Contents lists available at ScienceDirect

Journal of Air Transport Management

journal homepage: www.elsevier.com/locate/jairtraman

Safety performance monitoring and measurement of civil aviation unit

Weishi Chen^{*}, Jing Li

Airport Research Institute, China Academy of Civil Aviation Science and Technology, Beijing, 100028, China

ARTICLE INFO

Article history: Received 23 December 2014 Received in revised form 22 August 2016 Accepted 28 August 2016

Keywords: Civil aviation Safety performance Analytic hierarchy process Delphi method

1. Introduction

Safety has always been the most significant issue for the operation of civil aviation units (CAU). In recent years, with the widely promotion of safety management system (SMS) in civil aviation, the operation of safety performance has brought new task and challenge to the CAUs (International Civil Aviation Organization, 2013). An SMS defines measurable performance outcomes to determine whether the system is truly operating in accordance with design expectations and not simply meeting regulatory requirements. In SMS, the safety performance indicators (SPIs) are used to monitor known safety risks, detect emerging safety risks and to determine any necessary corrective actions. The Federal Aviation Administration (FAA) publishes the performance and accountability report every year (Federal Aviation Administration, 2014). European Organisation for the Safety of Air Navigation (Eurocontrol) annually releases the performance review report on the assessment of air traffic management (ATM) in Europe (Eurocontrol, 2014). The FAA and Eurocontrol also jointly provide the comparison of ATMrelated operational performance of U.S. and Europe yearly (FAA & EUROCONTROL, 2014). In the academic field, Luo developed the risk assessment model and procedures for ATM (Luo et al., 2009a, 2009b). Shyur proposed a quantitative model for aviation safety risk assessment, where the model used data on both accident and

E-mail addresses: wishchen@buaa.edu.cn (W. Chen), chenwsh@mail.castc.org.cn (J. Li).

ABSTRACT

A method of safety performance measurement is proposed to monitor the safety management process of civil aviation unit (CAU) with a series of safety performance indicators (SPIs). All these SPIs are arranged in a three-level model based on the analytic hierarchy process (AHP) and Delphi method, called the DAHP model, which takes full advantage of the expert knowledge and quantitative calculation. The weight of each SPI is estimated by the DAHP model, while its score is monitored and measured quantitatively with the two values of the standard deviation and average values of the preceding historical data points. The proposed method was tested successfully on the real data of a regional CAU in China, reflecting the CAU's safety management state immediately and quantitatively.

© 2016 Elsevier Ltd. All rights reserved.

safety indicators to quantify the aviation risk which are caused by human errors (Shyur, 2008). Lee developed a quantitative model for assessing aviation safety risk factors as a means of increasing the effectiveness of safety risk management system by integrating the fuzzy linguistic scale method, failure mode, effects and criticality analysis principle (Lee, 2006).

In spite of plenty of standards, specifications and papers on SMS released by civil aviation organizations and scholars, these documents mainly focus on hazard identification and risk control, lacking of a detailed method for CAU's SPI monitoring and safety performance measurement. In this paper, a specific method is proposed for safety performance measurement based on the analytic hierarchy process (AHP) and Delphi method, called the DAHP model, providing the basement for CAU's safety management. The remaining of this paper is organized as follows. In Section 2, the DAHP model is described in detail. Then, Section 3 describes the standards for the SPI monitoring and measurement, and in Section 4, a case is discussed on the safety performance process of a regional CAU in China. Some conclusions close the paper in Section 5.

2. The DAHP model

In this section, the traditional AHP model is introduced firstly, and then the Delphi method is proposed to estimate the element values of the comparison matrix of the AHP model with the support of expert knowledge. Finally, the AHP and the Delphi method are combined, called the DAHP model, to measure the safety performance quantitatively.





^{*} Corresponding author. Office 505, China Academy of Civil Aviation Science and Technology, Jia31, Guangximen Beili, Chaoyang District, Beijing, 100028, China.

