregate individual indicator evaluation result into a composite indicator with consideration of the weight information. The comprehensive evaluation value represents the overall level of the evaluated object. Here, the mathematical model used for "aggregation" is the comprehensive evaluation model. Due to the entire cockpit is a complex system, including many subsystems, and there exist closely correlation and strong coupling between the subsystems. This makes the evaluation indicators for the comprehensive evaluation of cockpit ergonomics have significant interdependence.

Whereas the existing comprehensive evaluation models, such as fuzzy comprehensive evaluation method, principal component analysis method, and utility theory model, etc., are essentially a linear weighted method. Their results contain serious duplicated information and cannot reflect the true level of the object evaluated [14]. Although the neural network is a non-linear model, it is difficult to obtain enough high quality training samples, which greatly limits its application.

1.3 Innovation Work

The purpose of this paper is to address the limitations described above. With the feedback of the majority's opinion of experts, Delphi method can achieve group consensus through multiple rounds of consultation [15]. Inspired by Delphi method, the feedback mechanism was introduced in Order Relation Analysis (ORA) method to address the bias and uncertainty of expert subjective judgment.

The existing practice to handle the interdependence between evaluation indicators is to eliminate the overlapping information during the construction stage of the evaluation metrics [16], or to adjust the indicator weight according to the correlation between them [17], when ANP is advisable [18]. Considering the general process of comprehensive evaluation, it is apparent that another potential point to handle the redundant information between indicators lies in comprehensive evaluation model. However the related literatures are few. The improved radar chart proposed in this paper has integrated the correlation between indicators, which is a means to address the overlapping information through the development of an appropriate comprehensive evaluation model.

1.4 The Structure of the Thesis

The remainder of this paper is structured as follows. The second section introduces the process of Delphi-ORA method in detail, including the design of Loop control variable and feedback variable. The third section describes the improved Radar Chart. Finally, a conclusion and discussion is given.

2 Delphi-Order Relation Analysis Method

The order relation analysis (ORA) method proposed by Guo [19], also called G1 method, is a subjective weighting approach based on expert judgment. Compared with analytic hierarchy process (AHP), ORA method does not need consistency testing. Besides, the amount of Comparative judgment work has been reduced greatly. For these reasons, ORA method was chosen in this paper for improvement. A feedback mechanism was introduced in ORA method, which was inspired by Delphi method. The Delphi-ORA method requires organizers to do statistical processing on the data collected in previous advisory round and feedback the result to experts. Then experts are required to give their opinion again based on the feedback information. After several rounds of consultation, experts' opinion will eventually tend to be consistent.

2.1 ORA Method Process

ORA method firstly asks experts to sort the evaluation indicators according to their importance on evaluation object in a descending order. And then estimated values for the importance ratios between adjacent indicators should be given. The estimated ratios are always presented as qualitative linguistic values. During the subsequent data processing, a numerical value will be assigned to each linguistic value. Then through a simple mathematical treatment, the weight coefficient of each indicator can be calculated. The weighting process of ORA method consists of the following three steps.

Indicator Importance Sorting. For given evaluation object, if indicator u_i is more (not less) important than indicator u_j , then it can be documented as $u_i > u_j$. First, the expert need to choose the most important indicator he think among then indicators $\{u_1, u_2, \dots, u_n\}$, and record it as u_1^* . Then select the most important indicator from the remaining $n-1$ indicators to be as u_2^* . Repeat the steps above, after $n-1$ rounds, a descending sorting according to their importance of the n indicators can be derived:

$$u_1^* > u_{n2}^* > \dots > u_n^* \tag{1}$$

Where u_i^* is the i -th indicator after sorting ($i = 1, 2, \dots, n$).

Estimate the Importance Ratio. Suppose w_i^* is the weight coefficient of indicator u_i^* . Then the importance ratio (r_i) between indicator u_{i-1}^* and u_i^* can be represented as w_{i-1}^*/w_i^* (formula 2). Through expert judgment, a qualitative linguistic value of r_i can be obtained. Then a numerical value should be assigned to each linguistic level as appropriately. The assignment can refer to Table 1.

$$w_{i-1}^*/w_i^* = r_i, \quad i = n, n - 1, \dots, 2 \tag{2}$$

Table 1. Suggested linguistic levels and corresponding numerical values for r_i

Linguistic level terms (w_{i-1}/w_i)	Numerical values (r_i)	Explanation
Equally important	1.0	u_{i-1}^* and u_i^* are equally important
Slightly important	1.2	u_{i-1}^* is slightly more important than u_i^*
Moderately important	1.4	u_{i-1}^* is moderately more important than u_i^*
Very important	1.6	u_{i-1}^* is much more important than u_i^*
Extremely important	1.8	u_{i-1}^* is extremely more important than u_i^*

