The Deviation Characteristics of High-speed Train Running Time in Section and its Effect on Transport Capacity

Zhiming Yuan, Qi Zhang, Tao Wang and Feng Chen

Signal and Communication Research Institute, China Academy of Railway Sciences, Beijing, China, 10081
13810696163@139.com

Abstract

Starting from the phenomenon that a large number of trains' actual arrival & departure time deviate from specified time in train working diagram in the database of high speed railway dispatching system, the paper analyzes the defects of current train running time calculation method. Based on data mining theory, a section running time deviation calculation model for high speed railway train is proposed. Then the properties of train section running time deviation are studied in three different cases (non-stop both at the start and the end of section, stop at the end of section, and departure at the start of section) with the data from a Chinese railway bureau dispatching system, and its further influence on carrying capacity is analyzed. The results show that three deviation cases all fit Gauss distribution and the accumulative deviation time can increase the transport capacity for more than 1 million passengers per year.

Keywords: High Speed Railway; Train Running Time in Section (TRTS); Deviation; Data Mining; Gaussian distribution

1. Introduction

The high speed railway is a system, that are reconstructed (line linearization, gauge standardization) from existing lines and maximum operation speed reach 200km/h or more, or new lines with maximum operation speed achieve 250km/h or more. As a new mode of transportation, high speed railway has many advantages such as rapidity and punctuality, and has gained rapid development and widely attention from the world since being put into operation. After 10 years' development, the mileage of Chinese high speed railways in operation has reached more than 16 thousand kilometers.

Train running time in section (TRTS) is not only the standard of operation time between two adjacent stations or block posts, but also an important basis for drawing train working diagram and calculating railway line transport capacity. The influence factors for TRTS include the maximum operation speed of train, the horizontal and vertical designs of the lines, the quality of the train, *etc*. According to different situations on train operation process, TRTS can be divided into pure running time, additional time for stop, and additional time for departure. Pure running time is the time interval between two adjacent stations or block post when trains are under normal operation speed. Additional time for stop is the increase time between stop at a station and pass through it. Additional time for departure is the increase time between departure from a station and pass through it. Therefore, TRTS can be calculated in formula (1).

$$t_{sec} = t_{ps} + at_{start} + bt_{stop} \qquad (1)$$

Where, t_{sec} is TRTS (min), t_{start} is the additional time for start (min), t_{stop} is the additional time for departure (min). And a, b are parameters which can be determined in Formula (2) and (3).

$$a = \begin{cases} 1, when\,train\,stops\,at\,the\,start\,station\,or\,block\ post\ of\ the\ section \\ 0,\ other\ conditions \end{cases} \tag{2}$$

$$b = \begin{cases} 1, when\,train\,stops\,at\,the\,end\,station\,or\,block\ post\ of\ the\ section \\ 0, other\ conditions \end{cases} \tag{3}$$

According to the value combinations of a & b, four situations (as shown in Figure 1) can be get.

$\{0,0\}$: The train does not stop neither at the start nor the end station or block post of the section.

$\{1,0\}$: The train stops at the start station or block post and passes the end section or block post of the section.

$\{0,1\}$: The train passes the start station or block post and stops at the end station or block post of the section.

$\{1,1\}$: The train stops both at the start and the end station or block post of the section.

For high speed railway's rapidity requirement, the last situation that train stops both at the start and the end station or block post (the same function to station in train dispatching work) of the section is rare enough to neglect in actual operation, so in this paper, this situation is not in conference.
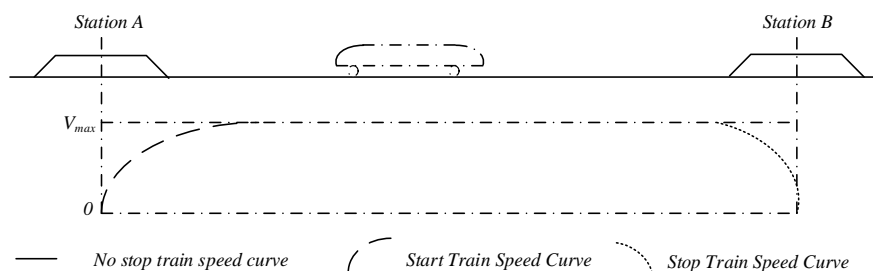
Figure 1. Schematic Diagram of High Speed Train Running in Section

