Correlation analysis method of track irregularity indexes

Shiyi Li  
China Railway Eryuan Engineering Group Co., Ltd., Chengdu, Sichuan, 610031, China  
email: 153147555@qq.com

Abstract. Research on the correlation analysis method of track irregularity indexes has an important guiding role in selecting targeted maintenance decision-making indexes and guiding maintenance decision making related to track irregularity state. In this study, the track irregularity index structures and track quality index were used to design the correlation analysis method of track irregularity indexes. To verify the effectiveness of this method, a total of 48 track inspection car historical data from September 2014 to February 2017 and November 2017 to December 2018 of the Shenmu-Shuozhou railway Hexi transportation division was used as the case study. The results show that the correlation analysis method proposed in this study can be used to effectively analyze the correlation between the track irregularity indexes. This is verified by comparing with the correlation analysis method based on the standard deviations of track irregularity indexes.

1. Introduction
Railway track irregularities are closely related to the safe operation of trains and track maintenance costs[1-2]. For the track segment of a specific length (generally 200 m), the track quality index ($TQI$) is usually used to describe the overall track irregularity state. The $TQI$ is the sum of the standard deviations of the seven single track irregularity indexes such as the gauge ($G$), the cant ($C$), the left longitudinal level ($LLL$), the right longitudinal level ($RLL$), the left alignment ($LA$), the right alignment ($RA$), and the twist ($T$) for the track segments. The $TQI$ can effectively reflect the track irregularity degradation, and each single track irregularity index can also reflect the degradation of the corresponding track irregularity. The larger the value of $TQI$ and each single track irregularity index for the track segment, the worse the track irregularity state and degradation[3]. The $TQI$ and the seven track irregularity indexes reflect different characteristics. Therefore, the research on the correlation analysis of track irregularity indexes (i.e., $TQI$, $G$, $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$) can be used to analyze the degree of correlation between various track irregularity indexes, and the influence of each single track irregularity index on $TQI$. On this basis, the track irregularity state can be evaluated more accurately and effectively, which has an important guiding role in selecting targeted maintenance decision-making indexes and guiding maintenance decision making related to track irregularity state.

Many studies have been carried out on the correlation analysis of track irregularity indexes. Yang et al.[4] analyzed the correlation between the $TQI$ and each single track irregularity index for the Shanghai-Nanjing intercity railway. And based on the correlation analysis, the contribution of each single track irregularity index to $TQI$ was analyzed. The results show that the $TQI$ has a significant positive correlation with the longitudinal level, the cant, and the twist, indicating that the change of $TQI$ is mainly caused by the change of vertical irregularities (including the longitudinal level, the cant, and the twist). Zuo and Xiang[5] analyzed the correlation between the longitudinal level, the alignment, and the cant
based on the historical data of the Zhengzhou-Wuhan railway. The results show that the three track irregularity indexes are weakly correlated, and can be approximately regarded as independent of each other. Jin and Zeng[6] analyzed the correlation between the left longitudinal level and the right longitudinal level, and the correlation between the left alignment and the right alignment, based on the track inspection car historical data of the Beijing-Tianjin intercity railway. Lasisi and Attoh-Okine[7] analyzed the correlation between the track irregularity indexes such as the left alignment and the right alignment, and based on the correlation analysis, the effectiveness of the main components of track irregularity indexes selected in the study were analyzed and verified.

The standard deviations were used as the values of each single track irregularity index in previous studies, which are only the absolute amount of each index. And the consideration of overall structure of the seven track irregularity indexes (i.e., $G$, $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$) is lacking. The conclusions in these previous studies can only explain the relationship between the changes of the absolute amount of these track irregularity indexes, but cannot fully reflect the contribution of each single track irregularity index to $TQI$, and the influence on $TQI$, as well as the degree of correlation between the $TQI$ and the seven track irregularity indexes. In this study, track irregularity index structures[2, 8] and $TQI$ were used to design the correlation analysis method of track irregularity indexes for the 200 m tangent segments. The track inspection car historical data of the Shenmu-Shuozhou railway was used as the case study. And the effectiveness of the correlation analysis method in this study is verified by comparing with the correlation analysis method based on standard deviations of each single track irregularity index.