Calculate Weight Coefficient w_i . Formula (3) and (4) can be used to calculate the weight coefficient of each indicator in sorting (1)

$$w_n^* = \left(1 + \sum_{k=2}^n \prod_{i=k}^n r_i\right)^{-1} \tag{3}$$

$$w_{i-1}^* = r_i w_i^*, \quad i = n, n-1, \dots, 2 \tag{4}$$

Therefore, weight coefficients of sorted indicators can be obtained,

$$\mathbf{W}^* = (w_1^*, w_2^*, \dots, w_n^*) \tag{5}$$

Then adjust the order of each element in weight vector \mathbf{W}^* , we can get the weight vector corresponding to the original indicator set $\{u_1, u_2, \dots, u_n\}$,

$$\mathbf{W} = (w_1, w_2, \dots, w_n) \tag{6}$$

2.2 Information Feedback Variable

The consultation result of former round is provided to experts through information feedback variable (IFV). Assuming there are m experts participating in the indicator weight consultation of ORA method, then after the first round consultation, we can get m pairs of different sorting and corresponding estimated importance ratios (see formula 7) for indicator set $\{u_1, u_2, \dots, u_n\}$.

$$\left\{ \begin{array}{l} u_{i1}^* > u_{i2}^* > \dots > u_{in}^* \\ (r_{i2}, r_{i3}, \dots, r_{in}) \end{array} \right., \quad (i = 1, 2, \dots, m) \tag{7}$$

Thus, the weighting result for indicator set $\{u_1, u_2, \dots, u_n\}$ of the i -th expert can be calculated using ORA method.

$$\mathbf{W}_i = (w_{i1}, w_{i2}, \dots, w_{in}), \quad (i = 1, 2, \dots, m) \tag{8}$$

Where, w_{ij} denotes the weight coefficient assigned to the j -th indicator in $\{u_1, u_2, \dots, u_n\}$ by the i -th expert.

The arithmetic mean of \mathbf{W}_i can be calculated as formula (9), which is regarded as the ultimate weight vector of the first round consultation.

$$\mathbf{W} = \frac{1}{m} \sum_{i=1}^m \mathbf{W}_i = \left(\frac{1}{m} \sum_{i=1}^m w_{i1}, \dots, \frac{1}{m} \sum_{i=1}^m w_{in} \right) = (w_1, w_2, \dots, w_n) \tag{9}$$

According to the size of each element (indicator weight) in the weight vector \mathbf{W} , the first round average sorting of the n indicators could be determined. And this average sorting is selected as the information feedback variable.

$$u_1^* > u_2^* > \dots > u_n^* \tag{10}$$

2.3 Loop Control Variable

Loop control variable is used to test the consistency of expert judgment, upon which the weight consultation process is determined whether to be ended. In this paper, an average ordering deviation index (AODI) based on the number of reverse order is proposed to represent the consistency/divergence level of expert group ordering.

An ordered array composed of natural numbers $(1,2,\dots,n)$ is called an-order permutation, denoted as j_1, j_2, \dots, j_n . An n-order permutation can most have $n!$ different permutations. Among them, the ascending permutation, i.e. $1,2,\dots,n$, is defined as standard permutation or natural permutation. In a permutation, if one bigger number is in front of a smaller one, then these two numbers form a reverse order. The total number of reverse orders in a permutation is defined as the reverse order number (RON) of the permutation, denoted as $\tau(j_1, j_2, \dots, j_n)$.

The average sorting $(u_1^* > u_2^* > \dots > u_n^*)$ obtained from a weight consultation round is regarded as the standard permutation $(1,2,\dots,n)$. For the i -th expert's sorting $u_{i1}^* > u_{i2}^* > \dots > u_{in}^*$, it corresponds to a n-order permutation and its RON can be denoted as $\tau(i)$. For a n-order permutation, the maximum RON is $n(n-1)/2$. The ratio of average RON of m permutations provided by expert group and the maximum RON is used to represent the consistency/divergence level of expert group ordering, i.e. the average ordering deviation index (AODI).

$$AODI = \frac{\sum_{i=1}^m \tau(i) / m}{n(n-1) / 2} \tag{11}$$

The smaller the ordering deviation is, the smaller AODI will be, and vice versa. If $AODI < 0.1$, the consistency of expert group ordering is acceptable and the weight consultation can be ended.

