

Research Article

Study of Track Irregularity Time Series Calibration and Variation Pattern at Unit Section

Chaolong Jia,¹ Lili Wei,² Hanning Wang,³ and Jiulin Yang⁴

¹ School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

² Chongqing Public Security Bureau, Chongqing 401147, China

³ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

⁴ China National Tendering Center of Mach. & Elec. Equipment, Beijing 100142, China

Correspondence should be addressed to Chaolong Jia; jiachaolong@bjtu.edu.cn

Received 15 July 2014; Accepted 5 October 2014; Published 4 November 2014

Academic Editor: Xiaobei Jiang

Copyright © 2014 Chaolong Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Focusing on problems existing in track irregularity time series data quality, this paper first presents abnormal data identification, data offset correction algorithm, local outlier data identification, and noise cancellation algorithms. And then proposes track irregularity time series decomposition and reconstruction through the wavelet decomposition and reconstruction approach. Finally, the patterns and features of track irregularity standard deviation data sequence in unit sections are studied, and the changing trend of track irregularity time series is discovered and described.

1. Introduction

Time series [1–3] is a statistical method of econometrics. Time series studies the changes showed by observation values of a certain variable in the system in chronological order during a given period and tries to find out the characteristics, future trends and laws over time and the laws are often the consolidated results of impacts by a variety of other factors. Time series does not study the interdependence causality between things, and the study is based on the assumption that some of the information which comes from the historical data can be used to explain the current situation and to predict the future of time series. This reflects an inertia of the development of things with time.

Generally, there are four changing features such as trend [4], periodicity [5], randomness [6], and synthesis [7] in time series and three types of time series prediction including point prediction which offers prediction values, range prediction which offers prediction value within one interval, and density prediction which offers the complete probability distribution of the prediction value.

Time series forecasting is an active study area, and there are a lot of literature on it. In the researching methods of time series forecasting, collection and analysis of historical

observations are used to determine the model and to capture the generating process of underlying data, and then the model is used to make prediction. This predictive method is very important in many fields and is widely used in the business, economic [8, 9], industrial [10, 11], engineering [12–14], science [15–21], and other communities. Scholars around the world have been committed to the development and improvement of time series forecasting model in the past few decades.

In the study of time series data, there are some prerequisites for time series modeling in order to make sure the results are accurate and the model is effective. These prerequisites include studying the characteristics of the object data, selecting representative data for study, controlling data quality by means of data correction, analyzing data composition in-depth internally, and discovering implied rules and characteristics in data. All these need to be further studied.

Railway network undertake the important task of passenger and freight traffic; its performance will play an important role for rail transport [22–25]. The railway track states directly determine the safety of train operation. The regularity of the track is not only an important indicator of the track state but also the basis for evaluation of train running quality. Where there is track irregularity, speed limit should be paid attention

to; otherwise, in some extreme occasions, overturning might occur. As a result, it is urgent for railway departments to study the law of track irregularity changes so as to master trends of track state changes and to take prevention measures [26, 27]. Various tracks state that inspection data is the most important resource and the accuracy of the data can not only truly reflect the state of the track but also is the basis for modeling and forecasting. Based on the importance of data, this paper identifies abnormal data and calibrates offset data and segment data in order to study track irregularity change trends. In this context, this paper analyses track irregularity data, explores the underlined rules of track irregularity change, predicts future trends, and, ultimately, provides data and models support of track state changes to relevant railway departments, to ensure railway transportation safety.

In this study, track irregularity data is provided by State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University.

2. Abnormal Data and Offset Data

The track irregularity inspection data cannot be found through long-term, continuous monitoring at a certain fixed monitoring site but through the inspection data of various subsections. A single detected data is not time series data, but repeated inspection data is. Meanwhile, each inspection point corresponding to the inspection data will have some offset, which is mainly caused by the inspection device. Since the inspection is dynamic, mileage offset exists in inspection data, so it requires manual correction for every 10 km during the operation of track inspection car. However, there are errors in manual correction, and, according to on-site work experience, this error range is essentially within 50 m, which is still a great error.

Track geometric irregularity data on the timeline at each measuring point should be a time-series data, but in real inspection process, the actual mileage and the mileage measured by track inspection car does not remain the same, and in some occasions the previous measuring points do not correspond to each other, so the result will be as follows: time series data should be constituted by the track irregularity data at the same location but at different time; however, in reality it is constituted by track irregularity data at different time and at different location.

Specifically, mileage offset can be divided into two cases. In the first case, in a single inspection, inspection data and mileage measuring point position correspond to each other accurately, but there are differences between the corresponding measuring points of each time inspection data. In the second case, position of the measuring point corresponding to the inspection data does not correspond with the actual distance, and the actual data is the data corresponding to a position before or after the measuring point. In practice, it is difficult to distinguish these two cases and they can coexist.

3. Identify Abnormal Data

Data deviated from the normal value is commonly referred as abnormal data or outliers. In track state inspection process,

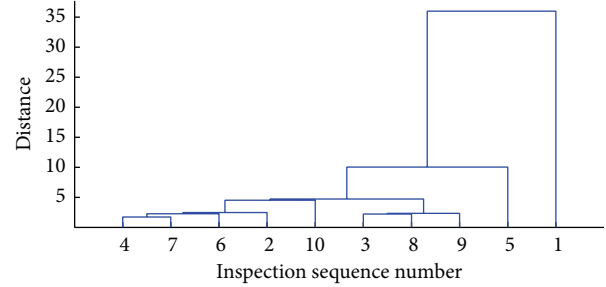


FIGURE 1: Pedigree chart of clustering.