2. Background and Related Work

2.1. Background

Current TRTS calculation method of high speed railway is based on the test method. The test train runs in different mode repeatedly before the high speed railway line put into operation, and the test data are collected. After data analysis, the pure running time, additional time for stop, and additional time for departure are obtained. Then the train working diagram can be drawn based on the results above. This method for calculating TRTS is on the conditions of test running, in which the train type is lack of variety and the running condition is ideal. Due to the lack of actual operation practice, especially the situation of a large number of trains in operation, after the line was put into service, a lot of trains' arriving & departing time deviate from the time specified in the train working diagram are found in the database of dispatching system, which means that the real TRTS doesn't match the time in the train working diagram.

According to formula (1), influenced by climatic and other conditions, the train pure running time t_{ps}, additional time for stop t_{stop}, and additional time for departure t_{start} may diverge from the train operation diagram. Quantitative analysis of this deviation by mathematical method has positive significance for improving the dispatch accuracy and transport efficiency of high speed railway.

2.2. Related Work

In recent years, with the large-scale construction and operation of high speed railway (especially in China), a boom of researching has been set off. On the field of high speed trains operation, Xuesong Feng (2011) calculated the energy consumption and transport efficiency of different speed target values from the point of train traction, and recommended its economic speed. Rémy Chevrier *et al*. (2013) established a bi-objective optimization model, which considers both the TRTS and energy consumption. Then the model is solved on evolutionary algorithms, and provides a recommend algorithm for high speed railway train diagrams. Gerben M. Scheepmaker *et al*. (2015) studied the relationship between train control efficiency and the reserving time in train operation, then built the EZR model which based on optimum control theory, and applied it to local railways in the Netherlands. José Luis Espinosa-Aranda *et al*. (2015) established a high speed train scheduling model based on user choice.

About the researches for TRTS, Bangmo He *et al*. (1981) analyzed the influence factors of TRTS under the condition of the existing trains with types of both passenger and freight, and according to the actual data (including additional time for departure& stop) of TRTS which collected by train monitors, calculates the deviations of TRTS based on the statistics method. Basing on the dynamic performance of the TRTS, Weixiong Cha *et al*. (2001) used interval estimation of random variable under the certain level of significance in mathematics statistics, and analyzed the time standard of the TRTS and effects of the cushion time of train operation diagram. Thus, this work would offer effective reference for train scheduling and train operation adjustment. Xingchen Zhang *et al*. (1996) analyzed the deviation of the TRTS presented as slanting distribution by the method combining the theoretical analysis with the measure and calculation in practice, and described it with left slant β-function distribution. This provided the convenience for the simulation of train operations and for other train operating calculations. According to analyze the observed data of the running test before the opening of the high-speed railway, Xi Ling (2012) described the TRTS with normal distribution. Depending on different types of train, Jun Wang *et al*.(2011) used the statistical analysis method to categorized the recorded data`s rules of concentration, dispersion and distribution of monitoring device of the train running, and fitted the curve regression model to revise the parameters of the heavy haul railway running map.

Above all, the recent research mainly focused on the operation efficiency and energy consumption of high speed railway. The studies on TRTS were relatively rare and old. The data which collected by train monitors, have defects on sample size and randomness, when the data got from the testing experiment lack the actual running test. At present, there is no research of TRTS of high speed railway train based on the actual train running data.

3. Calculation Model

In theory, the train should be strictly obey the section running time in working diagram, but due to the impact of various factors, actual section running time deviate from the theory. In high speed railway system, the following train track the leading train, under normal conditions, the block sections can ensure sufficient time and space interval and trains will not interfere with each other. Although the train may be late in this section, it should run according to the train working diagram in the next section. So, from the statistical view, it can be considered that the trains' running time in section is independent with each other.

