Neural network-based probability forecasting method of aviation safety

Zhu Zheqi¹, Ren Bo¹*, Zhang Xiaofeng¹, Zeng Hang¹, Xue Tao² and Chen Qingge¹

¹ Equipment Management and UAV Engineering College, Air Force Engineering University, Xi’an 710051, Shaanxi, China
² Zhengzhou Regional Military Representative Office of Air Force, 450000, Zhengzhou, China

E-mail: rabber2003@163.com

Abstract. Aviation safety forecasting is of great significance for accident prevention. At present, aviation safety forecasting is mainly deterministic forecasting, ignoring the impact of various uncertainties on forecasting. In this paper, on the basis of deterministic forecasting, the forecasting of aviation safety probability is carried out based on the uncertainty of neural network point forecasting value. The uncertainty of aviation safety forecasting is described by three ideas: the numerical statistical characteristics of point forecasting value, the probability density fitting of point forecasting value and the distribution of error. The interval forecasting result is obtained, which can better understand the uncertainty and risk of forecasting quantity. Taking the aviation safety data of civil aviation from 1993 to 2008 as an example, the results show that this method can provide the most possible range of aviation safety forecasting under certain uncertainty level, and is more conducive to the modeling of aviation safety uncertainty.

1. Introduction
Aviation industry has high technology content and high safety risk. Once there is an aviation accident, there will be a large number of deaths and a strong destructive force, which will cause a huge negative impact. Therefore, it is an important topic for experts and scholars to avoid aviation accidents and improve aviation safety level to reduce the accident rate. Aviation safety forecasting studies the internal relationship between aviation accidents and causal factors, reveals the law of accidents, and realizes the forecasting of future trends [1]. Aviation safety forecasting includes time series forecasting and causal forecasting. Among them, the causal forecasting is based on the analysis of accident causes, focusing on the relationship between the causal factors and accidents. Luo fan, Zhang Sheng [2-4], etc. analyzed the causal relationship between aviation accidents and air traffic control system, obtained the causal factors such as organization management, environment, human error, equipment, etc., and constructed the aviation accident forecasting model. Salah and Gurjeet [5] studied the aviation safety indicators, and believed that pilots, maintenance, safety management system and active safety culture were the main factors affecting aviation safety. In addition, time series forecasting uses the statistical characteristics of unsafe event time series to reflect the correlation of unsafe events in time sequence and estimate the future trend of event series. Common methods include regression analysis [6], exponential smoothing [7], grey forecasting [8], artificial
intelligence [9], etc. Gan Xusheng and Duanmu Jingshun [10] used autoregressive moving average method to study flight accidents. Juan [11] uses spline interpolation function to analyze the data of Taiwan flight safety management system, studies the relationship between human factors and aviation safety, and improves the applicability of nonlinear data. Lv Xuemei [12] and others used the grey model to predict the ten thousand hour rate and accident symptoms of serious flight accidents in the future.

However, the two kinds of aviation safety forecasting belong to the deterministic forecasting [13], and the influence of the uncertainty is not fully considered (the uncertainty of the causal factors, the uncertainty of the forecasting model and the algorithm, the uncertainty of the accident environment, the resource conditions, the technical development status, etc.). At present, the deterministic point forecasting of aviation safety forecasting can not accurately reflect the randomness of aviation safety, and only one deterministic point forecasting curve is not enough to describe the uncertainty of aviation safety [14]. Moreover, the uncertainty in the transmission and output of aviation safety model greatly affects the accuracy of aviation safety forecasting and early warning. Confidence interval estimation [15] is a common method to describe the uncertainty of real value relative to forecasted value. Confidence interval estimation uses confidence interval to express the accuracy of estimation and confidence degree to express the reliability of estimation [16], which can better describe the uncertainty of aviation safety forecasting and assist the decision-making of aviation operation.