The method of AHP is based on the idea that a complex problem can be effectively examined if it is hierarchically decomposed into its components (Saaty, 1980, 2008). Thus, AHP provides a holistic view of the problem. AHP begins with the top level in the hierarchy that reflects the main objective. An element at a higher level of the hierarchy is said to be the governing element for those elements at the lower level. Elements at a certain level are compared against each other with reference to their effect on the governing element. Let us consider the elements $E_1, E_2, ..., E_n$ of some level in a hierarchy and denote their normalized weights by $w_1, w_2, ..., w_n$, respectively. The value of w_i reflects the degree of importance of the E_i element. The first step in the calculation of w_i is to derive pairwise comparisons between the *n* elements. These pairwise comparisons are structured into an $n \times n$ matrix called a comparison matrix

$$\mathbf{A} = \begin{array}{cccc} E_1 & E_2 & \dots & E_n \\ E_1 & \begin{bmatrix} a(1,1) & a(1,2) & \dots & a(1,n) \\ a(2,1) & a(2,2) & \dots & a(2,n) \\ \vdots & \vdots & & \vdots \\ a(n,1) & a(n,2) & \dots & a(n,n) \end{bmatrix} .$$
(1)

Elements of the matrix **A** can be derived using a nine-scale approach. The values of a(i, j) represent the importance comparison between the elements of E_i and E_j . More specifically, the value of a(i, j) is set to 1, 2, 3, ...and 9. Also, a(j, i) = 1/a(i, j) for all j = 1, 2, ..., n. In the nine-scale approach, if the element of E_i is more important than E_j , the value of a(i, j) is set to 2, 3, ...and 9. Conversely, if the element of E_i is more important than E_j , the value of a(i, j) is set to 2, 3, ...and 9. Conversely, if the element of E_i is more important than E_j , the value of a(i, j) is set to 1/2, 1/3, ...and 1/9. In case that the importance of the two elements are the same, the value of a(i, j) is set to 1. The weight of E_i , w_i , is the averaged and normalized value of all the elements in its row of the matrix **A**.

2.2. The Delphi method

The Delphi method is a structured communication technique, originally developed as a systematic, interactive forecasting method which relies on a panel of experts (Linstone and Turoff, 1975). In the standard version, the experts answer questionnaires in two or more rounds. After each round, a facilitator provides an anonymous summary of the experts' forecasts from the previous round as well as the reasons they provided for their judgments. Thus, experts are encouraged to revise their earlier answers in light of the replies of other members of their panel. It is believed that during this process the range of the answers will decrease and the group will converge towards the "correct" answer. Finally, the process is stopped after a pre-defined stop criterion and the mean or median scores of the final rounds determine the results. Two key issues of the Delphi method in performance measurement are discussed in the following paragraphs, including the design of expert questionnaire and the selection of experts.

In the expert questionnaires, the background knowledge of safety performance measurement is introduced firstly and then the Delphi method. Setting of the element values in the comparison matrices on all the levels is the main content. Additional questions are required to answer on the SPI rating standards, which is the score deduction standard in this paper. Moreover, suggestions should also be provided for the SPI selections.

Expert selection is another significant issue in the Delphi method, which should follow the principles of authority and universality. The experts should cover the specific fields of civil aviation, such as the operators of airlines, airports and air traffic service (ATS). The number of experts should also be properly set. A small number restricts the representativeness in subjects and area, while a large one results in management difficulties. Generally, it is suitable to invite 10-20 experts to answer the questionnaires.

2.3. DAHP

With the combination of the Delphi method and AHP, the DAHP model can take full advantage of the expert knowledge and quantitative calculation, overcoming the poor authority in the simple use of AHP. In this paper, the CAU SPIs are arranged on a three-level DAHP model. The weights of the former two levels could be estimated by the AHP method, where the element values of the comparison matrix on every level are estimated by the experts with the Delphi method.

Therefore, the synthetic weights of the sub-SPIs on the second level w_{ll}^{1-2} are calculated as

$$w_{ij}^{1-2} = w_i^1 \times w_j^2 \tag{2}$$

where w_i^1 denotes the weight of the indicator *i* on the first level, and w_j^2 denotes the weight of the indicator *j* on the second level belong to the indicator *i* on the first level. The weights of all the SPIs on the former two levels are estimated by the Delphi method described in Subsection 2.2. On the third level, the weights of all the indicators are equal in value. Then, all the weight elements are arranged in the weight vector \mathbf{W}^{1-2} . The details of the three-level DAHP model will be discussed in Section 3.