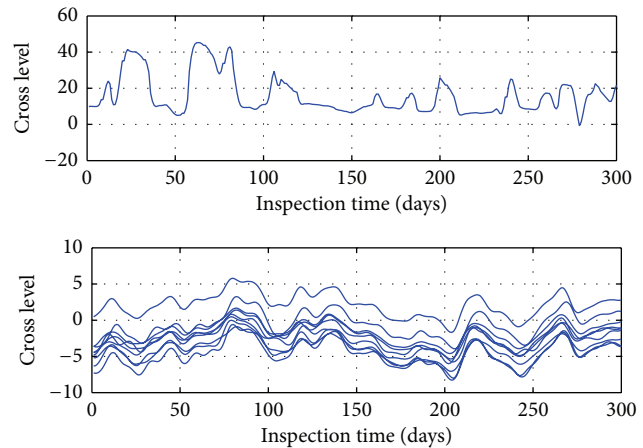


FIGURE 2: Results of gauge irregularity cluster.

abnormal inspection data values easily occur due to inspection equipment, locomotives working conditions, and other factors. The anomalies of track irregularity data include two types: overall anomalies and local anomalies.

3.1. Overall Abnormal. The track inspection data between October 22, 2007 to June 11, 2008, Beijing-Kowloon line, K500+000–K500+100 unit section is selected as the study object. Outlier curve and normal curve are separated through cluster analysis, and two cluster centers clustering results can be obtained, and outliers track state is detected.

Pedigree chart of previous gauge irregularity inspection waveform data by cluster analysis is shown in Figure 1.

Gauge irregularity cluster results are shown in Figure 2. The following chart is normal data, and the previous chart shows the abnormal value.

3.2. Local Anomaly. There will be local phenomenon outliers in track inspection data. In this case, the abnormal data often accounts for a small portion of all the data, but there is a larger difference in amplitude than other normal data.

In recognition of abnormal data, this paper proposes the ratio of difference between track irregularity values at adjacent measuring points to difference between interval lengths at adjacent measuring points (usually roughly 0.25 m). It is defined as an abnormal degree in this paper, and abnormal

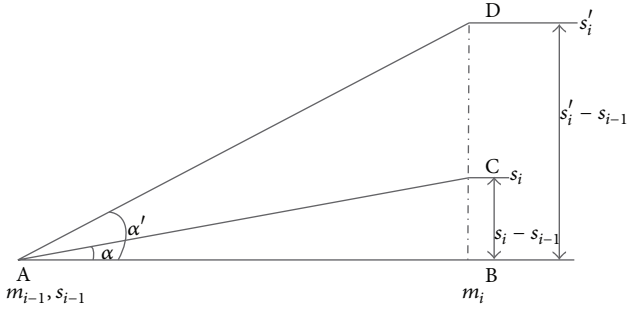


FIGURE 3: Schematic diagram of track irregularity abnormal state change.

degree is used to determine and identify outliers' values. The abnormal degree formula is shown as follows:

$$d_i = \frac{s_i - s_{i-1}}{m_i - m_{i-1}}. \quad (1)$$

In the formula, d_i is abnormal degree, s_i is track irregularity value at measured point i , s_{i-1} is track irregularity value at measured point $i - 1$, m_i is mileage values of measuring point i , and m_{i-1} is mileage values of measuring point $i - 1$.

The geometric form of formula (1) is shown in Figure 3. In the formula, abnormality degree is the tangent ($tg\alpha$) in Figure 3. The judgment of track irregularity outlier's recognition is shown in the following.

(1) *Normal Value*. When $tg\alpha < k$, it indicates that the state of track irregularity amplitude variations is among the normal range of variation, and in this case, some injuries such as broken rail will not appear.

(2) *Outlier Value*. When $tg\alpha \geq k$, it indicates that the track irregularity state change has exceeded the normal variation amplitude range, and in this case, the track may have serious injuries, such as broken rail.

In Figure 3, $tg\alpha' = k$ is the turning point of state exception changes.

The inspection data of Beijing-Kowloon line in 459 km-460 km mileage section in February 2009 is selected for the study, and the presence of local outliers can be found. The abnormal value of inspection data is shown in Figure 4.

By studying a large number of data, it can be found that, under normal circumstances, most distribution of d_i is $[-0.02, 0.02]$; that is, the range can be set to $[-0.02, 0.02]$.

The reasons of the occurrence of abnormal data can be grouped into two categories after analysis: inspection equipment problems (when track inspection car is in abnormal situation, abnormal data will occur); the difference of inspection objects, such as data, when track inspection car through the main line is different from that through turnout.

Abnormal data causes mutations and it must be eliminated. Restoration and correction to abnormal data can improve the effectiveness of the data in analysis, except for the interference of outliers, and then accurate characteristics of track state changing trends can be discovered.

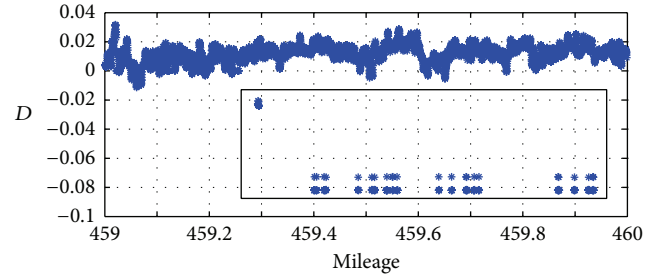


FIGURE 4: Local outlier values of inspection data in February 23, 2009.