2. Design of correlation analysis method of track irregularity indexes

2.1. Design ideas of correlation analysis method

For the 200 m tangent segments[9], using the track irregularity index structures and $TQI$ as research subjects, the correlation between various track irregularity indexes was studied, and the correlation between the seven track irregularity indexes (i.e., $G$, $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$), the correlation between each single track irregularity index and $TQI$, and the degree of influence of each single track irregularity index on $TQI$ were analyzed.

The track irregularity index structures refer to the composition of the seven track irregularity indexes (i.e., $G$, $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$) for the 200 m segments within a certain period of time, denoted by $P=(p_1, p_2, \cdots, p_7)$, and $\sum_{i=1}^{7} p_i = 1 \ (0 < p_i < 1)$. $p_1$, $p_2$, $p_3$, $p_4$, $p_5$, $p_6$, and $p_7$ are the seven single track irregularity index structures of $G$, $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$, respectively, which represent the proportions of the standard deviations of the seven track irregularity indexes to $TQI$. The track irregularity index structure reflects the contribution of each single track irregularity index to $TQI$. The larger the value of the single track irregularity index structure, the greater the contribution of the single track irregularity index to $TQI$.

2.2. Correlation analysis method of track irregularity indexes

The one-to-one correspondence between the track irregularity index structures and $TQI$ can be established for the 200 m segments according to the mileage information. The distance correlation analysis was used to measure the correlation between track irregularity indexes in this study. The distance correlation analysis is often used to measure the similarity or dissimilarity between variables. The dissimilarity measurement is expressed by calculating the distances between variables, and the similarity measurement is expressed by calculating the Person correlation coefficient[10]. The similarity measurement was selected in this study. By inputting one-to-one corresponding track irregularity index structures $P=(p_1, p_2, \cdots, p_7)$ and $TQI$ (denoted by $P_k$), the Person correlation coefficients $r_{jk} \ (j=1, \cdots, 8; k=1, \cdots, 8)$ between the track irregularity index structures $P_i \ (i=1, 2, \cdots, 7)$ and $TQI$ (i.e., $P_k$) can be output, which can effectively measure the correlations between $P_i \ (i=1, 2, \cdots, 7)$ and $TQI$ (i.e.,
The correlation coefficients $|r_{jk}| \leq 1 \ (j = 1, \ldots, 8; \ k = 1, \ldots, 8)$ were calculated by using distance correlation analysis. And the closer the value of $|r_{jk}|$ being to 1, the higher the correlation between $p_j$ and $p_k$. The closer the value of $|r_{jk}|$ being to 0, the lower the correlation. $r_{jk} > 0$ indicates that $p_j$ and $p_k$ are positively correlated, that is, $p_j$ and $p_k$ have a tendency to increase or decrease at the same time. $r_{jk} < 0$ indicates that $p_j$ and $p_k$ are negatively correlated, that is, $p_j$ and $p_k$ have a tendency that one increases while the other decreases. $r_{jk} = 0$ indicates that there is no correlation between $p_j$ and $p_k$.

According to the results of distance correlation analysis, the correlation between the track irregularity index structures can be measured, and the correlation between the track irregularity index structures and TQI can be measured intuitively. If the single track irregularity index structure and TQI are positively correlated, it means that TQI and the proportion of the single track irregularity index to TQI have a tendency to increase (or decrease) at the same time. It can be seen from the relationship between TQI and the single track irregularity index structure that the value of the single track irregularity index increases (or decreases) correspondingly when TQI and the single track irregularity index structure increase (or decrease) at the same time, which also reflects the tendency of TQI and the single track irregularity index to increase (or decrease) at the same time, and the correlation between the single track irregularity index and TQI can be effectively reflected. Moreover, the correlation coefficient between the single track irregularity index structure and TQI is positive and its value is larger, indicating that the contribution of the single track irregularity index to TQI and the influence on TQI are greater, and the development and changes of TQI can be reflected by the single track irregularity index much better. Therefore, the track irregularity indexes that are closely related to TQI can be screened out. And the track irregularity indexes corresponding to the track irregularity index structures that are positively correlated with TQI, which can be served as a basis for guiding maintenance decisions (such as tamping operation decision-making) related to track irregularity state in specific scenarios.

3. Case study
The Shenmu-Shuozhou railway was used as the case study to analyze the correlation between track irregularity indexes.