3. Safety performance measurement method

Safety performance results provide objective evidence for the regulator to assess the effectiveness of the CAU's SMS and to monitor achievement of its safety objectives. The CAU's SPIs should consider factors such as the safety consequences (result indicators), safety management and safety operation (process indicators), which are selected and developed in consultation with the CAU's regulatory authority. In this section, we propose a method for CAU's safety performance monitoring and measurement. The SPIs and associated targets should be accepted by the regulator responsible for the CAU's authorization, certification or designation. In Subsection 3.1, the SPI scores are measured with the proposed standard. Then, in Subsection 3.2, after the score measurement, all the SPIs are arranged in a three-level DAHP model framework to calculate the safety performance. In the three-level DAHP model, the safety performance indicators are arranged on a three-level framework. The weights of the indicators on each level are calculated with the comparison matrix of AHP, while the elements of the comparison matrix are decided by the experts after two or more rounds with the Delphi method.

3.1. SPI monitoring

In practice, the safety performance of an SMS could be expressed by quantitative SPIs on the third level of the DAHP model and their corresponding alert and target values. The CAU should monitor the performance of safety target indicators in the context of historical trends to identify any abnormal changes in safety performance. Likewise, target and alert settings should take into consideration of recent historical performance for a given indicator. Desired improvement of targets should be realistic and achievable for the CAU.

The target setting is a desired percentage improvement (in this case 5%) over the previous year's data point average. The

alert level for a new monitoring period (current year) is based on the preceding period's performance (preceding year), namely its data points average and standard deviation (SD). The three alert lines are average + 1 SD, average + 2 SD and average + 3 SD. An alert (abnormal/unacceptable trend) is indicated if any of the conditions below are met for the current monitoring period (current year): any single point is above the 3 SD line, 2 consecutive points are above the 2 SD line, or 3 consecutive points are above the 1 SD line. When an alert is triggered (potential high risk or out-of-control situation), appropriate followup action is expected, such as further analysis to determine the source and root cause of the abnormal incident rate and any necessary action to address the unacceptable trend.

If a more quantitative performance summary measurement is desired, appropriate score deduction may be assigned to each Yes/ No outcome for each target and alert outcome. The indicator rating standards could be obtained by the Delphi method. Suppose that the total score for each sub-sub-SPI is 100 on the third level, some score deduction standards are suggested:

- 1) 4 deducted when the 1 SD alert level breached once;
- 2) 5 deducted when the 2 SD alert level breached once;
- 3) 6 deducted when the 3 SD alert level breached once;
- 4) 40 deducted when target not achieved.

The final score for each sub-SPI on the second level could be calculated by

$$G_{ij}^{1-2} = 100 - G_{ij-d}^3 / N \tag{3}$$

where G_{ij}^{1-2} denotes the score for one of the sub-SPIs on the second level, and G_{ij-d}^{3} denotes the total deduction score for its *N* sub-sub-SPIs on the third level.

A range of SPIs of safety consequences as well as safety management and safety operation provide a more comprehensive insight into the CAU's safety performance. The SPIs are essentially data trending charts that track occurrences in terms of event rates. The SPIs of safety consequences could be measured by the rate of accidents and incidents (per 10 k flights), while those of safety management and safety operation could be measured by the work completion rate and equipment intact rate. This will ensure the comprehensive monitoring of outcome indicators and process indicators. Then, all the score elements are arranged in the score vector \mathbf{G}^{1-2} .

3.2. Safety performance measurement

The first level includes three indicators of safety targets, safety management and safety operation. On the second level, the indicator of safety consequences governs two sub-SPIs of the rate of accidents and incidents, as well as the rate of unsafe events. The sub-SPIs of safety management and safety operation are composed of several relative sub-SPIs. Furthermore, each sub-SPI on the second level includes several items to express the safety performance.

Once the SPIs and their corresponding targets have been defined, the performance outcome of each indicator should be updated and monitored on a regular basis. The target and alert level for each quantitative indicator may be tracked for their respective performance status. The grades of the qualitative indicators are given by the experts. A consolidated summary of the overall performance outcome of the complete SPIs package may also be aggregated for a given monitoring period. The overall performance score *P* is calculated as

$$P = \operatorname{sum}\left(\mathbf{G}^{1-2} \cdot \mathbf{W}^{1-2}\right) \tag{4}$$

where all the elements in \mathbf{W}^{1-2} and \mathbf{G}^{1-2} are calculated with Equations (2) and (3), respectively. In addition, the overall score and all sub-scores could be divided into different levels of excellent ($P \ge 95$), good ($90 \le P < 95$), pass ($80 \le P < 90$), and fail (P < 80).