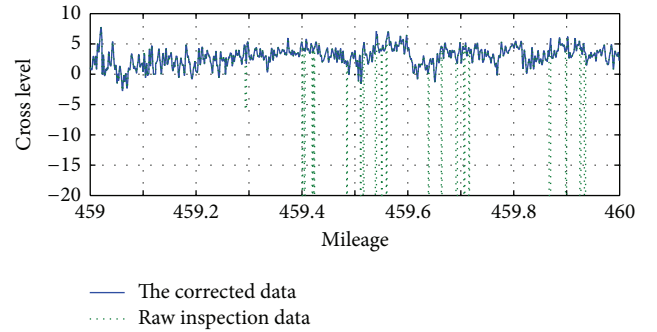


FIGURE 5: Comparison between revised local outliers data and original value in February 23, 2009.

4. Abnormal Data Treatment

In case of outliers, there are two measures for treatment: amendment and abandoned.

Deprecated case refers to the situation when data is covered by outliers in large area and the actual value is difficult to be restored and thus has to be abandoned in study. Usually, abandonment is seldom to be seen, and only exception data occurs sometimes. As long as the abnormal data is corrected, it can still be used for research.

As the track is continuous physically and spatially, track geometry irregularity changes along mileage direction show continuous features. According to this continuity character, it can be corrected by linear interpolation abnormal data. After correction of outliers, the comparison between the original data and revised local anomaly value in inspection data in February 23, 2009 is shown in Figure 5.

Local details of correction data are shown in Figure 6.

5. Data Correction

The practice of using mileage offset data to analyze track state at specified measuring point not only brings large deviation and does not reflect the true state but also is of no significance. So offset correction is needed.

There are two types of data correction: absolute correction and relative correction.

- (1) Absolute correction refers to the situation when the mileage that each measuring point corresponds to after correction is the accurate mileage. As is shown

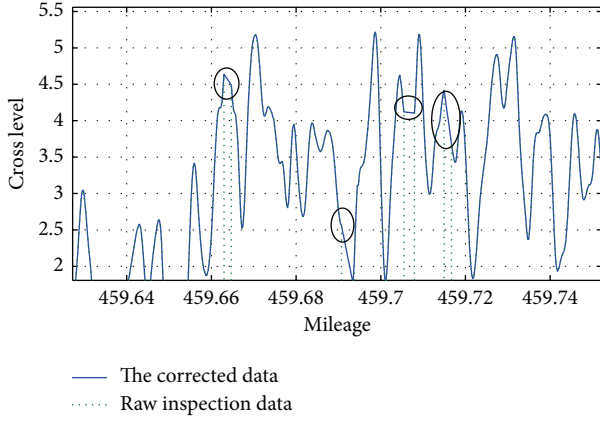


FIGURE 6: Details of the correction data.

in Figure 7, the actual mileage data is set for the reference point data, and other data corrects the mileage referring to it. In practice, it needs to know the precise mileage data of the measuring points in precise calibration, but it is difficult to be realized in fact, and it has little significance to research and practical application.

- (2) The relative correction refers to the situation that all measuring points of each inspection data after correction are pointing at the same mileage. As is shown in Figure 8, each inspection data takes t_1 mileage point data as the reference data, and other data corrects the mileage referring to it. But the mileage point may shift with the actual mileage points.

Both data after the above two types of correction can be used to do the time series data analysis, and there is little difference in practice. The latter is used in this paper.

The goal of mileage correction is to find each measuring point track irregularity status trends over time. Without mileage correction, the correspondent mileage of the all previous inspection data at each correspondent point is not the same with the actual mileage. This is similar to the practice that using the time series data consisted of data at different points to analyze the state changes of a certain point, and this will inevitably lead to inaccurate results.

In this paper, the idea of track space irregularity waveform similarity matching is applied to track irregularity mileage correction of sections. Typically, similarity distance is used to judge the similarity between two sequences. Euclidean distance and non-Euclidean distance are two types of methods to measure the distance.

Because Euclidean distance needs strict correspondence between all points of the sequence in the process of computing, and as a result, the following situation will appear: even a slight shift in the mileage of the inspection data will also make Euclidean distance between the two sections become large. Hence the deficiencies of Euclidean distance needs to be overcome.

In order to solve the problems of drift and noise data in track inspection car mileage data, this paper presents time series correction method based on trend similarity level.

The gauge inspection data in February 20, 2008, to November 13, 2008, Beijing-Kowloon line, section of K500+000–K500+075 km is selected for the study. The distribution of gauge inspection data of two adjacent sections before correction is shown in Figure 9.

The distribution of gauge irregularity inspection data details between two inspections on July 24, 2008, and August 16, 2008, is shown in Figure 10.

As can be seen from Figure 10, the gauge data have a certain offset compared to the corresponding mileage data.

There are three types of changing trends in adjacent track irregularity time series data elements: rising, falling, and flat. While $x_{j,t_i} > x_{j,t_{i-1}}$ ($1 \leq i-1 < i \leq n$), the data changing trend is upward; while $x_{j,t_i} < x_{j,t_{i-1}}$ ($1 \leq i-1 < i \leq n$), the data changing trend is downward; while $x_{j,t_i} = x_{j,t_{i-1}}$ ($1 \leq i-1 < i \leq n$), the data changing trend is flat. As the research is carried out on the same section repeatedly, all inspection data should reflect similar trends of the track irregularity state.

According to the idea of similar trends, data correction on track irregularity time series is done. There are four steps of data correction.