Based on the data mining theory (Jiawei Han *et al*, 2012), after some pre-processing of data integration, data cleaning and data discriminating (such as removing the abnormal data caused by temporary stop) of the train section running data, according to the

characteristics of the train section running data, the paper make the calculation model as follows:

(1) Train section running data structure construction

In China, the national railway lines are divided into 18 parts and managed by 18 railway bureaus. But for the complex investors, the high speed railway line companies are independent to the 18 railway bureaus. However it is very hard for them to set up a whole dispatching department, they usually adopt entrust operation mode. The high speed railway lines are divided into different parts according to the area of jurisdiction of 18 railway bureaus, and are entrusted to them. So, a dispatching department of a railway bureau always manages many parts of high speed railway lines.

In the database of dispatching system, there are many trains running on many parts of high speed railway at the same time. At least one recording with information of arriving or passing will generate when a train is getting through a station. Although it looks very complex, it can be divided into three keywords: Train Number, Date and Stop Station. Every train number has several stop stations according to time series, and most train numbers run every day. Based on the unique identity of each train, according to the sequence of operation and the characteristics of date repeatability, a three-level-list data structure consisting of train ID, date, stopping sequence is constructed.(as showed in Figure 2).
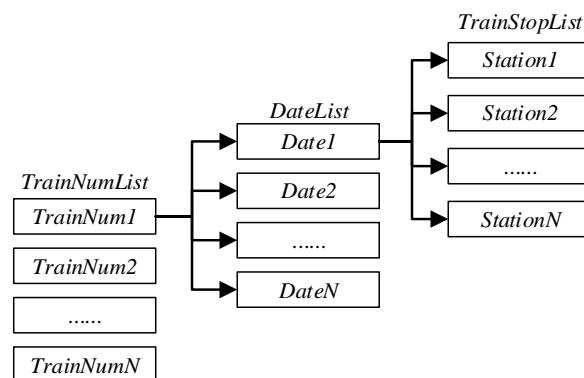
Figure 2. Data Structure of Train Running

(2) Train's trip list extraction

In dispatching system's database, the data are recorded according to the arrival and departure time of trains in stations. It should be dealt to get every train's every day's trip list in chronological order. The algorithm pseudo-codes are as followed.

For i from 1 to TrainRecordTable.Count
bSame ← False
For j from 1 to TrainNumList.Count
If TrainRecordTable(i).TrainNum =
TrainNumList.TrainNum Then
bSame ← True
Exit For
End If
End for
If (bSame = False) Then
TrainNumList.add (TrainRecordTable(i).TrainNum)
End If
End for

(3) Repeated records elimination

In the dispatching system, a dispatching desk is responsible for the operation of a section in the whole line, but for the coordination of different dispatching desks, the visual section will be extended from both end of the basic section. Although no adjustment can be done for the operation of the extending part, the running state of trains can be checked and recorded. Therefore, in the dispatch system database, repeated records of the trains will generate, and these data must be removed. The algorithm pseudo-codes are as followed.

For i from 1 to TrainNumList.Count
For j from StartDate to EndDate
strDate =format("yyyy-mm-dd",j)
*Sql="select * from [Sheet1$] where ArriTrainNun='" & TrainNumList(i).TrainNum &*
"'" & "and ArriDate='" & strDate & "'" & "order by ArriTime"
Sheets("Sheet2").[A1].CopyFromRecordset Conn.Execute(Sql)
iTrainStopNum = Sheet2.UsedRange.Rows.Count
If (iTrainStopNum > 1) Then
For k = 1 To iTrainStopNum - 1
If Sheet(2).Cells(k, 8).Value= Sheet2.Cells(k + 1, 8).Value Then
Worksheets(2).Range(Worksheets(2).Rows(k + 1), Worksheets(2).Rows(k + 1)).Delete
iTrainStopNum = iTrainStopNum - 1
End If
End for
iTrainStopNum = Sheet2.UsedRange.Rows.Count
Sheet2.UsedRange.Copy
ActiveSheet.Paste Destination:=Worksheets("Sheet3").Range("a" & iAddupNum)
Worksheets(2).Range(Worksheets(2).Rows(1),
Worksheets(2).Rows(iTrainStopNum)).Delete
iAddupNum = iAddupNum + iTrainStopNum
*End If*
End for
End for
End