4. A case study

To test the proposed method for safety performance monitoring and measurement with the three-level DAHP model, the case of a regional air traffic service (ATS) unit in China is studied with the real data of 2013.

4.1. Indicator system design

As shown in Table 1, the ATS unit safety performance considers three kinds of indicators on the first level, including the SPIs of safety consequences, safety management and safety operation.

On the second level, the indicator of safety consequences governs two sub-SPIs of the accident rate and the incident rate, and the unsafe event rate. The indicator of safety management is composed of eleven sub-SPIs, such as the planning and holding of safety meeting, the result of safety audit and the implement of safety responsibility et al. The performance of safety operation depends on four fields of air traffic control, aviation information, communication & navigation and aviation weather. Moreover, every sub-SPI is composed of a series of sub-sub-SPIs on the third level to check the progress of implementation. The indicator weights and an example of the monitoring and measurement process will be discussed in the following two sub-sections.

4.2. Weight calculation with the DAHP model

Ten experts were invited from the regional ATS units to answer the questionnaires. Their answers of the comparison matrices on the two levels converged after three rounds.

On the first level, the comparison matrix of the three indicators is as follows:

	1	2	3		
1	[1	1	1		
1 2 3	1 1 1	1	1	·	
3	1	1	1		
3	ΓI	I	I		

As shown above, the three indicators on the first level are equally important, so their weights are all 33.3%. The weights are provided in Table 1. It is revealed that the weight of the process indicators accounts for 2/3, exceeding that of the result indicators which is 1/3.

On the second level, the indicator of safety targets governs two sub-SPIs of accident and incident rate, as well as unsafe event rate, whose comparison matrix is

$$\begin{array}{c} 1.1 & 1.2 \\ 1.1 & \begin{bmatrix} 1 & 2 \\ 1/2 & 1 \end{bmatrix}. \end{array}$$

Therefore, their sub-weights are 66.7% and 33.3%, respectively. The indicator on the first level is index by "i" while its sub-SPI on the second level indexed by "i,".

W. Chen, J. Li / Journal of Air Transport Management 57 (2016) 228-233

Table 1	
Indicator weights of the former two levels.	

SPIs on the 1st level	Weight (%)	Sub-SPIs on the 2nd level	Sub-weight (%)	Synthetic weight (%)	Numbers of sub-sub-SPIs on the 3rd level
1. Safety consequences	33.3	1.1 Accident and incident rate	66.7	22.2	2
		1.2 Unsafe event rate	33.3	11.1	5
2. Safety management	33.3	2.1 Safety meeting	11.1	3.7	6
		2.2 Safety assessment	11.1	3.7	6
		2.3 Safety information management	11.1	3.7	11
		2.4 Safety education and training	11.1	3.7	9
		2.5 Incident investigation	11.1	3.7	4
		2.6 Emergency	11.1	3.7	8
		2.7 Safety document management	11.1	3.7	5
		2.8 Safety audit	22.2	7.4	4
3. Safety operation	33.3	3.1 Air traffic control	25	8.3	40
		3.2 Aviation information	25	8.3	36
		3.3 Communication & navigation	25	8.3	42
		3.4 Aviation weather	25	8.3	51

Similarly, the comparison matrices of the sub-SPIs of the safety management and safety operation are

	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	
2.1	<u>۲</u> 1	1	1	1	1	1	1	1/2 -	
2.2	1	1	1	1	1	1	1	1/2	
2.3	1	1	1	1	1	1	1	1/2	
2.4	1	1	1	1	1	1	1	1/2	and
2.5	1	1	1	1	1	1	1	1/2	
2.6	1	1	1	1	1	1	1	1/2	
2.7	1	1	1	1	1	1	1	1/2	
2.8	2	2	2	2	2	2	2	1 _	
	3.1	3.2	3.3	3.4					
3.1	<u>۲</u> 1	1	1	ך 1					
3.2	1	1	1	1 .					
3.3	1	1	1	1					
3.4	[1	1	1	1					

The corresponding sub-weights are shown in Table 1. The synthetic weights are also calculated with Formula (2). The numbers of the sub-sub-SPIs on the third level are also revealed in Table 1.

4.3. Examples of SPI monitoring and performance measurement

As for the SPI list of the ATS unit, twelve ATS experts of the different operation fields were invited to answer the questionnaires. Their answers converged after three rounds. Three examples of the ATS unit's sub-sub-SPI measurements and monthly monitoring chart are shown in Table 2 and Fig. 1, belonging to three kinds of SPIs.