3 Improved Radar Chart

3.1 Radar Chart and Comprehensive Evaluation

Radar chart is also known as spider diagram, which consists of several concentric circles (or polygons) and some axes starting from the circle center. Each axis in the radar chart denotes an indicator, while each concentric circle (or polygon) represents a certain indicator level. Radar chart comprehensive evaluation model is a kind of indicator value aggregation method based on the extraction of feature variables of radar chart. Radar chart comprehensive evaluation model is a combination of graphical evaluation method and digital evaluation method, which is greatly suitable for an entire and overall evaluation of complex multi-attribute structure, and much more intuitive, as well (Fig. 1).

Huili Zheng described the general process of radar chart comprehensive evaluation [20], Liu and Chen Yong improved radar chart comprehensive evaluation respectively [21, 22]. The most commonly used feature variables in radar chart comprehensive evaluation are the area A and perimeter C of radar chart, which are calculated as:

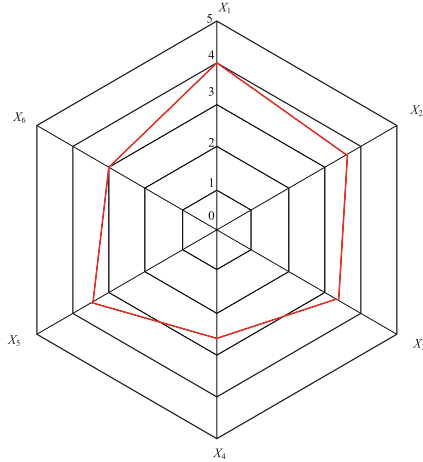


Fig. 1. The general form of traditional radar chart

$$\begin{cases} A = \sum_{i=1}^n \frac{1}{2} y_i y_{i+1} \sin \theta \\ C = \sum_{i=1}^n \sqrt{y_i^2 + y_{i+1}^2 - 2y_i y_{i+1} \cos \theta} \end{cases} \quad (12)$$

Where θ is angle between indicator axes, $\theta = 2\pi/n$. Area A represents the level of object being evaluated, while C reflects the balanced development of each index.

3.2 Improvement of Radar Chart

A fatal deficiency in traditional radar chart comprehensive evaluation method is that there is a lack of consideration of indicator weight and correlation between them. It is apparent in Eq. (12) that there are two factors affecting the results of radar chart comprehensive evaluation, i.e. indicator values and angles between indicator axes. Therefore, two points of improvement were conducted in this paper from the two aspects mentioned above, which will make radar chart comprehensive evaluation model more suitable for situations where strong mutual coupling exists.

Indicator Axis Value Integrated with Weight. For conventional radar chart, the evaluation value (or state value) of each indicator is used as indicator axis value immediately without consideration of indicator weight. In this paper, the indicator weight to the power of indicator evaluation value, i.e. $x_i^{w_i}$, is used as the indicator axis value. Thus the area formula in Eq. (12) will be transformed as follows.

$$A = \sum_{i=1}^n \frac{1}{2} x_i^{w_i} \cdot x_{i+1}^{w_{i+1}} \sin \theta_i \quad (13)$$

Formula (13) is actually a combination of linear weighted model and geometric weighted model, which has certain advantages of both.

Determination of θ_i . The angles between indicator axes in a conventional radar chart are usually the same, i.e. if there are n indicators, the axis angle will be $2\pi/n$. Correlation coefficients r_i between adjacent indicators are utilized as a basis for the determination of corresponding axis angle θ_i . The calculation formula are as follows.

$$k_i = \begin{cases} \sqrt{1 - r_i^2}, & (0 \leq r_i \leq 1) \\ \sqrt{1 - r_i^2}, & (-1 \leq r_i < 0) \end{cases} \quad (14)$$

$$\theta_i = 2\pi \cdot k_i / \sum_{j=1}^n k_j \quad (15)$$

Thus the angle between indicator axes will no longer be the average, but determined by correlation coefficients between adjacent indicators. Therefore, the greater the correlation between two indicators, the smaller the angle will be. As a result, their contribution to the composite indicator will be smaller, accordingly. Hence, the improved radar chart is no longer a regular polygon, but similar to the shape shown in Fig. 2. The evaluation value of indicator needs to be normalized, and the negative indicator (negatively correlated with composite indicator) should be transformed into a positive indicator (positively correlated with composite indicator) previously before determining their correlation.