(4) Difference of TRTS between actual train section running time and train working diagram Calculation

$$t_{sec}^{i'} = (t_{arri}^{i+1} - t_{dep}^i) - (t_{TDarri}^{i+1} - t_{TDdep}^i) \tag{4}$$

Where, $t_{sec}^{i'}$ is the deviation of running time in section (min)($i = 1, 2, ..., N - 1$); $t_{arri}^{i+1}, t_{TDarri}^{i+1}$ represent the actual arrival time and the set one in working diagram in station $i + 1$, respectively; t_{dep}^i, t_{TDdep}^i represent the actual departure time and the set one in working diagram in station , respectively.

(5) Statistical analysis on the deviation value and distribution of TRTS

Trains have three different situations on a section: non-stop both at the start and the end of a section, stop at the end of a section, and departure at the start of a section. The deviations for three situations are respectively estimated from the mean value, variance and standard difference, and the distribution curves are fitted and parameters are calibrated. The whole process of data analysis is shown in Figure 3.
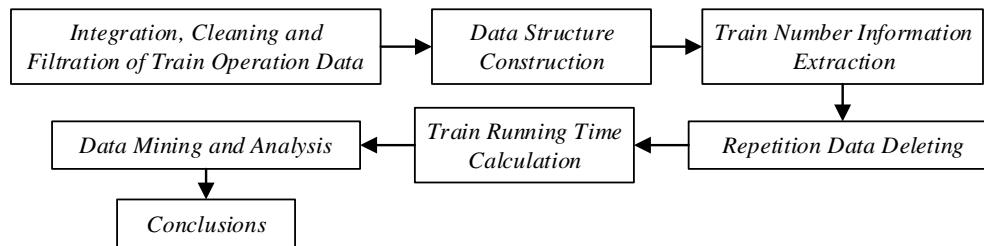
Figure 3. Deviation Analysis Process of TRTS

4. Data Analysis

The train operation data of high speed railway from a Chinese railway bureau (named A) are taken as sample, and comparison study are made for the deviations between recorded actual time and set time in train working diagram in three situations (including train non-stop both at the start and the end of a section, stop at the end of a section, and depart at the start of a section).The statistical results are shown in Table 1.

Table 1. Deviations of TRTS in Three Situations

Deviation(min)	$\leq$-4	-3	-2	-1	0	1	2	3	$\leq$4
Deviation percentage of non-stop (%)	0.008	0.355	6.725	33.416	43.406	14.742	1.312	0.032	0.003
Deviation percentage of end-stop (%)	0.264	2.251	11.373	29.553	32.988	17.630	5.190	0.698	0.053
Deviation percentage of start-depart (%)	0.494	3.440	12.948	27.169	31.861	19.302	6.332	1.083	0.158

As can be seen from Table 1, there is a big deviation in the interval running time on the whole, punctuality rate is low(less than 50%), which means: more than 50% of the trains didn't obey the section running time in train working diagram. With 0 as the median, taken the left and right deviation percentage into consideration, left deviation percentages exceed the right ones in all three situations, that means in the whole the actual TRTS are shorter than the ones planned in the train working diagram. During the three situations, non-stop situation has the lowest deviation, because lack of two factor t_{start} & t_{stop} make the fluctuation of TRTS to the gentlest. At the meantime, the additional time for stop is less than for departure, so the deviation in end-stop situation is smaller than start-departure situation.

In the case of non-stop, several models are used to fit the data in Table 1, and the indicators of fitting results are showed in Table 2.