3.1. Data preparation
The Shenmu–Shuozhou railway is a national state I level heavy-haul coal transportation special railway with double electrified lines in China, which runs from Daliuta Town in Shaanxi Province to Shuozhou City in Shanxi Province, with a 265.75 km long main line and a continuously increasing annual tonnage. As the annual tonnage arises, the workload of maintenance related to track irregularity state such as tamping presents a rising trend[11, 12].

This study used a total of 48 track inspection car historical data from September 2014 to February 2017 and November 2017 to December 2018 of the Shenmu–Shuozhou railway Hexi transportation division for verification. The Shenmu–Shuozhou railway did not use track inspection cars for inspection from March 2017 to October 2017. This study separated the up and down lines of the Hexi transportation division into 1022 consecutive segments based on the jurisdiction of the Hexi transportation division.

3.2. Correlation analysis results of track irregularity indexes
The track irregularity index structures $P = (p_1, p_2, \ldots, p_8)$ were calculated for the 1022 segments according to the mileage information and corresponding track inspection car historical data of the Shenmu–Shuozhou railway. Based on the correlation analysis method (denoted by W1) described in
section 2.2, the Person correlation coefficients between \( p_i \) \( (i=1,2,\cdots,7) \) and \( TQI \) \( (P_h) \) were calculated by using the distance correlation analysis in the SPSS statistical tool, with the track irregularity index structures \( P=(p_1, p_2, \cdots, p_7) \) and \( TQI \) \( (P_h) \) as input. Table 1 shows the correlation coefficients.

| Item | \( p_1 \) | \( p_2 \) | \( p_3 \) | \( p_4 \) | \( p_5 \) | \( p_6 \) | \( p_7 \) | \( p_8 \) |
|------|-----|-----|-----|-----|-----|-----|-----|-----|
| \( p_1 \) | 1.000 | -0.251 | -0.417 | -0.434 | -0.083 | -0.038 | -0.144 | -0.268 |
| \( p_2 \) | -0.251 | 1.000 | -0.096 | -0.108 | -0.486 | -0.509 | 0.738 | 0.017 |
| \( p_3 \) | -0.417 | -0.096 | 1.000 | 0.736 | -0.425 | -0.444 | -0.126 | 0.095 |
| \( p_4 \) | -0.434 | -0.108 | 0.736 | 1.000 | -0.414 | -0.426 | -0.136 | 0.086 |
| \( p_5 \) | -0.083 | -0.486 | -0.425 | -0.414 | 1.000 | 0.747 | -0.531 | 0.128 |
| \( p_6 \) | -0.038 | -0.509 | -0.444 | -0.426 | 0.747 | 1.000 | -0.551 | 0.140 |
| \( p_7 \) | -0.144 | 0.738 | -0.126 | -0.136 | -0.531 | -0.551 | 1.000 | -0.132 |
| \( p_8 \) | -0.268 | 0.017 | 0.095 | 0.086 | 0.128 | 0.140 | -0.132 | 1.000 |

From Table 1, it can be seen that there are positive correlations between \( TQI \) and the track irregularity index structures of \( C \), \( LLL \), \( RLL \), \( LA \), and \( RA \), with the corresponding correlation coefficients 0.017, 0.095, 0.086, 0.128, and 0.140 respectively, that is, \( TQI \) and the track irregularity index structures of \( C \), \( LLL \), \( RLL \), \( LA \), and \( RA \) have a tendency to increase (or decrease) at the same time. And the track irregularity index structures of \( LA \) and \( RA \) have the highest positive correlation with \( TQI \). The correlation coefficients between \( TQI \) and the track irregularity index structures of \( G \) and \( T \) are -0.268, -0.132, that is, the track irregularity index structures of \( G \) and \( T \) have a tendency to decrease with the increase of \( TQI \). indicating that the gauge \((G)\) and the twist \((T)\) are relatively stable and have little influence on the development and changes of \( TQI \).