In view of this, this paper explores the uncertainty factors of the forecasting results to carry out the aviation safety forecasting based on the certainty point forecasting. Firstly, the results are forecasted many times by BP neural network. Aiming at the forecasting results, the most possible confidence interval under a certain confidence probability is obtained by the uncertainty description method, and the changes of the aviation safety forecasting results caused by the uncertainty factors are quantified, so as to determine the reliability degree of aviation safety forecasting value included in the interval, and better understand the uncertainty and risk that the forecasted quantity may face in the future change[19].

2. Point forecasting of aviation safety based on Neural Network

2.1. Aviation safety forecasting.

The research of aviation safety forecasting mainly includes: aviation safety risk forecasting, aviation safety early warning and accident prevention.

Among them, aviation safety forecasting reveals the law and trend of accident development by using management, mathematics and statistics methods, based on the cause and change law of unsafe events, through statistical analysis of the data of unsafe events that have occurred. Time series forecasting studies the influence relationship between the occurrence rule of accidents at the previous time and the occurrence trend of accidents in the future. By using the statistical rule of unsafe event time series, it reflects the correlation of unsafe events in time sequence and estimates the future trend of event series. It is defined as follows:

$$Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \cdots + \phi_k Z_{t-k} + \xi_t + \mu_1 \xi_{t-1} + \mu_2 \xi_{t-2} + \mu_3 \xi_{t-3} + \cdots + \mu_k \xi_{t-k}$$ (1)

Among them, $Z = (z_1, z_2, z_3, \cdots, z_n)$ represents the aviation unsafe event at time $t$, $\xi_t$ is a stationary time series, and $\xi_t$ is a white noise series.

At present, the phase time series forecasting is only aimed at the time series of aviation unsafe events, ignoring the influence of time series correlation on unsafe events. Moreover, neither causal forecasting nor time series forecasting takes into account the wide range of uncertainties in causal factors, the impact of errors caused by uncertainty on forecasting, nor the evaluation of forecasting and early warning output based on uncertainty.
2.2. **Point forecasting of aviation safety based on BP neural network.**

The error back propagation algorithm of artificial neural network is a non-linear geometric mathematical calculation equation that can determine the nonlinear mapping relationship between the input value of artificial neural network and the expected network output without prior design. Only through its own artificial intelligence training, learn the calculation rules of a certain artificial neural network, and the correct calculation is closest to the actual expectation of a given artificial neural network, it is possible to get the nonlinear results closest to the actual expected output value of the artificial neural network.

By constructing the input-output relationship between the factor data and the result data that affect the civil aviation security situation, and using BP neural network to learn some rules, the point forecasting value of the forecasting target can be obtained. Moreover, since the initial weights and thresholds of the network are generated randomly, the results of each training are different, which provides uncertain factors.

3. **Probability forecasting of aviation safety**

3.1. **Basic theory of probability forecasting.**

Probability forecasting can be understood as a general term of interval forecasting and density forecasting, which is different from point forecasting (expectation forecasting). Interval forecasting refers to the forecasting of the target's confidence interval at a certain time point or under certain conditions in the future, which can obtain the approximate fluctuation range of the forecasted target; while density forecasting refers to the forecasting of all probability information of the forecasted target at a certain time point or time period in the future, which can quantitatively describe the probability of the forecasted object taking a certain value.

The conventional point forecasting method will transform the forecasting object into the determinate variable, so the determinate forecasting value can be obtained from it. In the process of analysis, it is assumed that the forecasting value is consistent with the actual value at the forecasting time, and the forecasting result obtained has no corresponding probability characteristics. But the viewpoint of probability forecasting is different. It thinks that the forecasting object can't be determined, so it needs to describe the data sequence of forecasting object based on random sequence, that is to say, the value of forecasting object at the time of forecasting has uncertainty. Generally, the goal can be achieved by using probability forecasting method, that is, in a certain range, the cumulative probability of all forecasted objects can be obtained. Of course, the confidence level (cumulative probability) can be determined first, and then the corresponding forecasting interval can be output through probability forecasting.