First Step: Trend Data Transformation. Gauge irregularity data is selected for the study. Assume the inspection time series data, whose length is n , consisted of n measurement points in the unit section as follows:

$$\begin{aligned}
 X_1 &= (x_{1,t_1}, x_{1,t_2}, \dots, x_{1,t_i}, \dots, x_{1,t_n}), \\
 X_2 &= (x_{2,t_1}, x_{2,t_2}, \dots, x_{2,t_i}, \dots, x_{2,t_n}), \\
 &\vdots \\
 X_j &= (x_{j,t_1}, x_{j,t_2}, \dots, x_{j,t_i}, \dots, x_{j,t_n}), \\
 &\vdots \\
 X_n &= (x_{n,t_1}, x_{n,t_2}, \dots, x_{n,t_i}, \dots, x_{n,t_n}).
 \end{aligned} \tag{2}$$

In this formula, X_j is inspection sequence data formed of the j th inspection of the section and X_{j+1} is inspection sequence data formed of the $j+1$ th inspection of the section. As there is mileage offset in track inspection data, inconsistencies exist in mileages of the measuring points corresponding to the two sequences.

Trend processing methods of data are as follows.

First, define the trend series X'_j , $X'_j = (x'_{j,t_1}, x'_{j,t_2}, \dots, x'_{j,t_i}, \dots, x'_{j,t_{n-1}})$.

Then, series X_j is transformed into a series trend X'_j .

When $x_{j,t_{i+1}} > x_{j,t_i}$ ($1 \leq i \leq n-1$), $x'_{j,t_i} = 1$.

When $x_{j,t_{i+1}} < x_{j,t_i}$ ($1 \leq i \leq n-1$), $x'_{j,t_i} = -1$.

When $x_{j,t_{i+1}} = x_{j,t_i}$ ($1 \leq i \leq n-1$), $x'_{j,t_i} = 0$.

The process of sequence X_{j+1} is the same with X_j .

Here, x'_{j,t_i} is the element of the trend series X'_j , and the length of trend series X'_j is $(n-1)$. Elements of the trend series

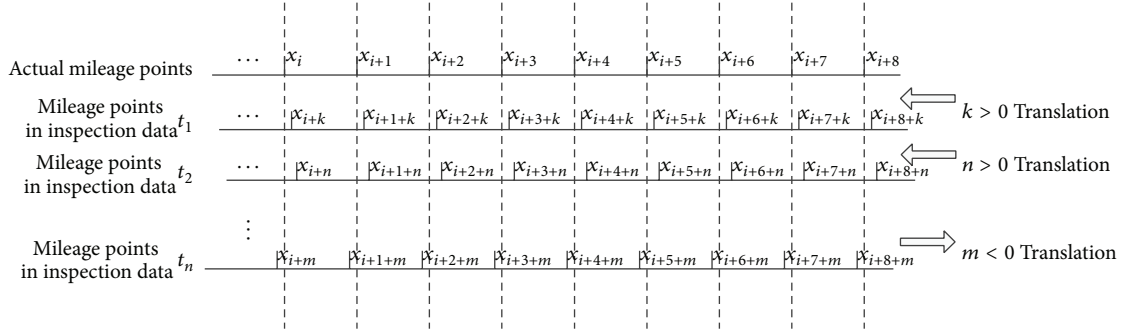


FIGURE 7: Schematic diagram of mileage absolute calibration.

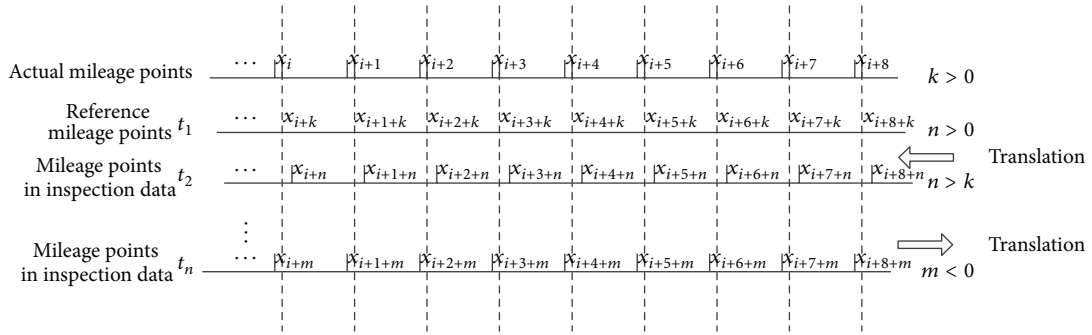


FIGURE 8: Schematic diagram of mileage relative calibration.

data X'_j and X'_{j+1} after conversation are composed of 1, 0, -1, such as

$$\begin{aligned} X'_j &= (1, 1, 1, -1, 1, -1, -1, 0, \dots), \\ X'_{j+1} &= (1, 1, -1, -1, 1, -1, 1, 0, \dots). \end{aligned} \quad (3)$$

Graphical representation of the process of sequence trend transformation is shown in Figure 11.

Step Two: Calculate the Similarity of Trend Sequences. To evaluate the similarity of the trend sequences X'_j and X'_{j+1} , the idea is as follows. When the number of equal corresponding elements between the trend sequences X'_j and X'_{j+1} is larger, the similarity of trends sequences X'_j and X'_{j+1} is higher.

Similarity Level Calculation. The resulting tendency sequence subtracts from each other to form a new sequence:

$$H = X'_{j+1} - X'_j. \quad (4)$$

Assume the number of elements in the sequence H is n and the number of 0 elements is h , and then define the similarity level between the trends sequences X'_j and X'_{j+1} as

$$l = \frac{h}{n} \times 100\%. \quad (5)$$

The larger the number of 0 elements in sequence H is, the greater the value l is and the higher the similarity of trends

sequences X'_j and X'_{j+1} is. This is the sequence's similarity level before correction.

Third Step: The Sequence Translation Transformation. Translation transformation includes left and right translation transformation, in which both left and right are relative to the reference sequence. Take X'_j as reference sequence, left and right translation transformation are carried out. The translation distance is the distance of m measuring and translation distance is set as $0.25m$ each time.