Table 2. Indicators of Different Fitting Models

Model	SSE	R-square	Adjusted-square	RMSE
Fourier	2.160e-2	0.893	0.828	6.573e-2
Gaussian	1.287e-6	1.000	1.000	4.632e-4
Smoothing Spline	8.211e-3	0.959	0.888	0.053

As showed in Table 2, all the test values of Gaussian fitting model perform better than those of better Gaussian model fit better than Fourier and Smoothing Spline models. The fitting Gaussian model is showed in Formula (5).

$$f(x_{ns}^t) = 0.462\exp(-((x_{ns}^t + 0.305)/1.221)\char`^2) \qquad (5)$$

Where, x_{ns}^t stands for the running time deviation in non-stop situation (min), $f(x_{ns}^t)$ stands for the distribution function of the running time in both ends non-stop.

The initial hypothesis, that the deviation of TRTS in Non-stop situation is fitting the Gaussian distribution with parameters of $\mu_{ns} = -0.305$, $\sigma_{ns} = 0.864$ is proposed. Then the TRTS deviation data are divided into 9 sections and degrees of freedom is 8 and confidence level is set to be 95%. Chi-square testing is used to testify the initial hypothesis and results are as follows.

$$\chi_{ns}^2 = \sum_{i=1}^{9} \frac{(A_{ns}^i - N_{ns}p_{ns}^i)^2}{N_{ns}p_{ns}^i} = 13.467 < 15.51 \qquad (6)$$

Where, χ_{ns}^2 stands for the Chi-square test value, A_{ns}^i is the observed sample number in the internal i, p_{ns}^i is the probability of the internal i, N_{ns} is the total sample size.

Based on the rules of Chi-square test, observed value χ_{ns}^2 (13.467) is smaller than 15.51(with 8 freedom degree and 95% confidence level), so the original hypothesis is accepted.

With the same method, deviation distributions of TRTS in both end-stop and start-depart situations are fitted based on Gaussian model. The distribution functions are showed in Formula (7) and (8).

$$g(x_{stop}^t) = 0.340\exp(-((x_{stop}^t + 0.306)/1.660)\char`^2) \qquad (7)$$

$$h(x_{start}^t) = 0.315\exp(-((x_s^t + 0.304)/1.791)\char`^2) \qquad (8)$$

Where, x_{stop}^t, x_{start}^t stand for the deviations of TRTS in end-stop and start-depart situations respectively (min), $g(x_{stop}^t)$, $h(x_{start}^t)$ stand for the deviation distribution functions of two situations respectively.

Thereby, hypothesize that deviations of TRTS for end-stop and start-depart situations fit the Gauss distribution respectively with $\mu_{stop} = -0.306$, $\sigma_{stop} = 1.174$ and $\mu_{start} = -0.304$, $\sigma_{start} = 1.267$. Then Chi-square tests are made to test the hypothesis and testing value are calculated in Formula (9) and (10).

$$\chi_{stop}^2 = \sum_{i=1}^{9} \frac{(A_{stop}^i - N_{stop}p_{stop}^i)^2}{N_{stop}p_{stop}^i} = 12.503 < 15.51 \qquad (9)$$

$$\chi_{start}^2 = \sum_{i=1}^{9} \frac{(A_{start}^i - N_{start}p_{start}^i)^2}{N_{start}p_{start}^i} = 14.994 < 15.51 \qquad (10)$$

Where, χ_{stop}^2 and χ_{start}^2 respectively stand for Chi-square test values for end-stop and start-depart situations, A_{stop}^i and A_{start}^i are the observed sample numbers in the internal i, p_{stop}^i and p_{start}^i are the probability of those internals, N_{stop} and N_{start} represent the total sample sizes.

Based on the rules of Chi-square test, testing value χ_{stop}^2 (12.503), χ_{start}^2 are smaller than 15.51(with 8 freedom degree and 95% confidence level), so the original hypotheses are accepted.

Some findings can be found from the above three fitting Gauss functions. Firstly, all three functions have similar parameter μ of 0.305 value with differences less than ± 0.001. That means the average values of TRTS's deviations have little difference from each other in three situations. However, the parameter σ changes a lot for the

largest gap up to 0.403min, which means the standard errors in three situations differ greatly. The non-stop situation has the lowest standard error (0.864min) for TRTS only have one component as pure running time. And the end-stop situation has lower standard error (0.864min) than start-depart for the additional time for stop is smaller than that for departure. For the departure additional time is larger than the stop additional time in actual train operation, this result matches the reality in railway operation very well.