3.2. **Research significance of aviation safety probability forecasting.**

At present, most of the aviation safety forecasting is the deterministic point forecasting. Using the deterministic forecasting model to predict the aviation safety data information can not transfer the probability reliability of the results, which is not suitable for the application of risk analysis and control. Some point forecasting methods have achieved good forecasting results under the corresponding specific conditions, but restricted by their inherent defects, their forecasting effect robustness is deficient in more general scenarios. Compared with deterministic forecasting, probabilistic forecasting is more valuable at a reasonable risk level because it can give a probabilistic conclusion.
4. Probability forecasting of aviation safety based on statistical characteristics of forecasting value

4.1. Normal distribution

The normal distribution was first obtained by Abraham de Moivre in the asymptotic formula for binomial distribution. C. F. Gauss derived it from another angle when he studied the measurement error. P. S. Laplace and Gauss studied its properties. It is a very important probability distribution in the fields of mathematics, physics and engineering, and has a great influence in many aspects of statistics.

If the random variable obeys a probability distribution with position parameter and scale parameter, and its probability density function is:

\[ f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \]  

Then the random variable obeys the normal random variable, namely: \( X \sim N(\mu, \sigma^2) \).

4.2. Realization steps of probability forecasting of aviation safety based on statistical characteristics of forecasting value.

For the description of data uncertainty, only getting its probability density function can not fully mine the information it contains, especially when the forecasting object is continuous data, the probability at a single point does not exist. Therefore, interval forecasting results are the most common representation of probability forecasting. When the forecasted value samples obey the normal distribution, the confidence interval is obtained when the reliability level of the forecasted results is 90%, that is to say, when the reliability level of the forecasted results is 90%, the future aviation safety trend is most likely to fall into the interval. The results of interval forecasting can also be solved by using the idea of function fitting without assuming the distribution function of forecasting value.

In conclusion, the specific steps of aviation safety probability forecasting based on the statistical characteristics of BP neural network forecasting values are as follows:

Step 1: Based on the established input-output relationship between the aviation safety influencing factors and the result factors, the output value under the given time condition is forecasted multiple times based on the neural network training of a certain rule.

Step 2: according to the statistical distribution characteristics of multi-point forecasting value samples, based on the distribution characteristics, under certain confidence conditions, get the interval forecasting results.

Step 3: fit the probability density curve of the forecasting value of multiple points to obtain multiple optimal forecasting values, and further fit the probability density curve of each optimal value.

Step 4: Based on the principle that the interval coverage PICP should be as close as possible to the preset rated confidence level PINC, the interval results are estimated again by integrating the probability density curve.
5. Forecasting of aviation safety probability based on error density

5.1. The generation of forecasting error of BP neural network.
There are many reasons for the prediction error of BP neural network.

Firstly, neural network is an intelligent approximate calculation. The output of BP neural network model according to its own training is a kind of inaccuracy, which can not be completely consistent with the actual standard value. Even though the model is quite accurate, the actual value is also uncertain, not the exact function relationship, so the difference caused by the model is inevitable, but it is also completely acceptable in the forecasting work.

Secondly, there are many algorithms and parameters that affect the neural network, such as the number of neurons in the hidden layer, whether the learning rate changes, whether the driving amount, and so on. At present, there is no specific method to calculate these factors. Generally, the more satisfactory network parameters can be obtained by trial method after many times of experiments.

Finally, data processing causes differences; in the process of data processing in Section 3.2, one of the ideas is to assume that multiple samples of forecasting results are subject to normal distribution, which is an ideal assumption in itself. Secondly, when fitting the probability density curve, the interpolation method is also used. It is a simple calculation for the result, and it will cause a series of errors.

5.2. Realization steps of aviation safety probability forecasting based on error value density.
For the forecasting of a certain time, using the same model to predict the error value of the previous nearby points has more reference value for the error evaluation of the time forecasting value. Therefore, it is considered to attach the error distribution before the forecasting time to the point forecasting value of the time to get the interval forecasting result.