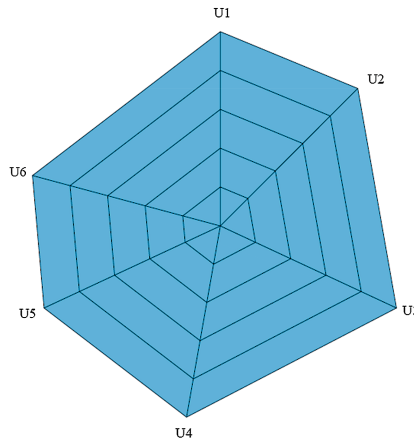


Fig. 2. Example form of improved Radar Chart

There are two methods recommended for determining the correlation coefficient r_i in this paper.

1. For the case that there exist multiple objects, supposing p objects, to be evaluated, employing the indicator set $\{u_1, u_2, \dots, u_n\}$ to evaluate these objects respectively, then a $p \times n$ -order evaluation matrix X can be obtained.

$$X = [X_1, X_2, \dots, X_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix} \tag{16}$$

Where x_{ij} represents the evaluation value of the i -th object, while X_j denotes the evaluation vector of the j -th indicator on the p objects. Thus, the Pearson correlation coefficient can be calculated and used to represent the correlation between indicators.

$$r_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{D(X_i)}\sqrt{D(X_j)}} \tag{17}$$

Where $\text{Cov}(X_i, X_j)$ is the covariance between X_i and X_j , $D(X_i)$ and $D(X_j)$ are the variance of X_i and X_j respectively.

2. For the case of only one evaluation object, such as the comprehensive evaluation of a specific alternative during the system design process, expert judgment is advisable for the acquisition of correlation coefficients. The correlation relationship between evaluation indicators may be divided into the following three conditions:

- Positive correlation: indicator B will be improved with the improvement of indicator A;
- Irrelevant: the improvement of indicator A has no effect on indicator B;
- Negative correlation: the improvement of indicator A will worsen indicator B.

When judging the correlations between indicators, the scale presented in Fig. 3 can be served as a reference, and the judgment results can be recorded in the questionnaire shown in Fig. 4 accordingly. In the subsequent data processing, an appropriate correlation coefficient value can be assigned to each level, and finally the mean of expert judgment results will be employed as the correlation coefficient between indicators. This part of work can be carried out simultaneously with the weight consultation process.

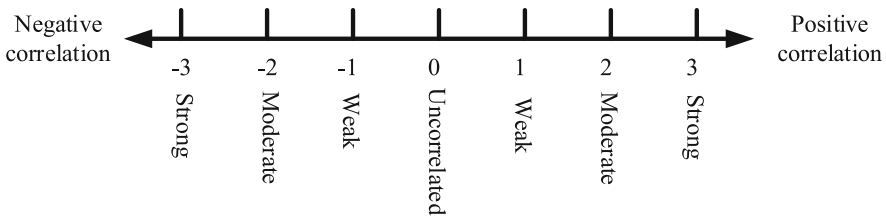


Fig. 3. The qualitative grade division of indicator correlation

X_1								
X_2								
X_3								
X_4								
X_5								
X_6								
X_7								
X_8								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8

Fig. 4. A questionnaire example for correlation coefficient survey

4 Conclusion

In this paper a new approach for comprehensive evaluation of cockpit ergonomic design was proposed. Though there are plenty of researches and applications about comprehensive evaluation/multiple attribute decision making, most of them have a lack of sufficient consideration of the subjectivity of expert judgment and the interdependence between evaluation indicators. The effort of this paper is focused on these two problems.

To mitigate the limitation of individual expert's knowledge structure and experience, and handle the uncertainty and randomness of subjective judgment, a feedback mechanism was introduced into ORA method for acquiring more reasonable weight coefficients. The new proposed method, named Delphi-ORA method, can promote expert opinions to a consensus through a multi-round consultation. In the multi-round consultation, the average sorting of expert group was utilized as information feedback variable, while the average ordering deviation index based on number of reverse order was defined as loop control variable. In terms of handling the interdependence between indicators, the common method is to ensure the independence between indicators during the development stage of indicator system, or to adjust indicator weights according to their correlations. From the perspective of comprehensive evaluation model, this paper proposed the improved radar chart to incorporate the consideration of interdependence between indicators. In the improved radar chart, angles between indicator axes are determined based on the correlation coefficients of adjacent indicators, and the indicator weight to the power of indicator evaluation value, i.e. $x_i^{w_i}$, is

used as the indicator axis value. Thus the improved radar chart comprehensive evaluation model is virtually a conjunction of linear and non-linear comprehensive evaluation model, which has incorporated the correlation between indicators. Therefore, the improved radar chart is very suitable for the comprehensive evaluation of complex and strongly coupled systems such as flight deck.