From the result above, it can be seen that the TRTS need to be adjusted, which will improve the accuracy of dispatching system, and increase the service level of high speed railway.

5. Effect on Transport Capacity

Carrying capacity for railway line is the maximum daily(24h) number of trains or train-pairs at which trains can reasonably be expected to cross or be serviced in various fix facilities on a railway line under prevailing locomotive type and operation condition (Yang Hao, 2015;Hu Siji, 2009). As a basis of train running operation, carrying capacity calculation has two important parameters, which are TRTS and trains tracking interval. Carrying capacity calculation in paralleled train working diagram for high speed railway is showed in Formula (11).

$$N_{hsr} = \frac{1440 - T_{main} - \sum \overline{t_{rts}}}{I_{tracking}} \tag{11}$$

Where, N_{hsr} stands for the carrying capacity in paralleled train diagram for high speed railway (pairs), T_{main} stands for maintenance time (min), $\overline{t_{rts}}$ stands for TRTS (min), $I_{tracking}$ stands for trains tracking interval (min), which is the sum of minimum interval and buffering time of working diagram and has a reference values of 5min based on the high-speed railway operation in the world.

The transport capacity has close relationship with carrying capacity. For high speed railway, transport capacity can be calculated by carrying capacity. Shows as formula (12).

$$C_{hsr} = N_{hsr} \times C_{train} \times r_{train} \times 2 \tag{12}$$

Where, C_{hsr} stands for the transport capacity of high speed railway (person time), C_{train} stands for train's capacity, r_{train} stands for average attendance rate of train (%).

One high-speed arterial line (named B) in Railway Bureau A is taken as an example. From the working facts it is assumed that a train formed by 16 vehicles, average booking rate is 75%, the length of line is 957km with 32 sections, and trains stop 4 times in the line. The cumulative deviation time T_{devia}^{total} for the line B is calculation in Formula (13).The explanation of variables can be found in Formula (5), (7) & (8).

$$T_{devia}^{total} = -[5 \times (\mu_{start} + \mu_{stop}) + 22 \times \mu_{ns}] = -9.76 \tag{13}$$

Furthermore, the pairs of trains which can be added into the line using the deviation time is calculated in Formula (14).

$$N_{train} = \frac{1440 - T_{main} - (\sum \overline{t_{rts}} + T_{devia}^{total})}{I_{tracking}}$$

$$= \frac{1440 - T_{main} - \sum \overline{t_{rts}}}{I_{tracking}} - \frac{T_{devia}^{total}}{I_{tracking}} = C_{hsr} - \frac{-9.76}{5} = C_{hsr} + 2 \qquad \text{(pairs)} \tag{14}$$

Finally, the increase on passenger capacity per year is showed in Formula (15).

$$C_{pessager} = 2 \times 1000 \times 0.75 \times 2 \times 365 = 1.095 \tag{15}$$

In order to achieve security requirements of high speed railway, daily overhauling of facilities like track, signal and catenary is necessary. For this purpose, vertical skylights time for maintenance are designed for high speed railway trains all around the world. With skylights, one daytime in train working diagram is divided into two parts: maintenance time and service time. Because the maintenance time cut the service time into two segments, vertical skylights will reduce the long-distance capability of high-speed railway, and longer the route, greater the loss (as shown in Figure 4). Therefore inadequate long-distance capacity is China's main difficulties in high-speed railway. Great economic benefit can be got if the cumulative deviation time of TRTS will be used to run long-distance trains.
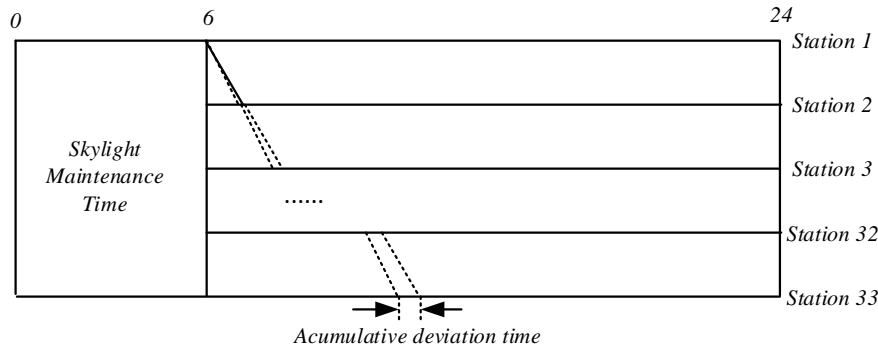
Figure 4. Schematic Diagram of the Effect of Deviation Analysis of TRTS