(1) *Left Translation Transformation.* Each time when X'_{j+1} is moved left for a measuring point distance, the operation would amputate the first element of X'_{j+1} and the last element of X'_j . In this case, after elements truncation, the two sequences X'_j and X'_{j+1} are of equal length. After one step shift operation, elements of the two sequences are corresponding to each other. Next, do the subtraction on the two new sequences, and then calculate the number of zero elements in the sequence formed by subtraction and then calculate the similarity level after the first left translation transformation.

The above process is repeated until m measuring points are moved left, and m similarity level values are achieved.

(2) *Right Translation Transformation.* The ideological of right translation transformation process is the same with the left. M similarity level values can be obtained after m times of right translation transformation.

Step Four: Data Correction. Trend similarity level values between the original trend sequences before the translation

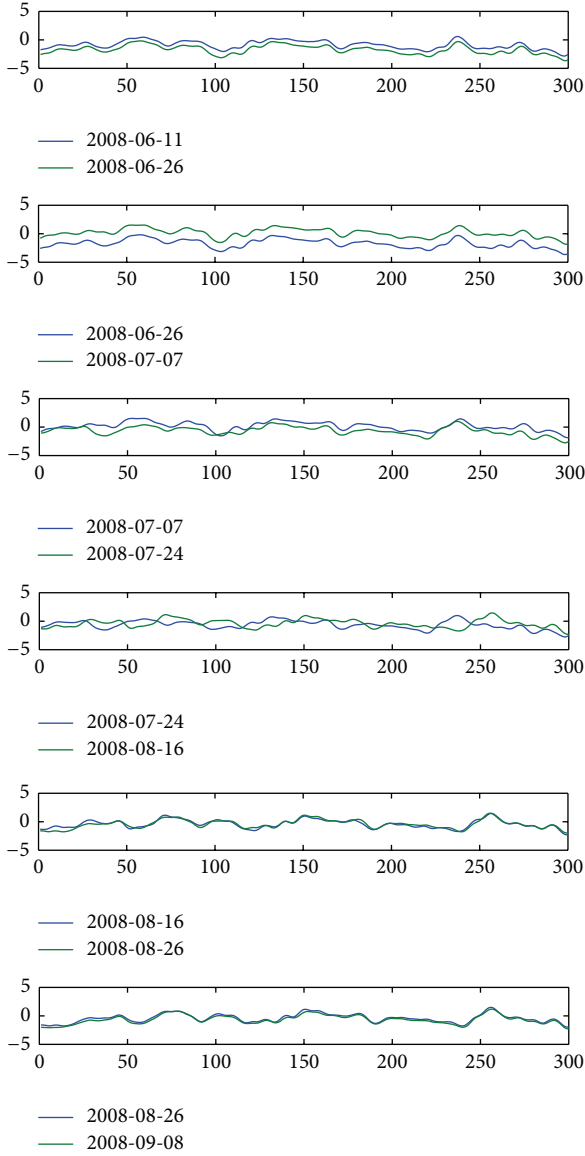


FIGURE 9: Distribution of gauge irregularity inspection data from February 2, 2008, to June 11, 2008, before mileage correction.

transformation and the similarity level values after m left and m right translation transformation are selected, and the maximum value of $2m + 1$ similarity level values is selected as correction criterion, which is as follows in detail.

- (1) If the trend similarity level before the translation transformation is the maximum, then the sequence needs no correction.
- (2) When l_{ik} (k steps left) is the maximum value of similarity level in m similarity level ($l_{i1}, l_{i2}, \dots, l_{im}$) of the left translation transformation, then X'_{j+1} and X'_j have the greatest similarity when X'_{j+1} moves the distance of k measuring points to the left. Since X'_{j+1} and X'_j are obtained by transformation of X_{j+1} and X_j ; therefore, the position of X_j and X_{j+1} has

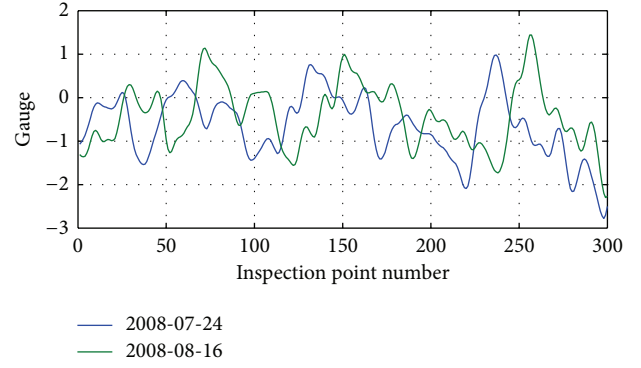


FIGURE 10: Distribution of gauge irregularity inspection data between July 24, 2008, and August 16, 2008.

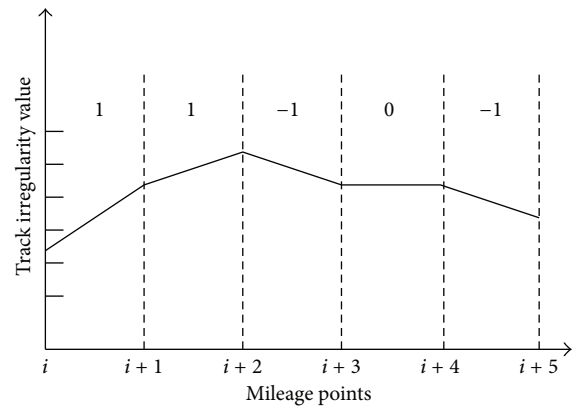


FIGURE 11: Trend analyses of time series data.

the maximum coherence after X_{j+1} moves the distance of k measuring points to the left, and data mileage between X_j and X_{j+1} is corrected to be aligned with each other.