In addition, from Table 1, the correlations between the seven track irregularity index structures can also be obtained. There is a significant positive correlation between the track irregularity index structures of \( LA \) and \( RA \), with the corresponding correlation coefficient 0.747, that is, the track irregularity index structures of \( LA \) and \( RA \) have a tendency to increase (or decrease) at the same time, indicating that the Shenmu–Shuozhou railway has better control over the \( LA \) and \( RA \). There is also a significant positive correlation between the track irregularity index structures of \( LLL \) and \( RLL \), with the corresponding correlation coefficient 0.736, indicating that the Shenmu–Shuozhou railway also has better control over the \( LLL \) and \( RLL \). However, the correlation coefficient between the track irregularity index structures of \( LLL \) and \( RLL \) is slightly smaller than the correlation coefficient between the track irregularity index structures of \( LA \) and \( RA \), indicating that the overall control of the \( LLL \) and \( RLL \) needs to be improved compared with the \( LA \) and \( RA \). Meanwhile, there is a significant positive correlation between the track irregularity index structures of \( C \) and \( T \), with the corresponding correlation coefficient 0.738, reflecting the close correlation between the \( C \) and \( T \), which is consistent with the views described in the literature[13] on the relationship between \( C \) and \( T \), and verifies the effectiveness of the correlation analysis method designed in this study.

Based on the above analysis, the \( C \), \( LLL \), \( RLL \), \( LA \), and \( RA \) are closely related to \( TQI \), that is, the five track irregularity indexes of \( C \), \( LLL \), \( RLL \), \( LA \), and \( RA \) have a relatively significant influence on the development and changes of \( TQI \). Therefore, the five track irregularity indexes of \( C \), \( LLL \), \( RLL \), \( LA \), and \( RA \) can be mainly considered as a basis for guiding maintenance decision making related to track irregularity state in specific scenarios.

3.3. Effectiveness verification of the correlation analysis method

In order to further verify the effectiveness of the correlation analysis method designed in this study, the total of 48 track inspection car historical data from September 2014 to February 2017 and November 2017 to December 2018 of the Shenmu-Shuozhou railway Hexi transportation division was used to compare the correlation analysis method designed in section 2.2 (i.e., W1) with the correlation analysis method (denoted by W2) which takes the standard deviations of track irregularity indexes as the input.
given in the literature[4].

Based on the correlation analysis method W2, the Person correlation coefficients between the seven track irregularity indexes (i.e., $G$, $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$) and $TQI$ were calculated by using the distance correlation analysis in the SPSS statistical tool with $TQI$ and the standard deviations of each single track irregularity index as input for the 1022 segments, according to the mileage information and corresponding track inspection car historical data of the Shenmu–Shuozhou railway. Table 2 shows the correlation coefficients.

Table 2 Correlation coefficients between track irregularity indexes and $TQI$ based on W2

| Item | $G$  | $C$  | $LLL$ | $RLL$ | $LA$  | $RA$ | $T$  |
|------|------|------|-------|-------|-------|------|------|
| $TQI$ | 0.446 | 0.845 | 0.845 | 0.844 | 0.752 | 0.748 | 0.844 |

According to the meaning of the correlation coefficient described in the literature[4], the single track irregularity index has a significant correlation with the $TQI$ when the correlation coefficient between them is greater than 0.5. From Table 1, it can be seen that there are positive correlations between $TQI$ and the track irregularity indexes of $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$. The results show that the development and changes of $TQI$ are mainly caused by the six track irregularity indexes of $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$ as using the correlation analysis method W2.

In order to illustrate the difference and effectiveness of the two correlation analysis methods W1 and W2, the track irregularity indexes closely related to $TQI$ based on W1 and W2 were selected respectively. And the indexes selected by using W1 are denoted as $TQI_1$, which include the five track irregularity indexes of $C$, $LLL$, $RLL$, $LA$, and $RA$. The indexes selected by using W2 are denoted as $TQI_2$, which include the six track irregularity indexes of $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$. The $TQI_1$ and $TQI_2$ are expressed as formulas (1) and (2) respectively.

$$TQI_1 = \sum_{n=1}^{6} \sigma_m$$  \hspace{1cm} (1)

$$TQI_2 = \sum_{n=1}^{6} \sigma_m$$  \hspace{1cm} (2)

where $\sigma_1$, $\sigma_2$, $\sigma_3$, $\sigma_4$, $\sigma_5$, and $\sigma_6$ are the standard deviations of the six track irregularity indexes of $C$, $LLL$, $RLL$, $LA$, $RA$, and $T$ for the 200 m segments, respectively.