6. Conclusions

The calculation of Train Running Time in the Section (TRTS) is the basis for train working diagram and transport capacity. Based on data mining theory, a TRTS deviation calculating model is built according to the data recording structure in high-speed railway dispatching system. The running data of high speed railway are taken from one of Chinese railway bureaus. Under three situations (including train non-stop both at the start and the end of a section, stop at the end of a section, and depart at the start of a section), studies are made on the distribution of TRTS's deviations between observed actual data and set data in train working diagram. The result shows that the deviations of TRTS in all three situations fit Gauss distribution with Chi-square test well. The standard errors of three situations increase orderly, which is reasonable in reality. On this basis, the effects of TRTS deviation on high-speed railway carrying capacity are analyzed. To quantize this effect, data from an arterial high-speed railway line have taken for example. Modification on the deviation of TRTS can improve transport capacity for 1 million passengers per year, which have significance for the high-speed railway development.

Acknowledgements

References

[1]   X. Feng, "Optimization of target speeds of high-speed railway trains for traction energy saving and transport efficiency improvement", Energy Policy, vol. 39, **(2011)**, pp. 7658–7665.

[2]   R. Chevrier, P. Pellegrini and J. Rodriguez, "Energy saving in railway timetabling: A bi-objective evolutionary approach for computing alternative running times", Transportation Research Part C, vol. 37, **(2013)**, pp. 20-41.

[3]   G. M. Scheepmaker and R. M.P. Goverde, "The interplay between energy-efficient train control and scheduled running time supplements", Journal of Rail Transport Planning & Management, **(2015)**, pp. 1-15

[4]   J.L. Espinosa-Arand, R. García-Ródenas, M. C. Ramírez-Flores, M. L. López-García and E. Angulo, European Journal of Operational Research, vol. 246, **(2015)**, pp. 772–786.

[5]   B. He, Y. Song and Z. Xu, "Deviation Analysis of Train Running Time in section", Railway Transport and Economy, no. 1, **(1981)**, pp. 57-63

[6]   W.-X. Zha and G.-L. Xiong, "Study on the Dynamic Performance of the Train Running Time in the Section", Systems Engineering, vol. 19, no. 1, **(2011)**, pp. 47-51

[7]   X. Zhang and A. Hu, "Analysis of β Function Distribution for Deviation of Train Running Time in Section", Journal of the China Railway Society, vol. 18, no. 3, **(1996)**, pp. 1-6

[8]L. Xi and W. Yao, "Research on Statistic Analysis Method of Operation Time Test on High-speed Railway Section", Railway Transport and Economy, vol. 34, no.9, **(2012)**, pp. 85-89.

[9]   J. Wang, X. Zhang, Y. Zhang and J. Chen, "Method of the Measurement and Correction of Train Diagram Parameters Based on Monitoring Data Mining".

[10] J. Han, M. Kamber and J. Pei, "Data Mining Concepts and Techniques (Third Edition)", M. Fan, X. Meng, China Machine Press, **(2012)**.

[11] Y. Hao, "Railway Transportation Organization (Third Edition)", China Railway Publishing House. Beijing, **(2015)**.

[12] S. Hu, "Railway Traffic Organization (Second Edition)", China Railway Publishing House. Beijing, **(2009)**.