- (3) When l_{rp} (p steps right) is the maximum value of similarity level in m similarity level ($l_{r1}, l_{r2}, \dots, l_{rm}$) of the right translation transformation, then X'_{j+1} and X'_j have the greatest similarity when X'_{j+1} moves the distance of p measuring points to the right. Due to the fact that X'_{j+1} and X'_j are obtained by transformation of X_{j+1} and X_j , therefore, the position of X_j and X_{j+1} has the maximum coherence after X_{j+1} moves the distance of p measuring points to the right, and data mileage between X_j and X_{j+1} is corrected to be aligned with each other.

According to experience, the value range of m is generally set from 40 to 100.

Two adjacent inspections sequences can be calibrated by translation transformation through finding the position of the maximum value of the similarity level of two adjacent sequences. If the overall mileage data of n times inspection data at section is calibrated, a certain time inspection data can be set as a reference data sequence (generally first inspection

TABLE 1: Statistics table of similarity level and mileage correction distance.

Number	Adjacent inspection time	The maximum similarity level	Translational direction	Translation distance	Overall translation distance
1	2008-02-20 and 2008-03-06	235/299	Left	2	Left 2
2	2008-03-06 and 2008-03-23	270/299	Right	1	Left 1
3	2008-03-23 and 2008-04-09	267/299	Left	1	Left 2
4	2008-04-09 and 2008-04-24	264/299	Right	1	Left 1
5	2008-04-24 and 2008-05-26	243/299	Inalienable	0	Left 1
6	2008-05-26 and 2008-06-11	247/299	Left	1	Left 2
7	2008-06-11 and 2008-06-26	256/299	Left	1	Left 3
8	2008-06-26 and 2008-07-07	239/299	Left	1	Left 4
9	2008-07-07 and 2008-07-24	231/299	Right	2	Left 2
10	2008-07-24 and 2008-08-16	220/299	Left	19	Left 21
11	2008-08-16 and 2008-08-26	232/299	Inalienable	0	Left 21
12	2008-08-26 and 2008-09-08	254/299	Inalienable	0	Left 21
13	2008-09-08 and 2008-09-24	254/299	Right	4	Left 17
14	2008-09-24 and 2008-10-10	260/299	Left	1	Left 18
15	2008-10-10 and 2008-11-13	234/299	Inalienable	0	Left 18
16	2008-11-13 and 2008-12-12	246/299	Inalienable	0	Left 18
17	2008-12-12 and 2008-12-25	244/299	Inalienable	0	Left 18

data is selected), and other sequences do translation transformation according to the position of the maximum value of the similarity level of two adjacent inspection sequences data. The statistics table of similarity level and translation transformation distance is shown in Table 1.

After calibration, the distribution of two adjacent gauge inspection data of sections is shown in Figure 12.

Distribution of gauge irregularity data of July 24, 2008 and August 16, 2008 is shown in Figure 13.

It should be noted that the mileage offset correction in this study here is a relative correction, because the first inspection sequence is set as a reference sequence in the correction process, and the mileage data is assumed to be with no offset. But the reality is that there is also mileage offset of the reference sequence compared to real mileage data. Therefore, this calibration process belongs to relative correction, but each inspection data is aligned with each other after calibration, which will provide a data base for

the research of each measurement point or smaller section within the time track status changes in the following studies.

6. Track Irregularity Time Series Data Wavelet Decomposition-Reconstruction

The wavelet transform [28–31] is a new rapidly evolving field of applied mathematics and engineering disciplines; it is a new branch of mathematics, which is the perfect crystal of functional analysis, Fourier analysis, sample transfer analysis, and numerical analysis. Data process or data series is converted into stages data series to find similar spectrum characteristics based on some special functions in the wavelet transform, so as to achieve a data processing. Wavelet transform is local transformation of space (time) and frequency, and it can effectively extract information from the signal and do multiscale detailed analysis to function or signal through stretching and panning arithmetic.

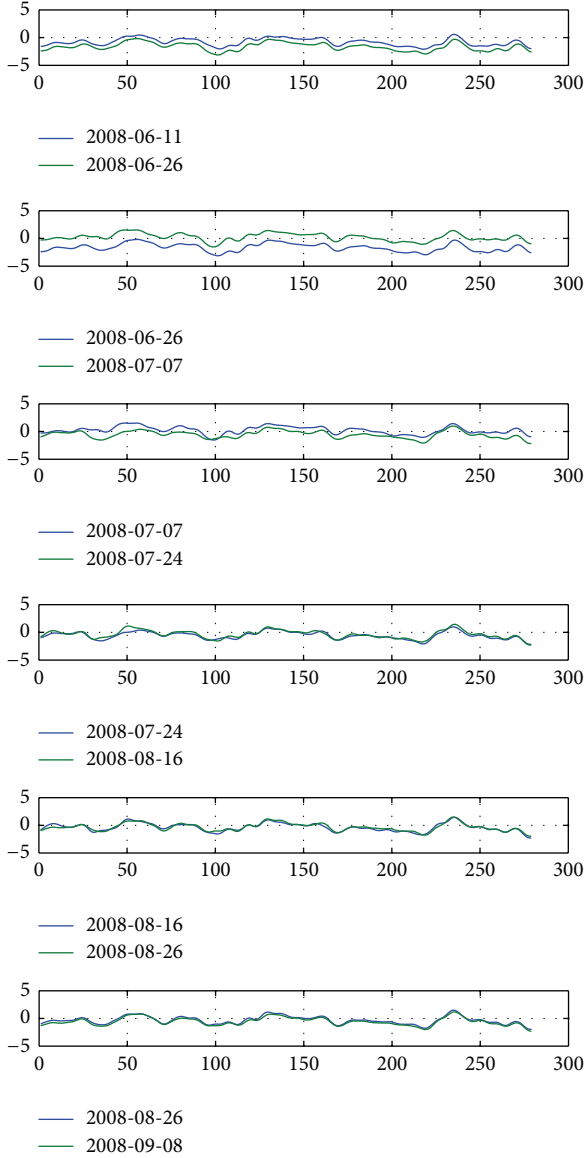


FIGURE 12: Distribution of gauge irregularity inspection data from February 20, 2008, to June 11, 2008, after mileage correction.