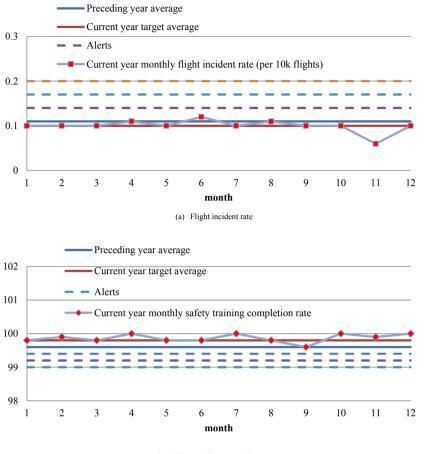
One sub-sub-SPI of the safety consequences considers the ATS operator monthly flight incident rate (per 10 k flights). One sub-sub-SPI of the safety management considers the ATS operator monthly safety training completion rate. One sub-sub-SPI of the safety operation considers the ATS operator monthly intact rate of the communication and navigation equipment. The SPI targets of 2013 are set according to the data average and SD of 2012 with 3%–5% improvements. The monitoring data shows that the targets of the three sub-sub-SPIs are all achieved. The alert levels of the first two sub-sub-SPIs are never breached, while the 1 SD and 2 SD alert levels of the third sub-sub-SPIs are estimated with the score deduction standards and Formula (3) proposed in Section 3.1.

In Fig. 1, the preceding year average and the current year target average are represented with solid lines, while the three sub-sub-SPI alert levels are represented with dashed lines. The current year target average is usually set according to the preceding year average with 3%–5% improvements, while the sub-sub-SPI alert levels are set on the basis of the preceding year average with a requirement reduction of 1 SD, 2 SD or 3 SD. The monthly sub-sub-SPI monitoring data are also plotted in the chart and compared with these five standard lines. The score deduction standards are suggested in Section 3.1. Fig. 1(a) shows the monthly monitoring result of the flight incident rate, where the current year target average is lower than the preceding year while the three alerts are higher. Fig. 1(b) and (c) illustrate the monthly monitoring result of the

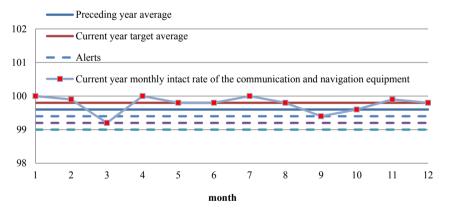
Table 2

Examples of the ATS unit's quantitative SPI measurements.

SPI description	SPI alert level criteria	iteria Times of alert level breached		vel	SPI target level criteria	Target achieved	Score
Safety consequences	_				-	-	
ATS operator monthly flight incident rate (per 10 k flights)	Average + 1/2/3 SD (annual reset)	1 SD 0		3 SD 0	<u>3%</u> improvement between each annual mean rate	Yes	100
Safety management							
ATS operator monthly safety training completion rate	Average-1/2/3 SD (annual reset)	1 SD 0	2 SD 0	3 SD 0	5% improvement between each annual mean rate	Yes	100
Safety operation							
ATS operator monthly intact rate of the communication and navigation equipment	Average-1/2/3 SD (annual reset)	1 SD 1	2 SD 1		$\frac{5\%}{100}$ improvement between each annual mean rate	Yes	91



(b) Safety training completion rate



(c) Intact rate of the communication and navigation equipment

Fig. 1. Examples of the monthly sub-sub-SPI monitoring chart of an ATS unit in 2013.

safety training completion rate and the intact rate of the communication and navigation equipment, where the current year target average is higher than the preceding year while the three alerts are lower.

Table 3 reveals the overall safety performance measurement of the ATS unit. Firstly, for each sub-SPI on the second level, the numbers of the times of alert level breached are counted, as well as the times of the target achieved (yes or no). Then, according to score deduction standards and Formula (3), the mean scores are calculated for each sub-SPI. Finally, based on Formula (4), the synthetic score of each sub-SPI is obtained according to its synthetic weight. Therefore, the total score is 96.25, representing that the safety performance level of this ATS unit is excellent. However, it is shown in Table 3 that the scores of six highlighted sub-SPIs on the second level are lower than the total score. Therefore, more attention should be paid to these sub-SPIs in the next monitoring period to improve its safety performance.