Therefore, it is easy to find out the weak aspect of system design, and then corresponding improvement can be taken. Although methods in this paper were proposed for the comprehensive evaluation of cockpit ergonomic design, they are also applicable to other areas.

References

1. Authority, C.A.: Global Fatal Accident Review 2002–2011. CAPI036 (2013)
2. Shyur, H.-J.: A quantitative model for aviation safety risk assessment. *Comput. Ind. Eng.* **54**(1), 34–44 (2008)
3. Boeing: Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations 1959–2013 (2014)
4. Stanton, N.A., Harris, D., Salmon, P.M., Demagalski, J.M., Marshall, A., Young, M.S., Dekker, S.W., Waldmann, T.: Predicting design induced pilot error using HET (human error template)-A new formal human error identification method for flight decks (2006)
5. Craig, J., Burrett, G.: The design of a human factors questionnaire for cockpit assessment. In: International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, pp. 16–20 (1999)
6. Shamo, M., Dror, R., Degani, A.: A multi-dimensional evaluation methodology for new cockpit systems. In: Proceedings of the Tenth International Symposium on Aviation Psychology, p. 120 (1999)
7. Li, Y., Yang, F., Wang, L., Yuan, X.: Study on comprehensive evaluation of cockpit ergonomics and its application. *J. Beijing Univ. Aeronaut. Astronaut.* **31**(6), 652–656 (2005)
8. Song, H., Sun, Y., Lu, Z.: Study on comprehensive evaluation of cockpit ergonomics. *Aircraf Design* **30**(4), 36–40 (2010)
9. Xu, J., Wu, W.: Multiple Attribute Decision Making Theory and Methods. Tsinghua University Press, Beijing (2006)
10. Guo, F.: Comprehensive Evaluation of Pilot Operation Procedures for Commercial Airliner. Beihang University, Beijing (2010)
11. Ma, J., Fan, Z.-P., Huang, L.-H.: A subjective and objective integrated approach to determine attribute weights. *Eur. J. Oper. Res.* **112**(2), 397–404 (1999)
12. Mukherjee, K.: Supplier selection criteria and methods: past, present and future (2014)
13. Ju, Y., Wang, A.: Emergency alternative evaluation under group decision makers: a method of incorporating DS/AHP with extended TOPSIS. *Expert Syst. Appl.* **39**(1), 1315–1323 (2012)
14. Pereira, J.M., Duckstein, L.: A multiple criteria decision-making approach to GIS-based land suitability evaluation. *Int. J. Geog. Inf. Sci.* **7**(5), 407–424 (1993)
15. Linstone, H.A., Turoff, M.: The Delphi Method: Techniques and Applications. Addison-Wesley, Reading (1975)
16. Li, Y., Yuan, X., Yang, C., Wang, L., Du, J.: Building of the index system for fighter cockpit ergonomics comprehensive evaluation. *ACTA Aeronautica et Astronautica Sinica* **2**, 148–152 (2005)

17. De Ambroggi, M., Trucco, P.: Modelling and assessment of dependent performance shaping factors through Analytic Network Process. *Reliab. Eng. Syst. Saf.* **96**(7), 849–860 (2011)
18. Wang, H.-J., Zeng, Z.-T.: A multi-objective decision-making process for reuse selection of historic buildings. *Expert Syst. Appl.* **37**(2), 1241–1249 (2010)
19. Guo, Y.: *Comprehensive Evaluation: Theory and Methods*. Science Press, Beijing (2002)
20. Zheng, H., Liu, C., Zhe, D.: Comprehensive evaluating method based on radar-graph. *J. Nanjing Univ. Posts Telecommun. (Nat. Sci.)* **21**(2), 75–79 (2001)
21. Chen, Y., Chen, X., Li, Z., Lin, Y.: Method of radar chart comprehensive evaluation with uniqueness feature. *Trans. Beijing Inst. Technol.* **30**(12), 1409–1412 (2010)
22. Liu, H., Liu, A., Zhang, B., Zhang, T., Zhang, X.: A fuzzy comprehensive evaluation method of maintenance quality based on improved radar chart. In: *ISECS International Colloquium on Computing, Communication, Control, and Management, CCCM 2008*, pp. 638–642 (2008)