To sum up, the specific steps of aviation safety probability forecasting based on BP neural network error value density are as follows:

Step 1: Based on step 3 of Section 3.2, statistics and sorting out the best fitting forecasting values of each group;
Step 2: compare the forecasted value with the real value to get the error values;
Step 3: fit the probability density curve of error value obtained in step 2, and obtain the interval forecasting result of error value under certain confidence level.
Step 4: add the error interval forecasting result to the point forecasting value to get the final interval forecasting result.
6. Example

6.1. Civil aviation safety case data.
The forecasting model proposed in this paper is applied to the forecasting of civil aviation safety. Taking civil aviation safety data from 1993 to 2008 as an example, the effectiveness of BP neural network in aviation safety probability forecasting is verified by its actual aviation safety data and aviation safety forecasting data. The aviation safety data of civil aviation statistics from 1993 to 2008 are normalized, and the uncertainty of the forecasted value (number of accident symptoms) from 2005 to 2008 is forecasted based on the historical data.

6.2. Forecasting of aviation safety data based on BP neural network.
Using BP neural network and 1993-2007 data as training set, the normalized value (real value is 0.6000) of the number of accident symptoms in 2008 was forecasted for 51 times, and the corresponding forecasting points were obtained as shown in table 1:

Table 1. forecasting results of normalized value of accident symptoms in 2008 Lists.

| No | Normalized forecast value | No | Normalized forecast value | No | Normalized forecast value |
|----|---------------------------|----|---------------------------|----|---------------------------|
| 1  | 0.7037                    | 18 | 0.5838                    | 35 | 0.6345                    |
| 2  | 0.6038                    | 19 | 0.5648                    | 36 | 0.6200                    |
| 3  | 0.5681                    | 20 | 0.6863                    | 37 | 0.4683                    |
| 4  | 0.6368                    | 21 | 0.6145                    | 38 | 0.5881                    |
| 5  | 0.6243                    | 22 | 0.7219                    | 39 | 0.5958                    |
6.3. Probability forecasting of aviation safety based on statistical characteristics of forecasting value.

For the 51 forecasted value samples obtained above, the frequency distribution histogram of the forecasted value is drawn, as shown in Figure 3.

![Figure 3. Distribution histogram of normalized forecasting value of accident symptoms in 2008.](image)

It can be seen from Figure 3 that the forecasted number of accident symptoms in 2008 roughly follows the normal distribution. According to this group of samples, the mean value, variance and standard deviation of samples are 0.6012, 0.00334 and 0.0578 respectively. In this example, the population variance is unknown. According to the method of determining the confidence interval of the single normal population mean, the unbiased estimator $S^2$ of $\sigma^2$ can be used instead of $\sigma^2$ to construct the pivot variable $T = \frac{(\bar{X} - \mu)}{(S/\sqrt{n})} \sim t(n-1)$. For a given confidence level of $1-\alpha$, the two-sided confidence interval of $1-\alpha$ of the mean $\mu$ is $\left[\bar{X} - t_{\alpha/2}(n-1) \cdot S/\sqrt{n}, \bar{X} + t_{\alpha/2}(n-1) \cdot S/\sqrt{n}\right]$. On this basis, when the confidence level $\alpha = 0.9$, the corresponding interval forecasting result of aviation safety forecasting is [0.5848, 0.6176], and the result of anti normalization is [117, 123].

The first 50 of the above 51 sample points are selected for processing to solve the problem of interval forecasting by density forecasting. In this paper, 50 pieces of data are mismatched and combined to form 10 groups of continuous data. By using the ksdensity core smooth density estimation function in MATLAB, the probability density curve of each group of data is fitted, and the
10 most likely forecasted values of forecasted values are obtained. On this basis, the above 10 most likely values are fitted with probability density function, and the results are shown in figure 4.

![Figure 4](image-url)  
**Figure 4.** Forecasting value density curve of normalized value of accident symptoms in 2008.