5. Conclusion

The theory of SMS has been the guidance of the CAU safety management in recent years. With the promotion of SMS, the

Table 3 Safety performance measurement of an ATS unit.

Sub-SPI	Times of alert level breached on the 3rd level			Times of target achieved		Mean score on the 2nd level (G_{ij}^{1-2})	Synthetic weight (%)	Synthetic score	
	1 SD	2 SD	3 SD	Yes	No				
1.1 Accident and incident rate	1	0	0	2	0	98	22.2	21.75	
1.2 Unsafe event rate	2	0	0	5	0	98.4	11.1	10.92	
2.1 Safety meeting	2	1	0	5	1	91.16	3.7	3.37	
2.2 Safety assessment	0	0	0	6	0	100	3.7	3.7	
2.3 Safety information management	2	2	0	11	0	98.36	3.7	3.63	
2.4 Safety education and training	3	1	0	8	1	93.66	3.7	3.46	
2.5 Incident investigation	0	0	0	4	0	100	3.7	3.7	
2.6 Emergency	1	0	0	8	0	99.5	3.7	3.68	
2.7 Safety document management	3	2	0	4	1	87.6	3.7	3.24	
2.8 Safety audit	1	0	0	4	0	99	7.4	7.32	
3.1 Air traffic control	8	5	2	38	2	96.27	8.3	7.99	
3.2 Aviation information	10	6	4	32	4	92.94	8.3	7.71	
3.3 Communication and navigation	15	9	5	40	2	94.88	8.3	7.87	
3.4 Aviation weather	12	8	5	48	3	95.33	8.3	7.91	
Total	60	34	16	215	14		100	96.25	

simple meeting of regulatory safety requirements should be changed to quantitative safety performance monitoring and measurement. In this paper, a specific method is proposed for safety performance measurement based on the DAHP model. Some conclusions are given as follows:

- 1) The combination of AHP and Delphi method takes full advantage of the expert knowledge and mathematical model.
- 2) The safety performance measurement is more feasible and quantitative with the monitoring of the SPIs of work completion rate and equipment intact rate.
- 3) It is more reasonable to monitor the SPIs under the standards of the target and alert level monthly than the subjective judgments of experts at the end of the year.
- 4) The SPIs, weights, score deduction standards and score division standard proposed in this paper are for reference only. The civil aviation units should make their own SPI system and management process according to their specific situations.

Acknowledgements

The research work was supported by the 2013 Special Task Item of CAAC and the National Key Research and Development Program (2016YFC0800406).

References

- Eurocontrol, May, 2014. Performance Review Report: An Assessment of Air Traffic Management in Europe during the Calendar Year 2013.
- FAA EUROCONTROL, Nov., 2014. Comparison of Air Traffic Management-related Operational Performance: U.S./Europe.
- Federal Aviation Administration, Dec., 2014. Fiscal Year 2014 Performance and Accountability Report.
- International Civil Aviation Organization, 2013. Safety Management Manual, third ed. Doc 9859.
- Lee, W.K., 2006. Risk assessment modeling in aviation safety management. J. Air Transp. Manag. 12 (5), 267–273.
- Linstone, H.A., Turoff, M., 1975. The Delphi Method: Techniques and Applications. Addison-Wesley, Reading, Mass. ISBN 978-0-201-04294-8.
- Luo, F., Xiao, W.J., Zhou, F.H., et al., 2009. Research on the Risk assessment model of hazards for air traffic management. In: Proceedings of the IASTED International Conference on Modelling, Simulation, and Identification.
- Luo, F., Xiao, W.J., Zhou, F.H., 2009. The procedures design of hazards management and control for air traffic management. In: Proceedings of the 6th International Conference on Innovation & Management.
- Saaty, T.L., 1980. The Analytic Hierarchy Process. McGraw-Hill, New York.
- Saaty, T.L., 2008. Decision making with the analytic hierarchy process. Int. J. Serv. Sci. 1 (1), 83–98.
- Shyur, H.J., 2008. A quantitative model for aviation safety risk assessment. Comput. Ind. Eng. 54 (1), 34-44.

Weishi Chen was born in 1982 in Tianjin, China. He received his Ph.D. degree from Beihang University in 2009. Now he is a Senior Engineer in Airport Research Institute, China Academy of Civil Aviation Science and Technology. His current research focuses on airport safety, risk assessment and safety management system.