“Wavelet” means the waveform with a small area, the limited length and 0 mean, in which “small” refers to the wavelet with decay, “wave” refers to its volatility, and its amplitude shocks in alternating positive forms and negative forms. Compared with the Fourier transform, wavelet transform is the localized analysis of the time (space) frequency. It does multistage subdivision gradually through stretching shift operation on the signal (function) and ultimately achieves time segments at high frequency and frequency segments at low frequency and can automatically adapt to the requirements of time-frequency signal analysis, and then can focus on any detail of the signal and thus can solve the difficult problem of Fourier transform. It has become a major breakthrough in the scientific method since Fourier transform, so wavelet transform is even called “mathematical microscope”.

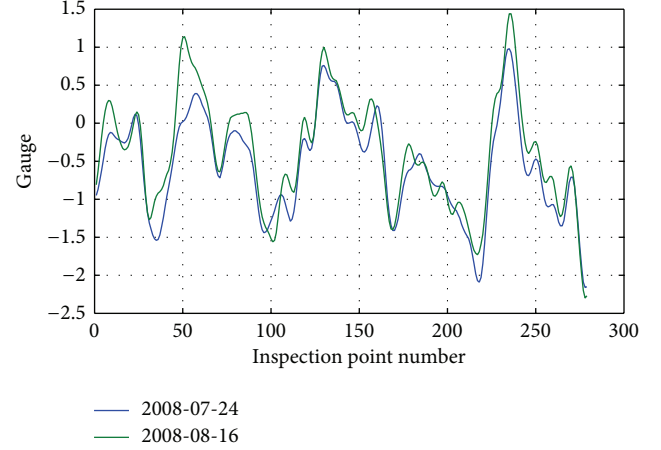


FIGURE 13: Distribution of details of correction of gauge irregularity inspection data between July 24, 2008, and August 16, 2008.

The decomposition of the function into the representation of a series of simple basis functions has an important significance both in theory and in practice. In this paper, Daubechies wavelet [32, 33] is used to do decomposition in track irregularity time series data, which is the general term for a series of binary proposed by the French scholar Daubechies, and multiscale wavelet decomposition of the signal can be done by it.

Assume a known signal

$$f(x) = \sum a_{j,k} \phi_{j,k}(x), \quad f(x) \in V_j. \quad (6)$$

The coefficients $\{a_{j,k}, k \in Z\}$ are known in the formula.

Now $f(x)$ is decomposed into two components of space V_{j-1} and space W_{j-1} :

$$f(x) = \sum a_{j-1,k} \phi_{j-1,k}(x) + \sum d_{j-1,k} \psi(x). \quad (7)$$

In a given situation of sequence $\{a_{j,k}\}$, respectively, the $(J-1)$ th approximate level sequence $\{a_{j-1,k}\}$ and $(j-1)$ th details level sequence $\{d_{j-1,k}\}$ can be calculated. According to two scale relations, it can be known that

$$\begin{aligned} \phi_{j-1,k} &= 2^{(j-1)/2} \phi(2^{j-1}x - k) \\ &= 2^{(j-1)/2} \sqrt{2} \sum_s h_s \phi(2(2^{j-1}x - k) - s) \\ &= \sum_s h_s 2^{j/2} \phi(2^j x - (2k + s)) = \sum_s h_s \phi_{j,2k+s}(x). \end{aligned} \quad (8)$$

Similarly, it can be calculated that

$$\psi_{j-1,k}(x) = \sum_s g_s \phi_{j,2k+s}(x). \quad (9)$$

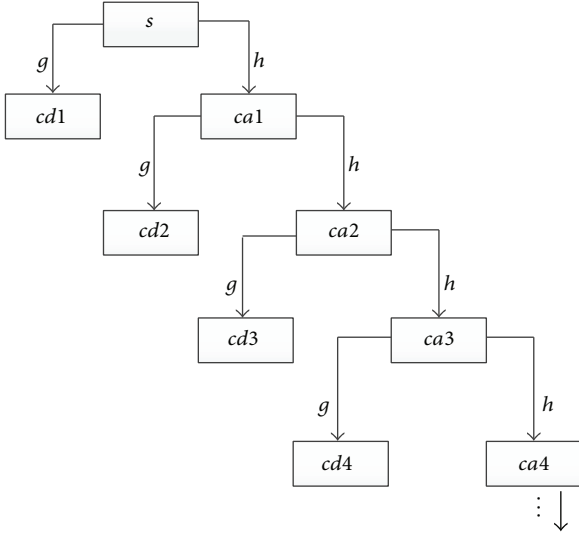


FIGURE 14: Process of wavelet decomposition.