According to the principle that the coverage rate of forecasting interval should be close to the preset rated confidence level as much as possible, the above curves are integrated by MATLAB, and two endpoints of the required interval are obtained. The two endpoints meet the requirements that the ordinate is equal and the integration value in the interval is the value of confidence level. Through many calculations, when the ordinate is 8, i.e. the upper and lower limits of the interval are 0.5766 and 0.6285 respectively, the integral value in the interval is 90%. Therefore, the interval [0.5766, 0.6285] is used as the interval forecasting result under the 90% confidence level. The interval forecasting result of accident symptoms in 2008 is [116, 125].

In the same process, under the condition of 0.9 confidence level, the forecasting results of the interval of the number of accident symptoms from 2005 to 2008 are obtained, as shown in Fig. 5 and Fig. 6.

![Figure 5](image-url)  
**Figure 5.** Forecasting results based on normal distribution characteristics of forecasted values.
Figure 6. Forecasting results based on probability density fitting of forecasting values.

6.4. Probability forecasting of aviation safety based on error density.
BP neural network is used to predict the normalized value of accident symptoms in 2005-2007. After fitting, 10 best forecasting values and their corresponding error values are obtained, as shown in Table 2.

Table 2. Error Statistics and Normalized Lists.

|    | Normalized forecast value | Error value   |    | Normalized forecast value | Error value   |    | Normalized forecast value | Error value |
|----|---------------------------|---------------|----|---------------------------|---------------|----|---------------------------|-------------|
| 1  | 0.52174                   | 0.05826       | 1  | 0.58472                   | -0.00028      | 1  | 0.56174                   | -0.01826    |
| 2  | 0.51722                   | 0.06278       | 2  | 0.58837                   | 0.00337       | 2  | 0.55845                   | -0.02155    |
| 3  | 0.52455                   | 0.05545       | 3  | 0.58506                   | 0.00006       | 3  | 0.56834                   | -0.01166    |
| 4  | 0.54707                   | 0.03293       | 4  | 0.57734                   | -0.00766      | 4  | 0.57113                   | -0.00887    |
| 5  | 0.57022                   | 0.00978       | 5  | 0.56678                   | -0.01822      | 5  | 0.57323                   | -0.00667    |
| 6  | 0.58266                   | -0.00266      | 6  | 0.56771                   | -0.01729      | 6  | 0.57559                   | -0.00441    |
| 7  | 0.59291                   | -0.01291      | 7  | 0.56877                   | -0.01623      | 7  | 0.57982                   | -0.00018    |
| 8  | 0.59132                   | -0.01132      | 8  | 0.56650                   | -0.01850      | 8  | 0.58598                   | 0.00598     |
| 9  | 0.56426                   | 0.01574       | 9  | 0.57364                   | -0.01136      | 9  | 0.56721                   | -0.01279    |
| 10 | 0.50773                   | 0.07227       | 10 | 0.59002                   | 0.00502       | 10 | 0.57279                   | -0.00721    |

The above 30 error values are fitted with probability density curve, as shown in figure 7.
Figure 7. Probability density curve of error value.

According to figure 7, when the ordinate is 5.4, the upper and lower limits of the corresponding error values are -0.0164 and 0.0602, and the integral value of the probability density curve in this interval is close to 90% (the actual accurate calculation value is 89.89%). The result shows that the confidence interval of error value at 90% confidence level is [-0.0164, 0.0602]. The final y-value in 2008 is [0.5589, 0.6355] at 90% confidence level, and the forecasting interval of the number of accident symptoms in 2008 is [112, 125] by inverse normalization.

Thus, the interval forecasting results based on error distribution from 2005 to 2008 are shown in figure 8.

Figure 8. Forecasting results based on the density distribution of error values.

7. Conclusion
The uncertainty of forecasted value seriously affects the accuracy of aviation safety forecasting. In this paper, the BP neural network model is used to predict the aviation safety trend. According to the statistical characteristics of multiple point forecasting, the probability density of point forecasting and the error distribution, the uncertainty is described on the basis of the deterministic forecasting results, and the given confidence is calculated respectively. It is more conducive to the uncertainty modeling of aviation safety, aviation safety risk decision-making and aviation safety risk management.
Generally speaking, this study basically achieves the goal of improving the forecasting accuracy: on the one hand, the range forecasting results are more in line with the objective reality, because any forecasting method is difficult to achieve zero error; on the other hand, the determination of the confidence level in advance controls the forecasting risk and ensures the reliability of the results.