According to the correlation analysis method designed in section 2.2, correlations between track irregularity index structures of $TQI_1$ and $TQI_2$ (i.e., the proportions of $TQI_1$ and $TQI_2$ to $TQI$) and $TQI$ can be calculated. By comparing the correlation coefficient of the track irregularity index structure of $TQI_1$ and $TQI$, with that of the track irregularity index structure of $TQI_2$ and $TQI$, the correlation analysis methods W1 and W2 can be compared. The larger the correlation coefficient, the better the corresponding correlation analysis method. Through statistical analysis and calculation, the correlation coefficient between the track irregularity index structure of $TQI_1$ and $TQI$ is 0.318, and the correlation coefficient between the track irregularity index structure of $TQI_2$ and $TQI$ is 0.268, indicating that the correlation between the track irregularity index structure of $TQI_1$ and $TQI$ is relatively more significant than that between the track irregularity index structure of $TQI_2$ and $TQI$. Although $TQI_1$ contains fewer track irregularity indexes than $TQI_2$, the correlation and influence of $TQI_1$ on $TQI$ are more significant than $TQI_2$, indicating that the development and changes of $TQI$ can be reflected by $TQI_1$ much better than $TQI_2$, which further verifies the effectiveness of the correlation analysis method (W1) designed in this study.

4. Conclusion and further research

(1) For 200 m tangent segments, using the track irregularity index structures and $TQI$ as research subjects, the correlation analysis method of track irregularity indexes was designed.

(2) The track inspection car historical data of the Shenmu–Shuozhou railway Hexi transportation division was used to verify the effectiveness of the correlation analysis method designed in this study. The results show that there are positive correlations between $TQI$ and the five track irregularity indexes
of $C$, $LLL$, $RLL$, $LA$, and $RA$. The correlation analysis method designed in this study was compared with the correlation analysis method based on the standard deviations of track irregularity indexes. The comparison results show that the correlation analysis method designed in this study can be used to analyze the correlations between the track irregularity indexes more effectively.

(3) In the future, the author will collect more comprehensive data, and further study the formulation of maintenance decisions related to track irregularity state based on the results of the correlation analysis of track irregularity indexes.

References

[1] Wang, N., Wang, F. (2015) Prediction for track irregularity based on gray interval prediction model. Railway Computer Application, 2015(01):1-3.
[2] Li, S., Liu, R., Wang, F. (2018) Prediction for track irregularity index structure based on grey compositional data model. Journal of Railway Science and Engineering, 15:1937-1942.
[3] Li, G. (2016) Research on the track quality index and its application in high speed railway dynamic acceptance. Journal of Railway Engineering Society, 33:45-50.
[4] Yang, F., Wang, T., Yang, J., Yin, F. (2016) Research on influence of mud-pumping and renovation of slab ballastless track subgrade on track irregularity. Railway Engineering, 2016(11):129-132.
[5] Zuo, Y., Xiang, J. (2006) Correlation analysis of track irregularities of Zhengzhou-Wuhan railway. Journal of Railway Science and Engineering, 2006(01):46-49.
[6] Jin, S., Zeng, Z. (2009) Analysis of bogl slab track irregularity of Beijing-Tianjin intercity high speed railway. Journal of Railway Science and Engineering, 2009(01):48-52.
[7] Lasisi, A., Attoh-okine, N. (2018) Principal components analysis and track quality index: A machine learning approach. Transportation Research Part C: Emerging Technologies, 91:230-248.
[8] Zhang, N., Liu, R., Wang, F., Li, S. (2018) Prediction for track vertical profile irregularity index structure of Shennu-Shuozhou railway. In: 2018 3rd IEEE International Conference on Intelligent Transportation Engineering. Singapore. pp: 80-84.
[9] Liu, R., Xu, P., Wang, F. (2010) Research on a short-range prediction model for track irregularity over small track lengths. Journal of Transportation Engineering, 136:1085-1091.
[10] Yan, Y., Jia, X. (2016) Synthetical analysis of factors affecting aviation maintenance errors based on distance correlation analysis method. Aeronautical Computing Technique, 46:84-87.
[11] Li, S., Liu, R., Wang, F. (2019) Correlation analysis between track irregularity indexes and quality of ballast tamping operation. Railway Engineering, 59:114-118.
[12] Qin, X., Bai, L. (2015) Comprehensive assessment model on track quality condition of Shennu-Shuozhou railway and study on its algorithm. Railway Engineering, 2015(12):130-133.
[13] Gao, L. (2014) Track engineering. Chongqing University Press, Chongqing.