Acknowledgments
Financial supports from the National Natural Science Foundation of China (No. NSFC71701210), the Natural Science Basic Research Plan of Shaanxi Province, China (No. 2019JQ-710), and Aviation Science Fund (No.20165196017) are gratefully acknowledged.

References
[1] Duanmu Jingshun, Ma Zhenyu and Chang Hong, Theoretical Methods of the Aviation Accident forecasting. Early-Warning and Prevention (Bei Jing: National Defense Industry Press)
[2] Luo Fan and Jia Gang. 2006 Factor Analysis of Early Warning Indexes for Airlines Organizational Management. J. Journal of Wuhan University of Technology (Information & Management Engineering) 10(28) 93–100
[3] Li Ke and Zhang Sheng. 2007 Management Information System of Safety Early Warning for an Air Traffic Control Center. J. Journal of Wuhan University of Technology (Information & Management Engineering) 29(8) 106–10
[4] Wang Bentao. 2014 Study on the Early Warning Model for Ground Service Personnel Human Error of Airport Airfield D Wuhan University of Technology
[5] Md. Salah Uddin Rajib and Luo Fan. 2015 A study on the critical factors of human error in civil aviation: An early warning management perspective in Bangladesh. J. Management Science Letters. 5 21–8
[6] Du Yi. 2010 Regression analysis of 10000 hours rate of civil aviation accidents. J. Journal of civil aviation safety of China 21(1) 42–4
[7] Du Hongbing and Qin Penghui. 2011 Time sequence extrapolative analysis model of short-term forecasting of U.S. general aviation accidents J. Journal of Safety and Environment. 11(3) 208–11
[8] Wang Yanyang and Cao Yihua. 2011 Analysis and prediction of civil aviation safety index results J. Journal of Beijing University of Aeronautics and Astronautics 37(10) 1223–8
[9] Ding Songbin and Wang Fei. 2006 Study on prediction method of civil aviation safety based on BP neural network J. Journal of CAAC 24(01) 53–6
[10] Gan Xusheng and Duanmu Jingshun. 2012 Grey time series combination prediction model of aviation equipment accident J. China Security Science Journal. 22(4):32–8
[11] Huan-Jyh Shyr. 2008 A quantitative model for aviation safety risk assessment. J. Computers and Industrial Engineering. 54 34–44
[12] Lu Xuemei. 2008 Study on prediction of unsafe events in Airlines D. China Civil Aviation University
[13] Liu Bin, Zhou Jingyang, Zhou Haiming and et al. 2012 An improved error distribution model for wind power prediction J. East China power. 40(2) 286–91
[14] Nan Xiaqiang and Li Quanzhan. 2013 Energy storage power and capacity allocation method considering wind power prediction error distribution J. Power automation equipment. 33(11) 117–22
[15] Wu Xiong and Wang Xiuli. 2013 Confidence interval estimation of combined output of new energy J. Power system automation 37(16) 7–12
[16] Xu man, Qiao Ying and Lu Zongxiang. 2011 Comprehensive evaluation method of short-term wind power prediction error J. Power system automation.35(12) 20-6
[17] Hyndman RJ. 1996 Computing and graphing highest density regions J. The American Statistician. 50(2) 120–6
[18] Kim JH, Silvapulle P and Hyndman RJ, 2007 Half-Life Estimation based on the Bias-Corrected Bootstrap: A Highest Density Region Approach J. Computational Statistics and Data Analysis, 51(7) 3418–432

[19] Zhang Guoyong, Wu Yonggang, Zhang Yang and et al, 2014 A simple probability interval prediction model of wind power chaotic time series J. Acta physica Sinica. 63(13) 422–30