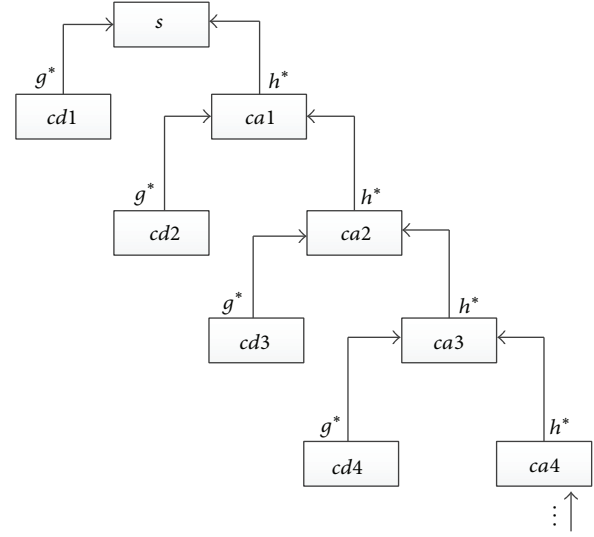


FIGURE 15: Process of wavelet reconstruction.

It can be inferred according to the above relation that

$$\begin{aligned}
 a_{j-1,k} &= \langle f(x), \phi_{j-1,k}(x) \rangle \\
 &= \left\langle f(x), \sum_s h_s \phi_{j,2k+s}(x) \right\rangle \\
 &= \sum_s \bar{h}_s \langle f(x), \phi_{j,2k+s}(x) \rangle \\
 &= \sum_s \bar{h}_s a_{j,2k+s} \\
 &= \sum a_{j,n} \bar{h}_{n-2k} \\
 &= a_j \times h'(2k).
 \end{aligned} \tag{10}$$

In the formula, $h'_k = \bar{h}_{-k}$.

Similarly, it can be calculated that

$$d_{j-1,k} = a_j \times g'(2k), \quad g'_k = \bar{g}_{-k}. \tag{11}$$

The second level decomposition $\{a_{j-2,k}\}$ and $\{d_{j-2,k}\}$ can be obtained after doing decomposition on the approximate sequence $\{a_{j-1,k}\}$ resulted from the first stage of decomposition results again, and the third level decomposition $\{a_{j-3,k}\}$ and $\{d_{j-3,k}\}$ can be obtained after doing decomposition on the approximate sequence $\{a_{j-2,k}\}$ resulted from the second stage of decomposition results, and so on, until the multiscale wavelet is decomposed into a specified stage.

The decomposition process is called Mallat Pyramidal algorithm as shown in Figure 15. Mallat algorithm is inspired by the famous Pyramidal algorithm [34] for image decomposition and combined with multiresolution analysis, proposing signal tower multiresolution decomposition and synthesis algorithms. It is named after the data structure which is a tower structure in decomposition process. Decomposition and reconstruction process is shown in Figures 14 and 15.

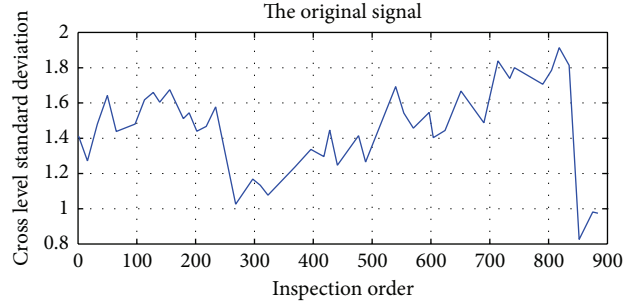


FIGURE 16: Original waveform signal of track irregularity.

S is the original signal in figure, $cd1$ and $ca1$ are detail sequence and approximate sequence after level 1 decomposition, and $cd2$ and $ca2$ are detail sequence and approximate sequence after level 2 decomposition, and so on.

Standard deviation reflects changes of the deviation from mean of track irregularity. When distribution of irregularity around the mean value is more discrete, the representation of average is poorer, and track irregularity will be in poorer state. Conversely, the smaller the standard deviation is, the smaller the variation between the track irregularity values is, the denser the irregularity distribution around the mean value is, and the better the representation of the mean and track state is.

Changes of track irregularity standard deviation series data of the Beijing-Kowloon line K449+000–K450+000 sections within 44 times inspection are selected as the research object, and the original signal is shown in Figure 16.

Daubechies wavelet is chosen in signal decomposition of track irregularity standard deviation time series data, with the decomposition depth 3. Mallat tower algorithm is used for decomposition and reconstruction of track irregularity standard deviation time series. After wavelet decomposition, 1, 2, and 3 layers are the waveform signal (high frequency) of details, respectively, represented by $D1$, $D2$, and $D3$; and

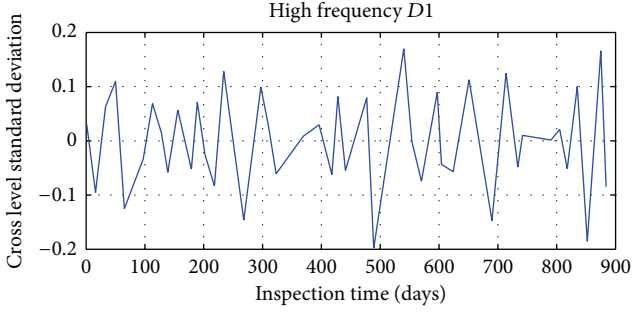


FIGURE 17: The first layer detail waveform signal of track irregularity (HF).

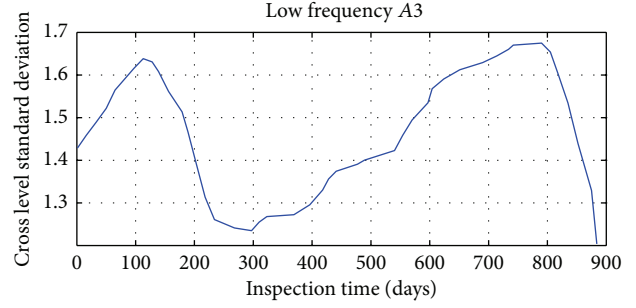


FIGURE 20: The third layer approximation waveform signal of track irregularity (LF).

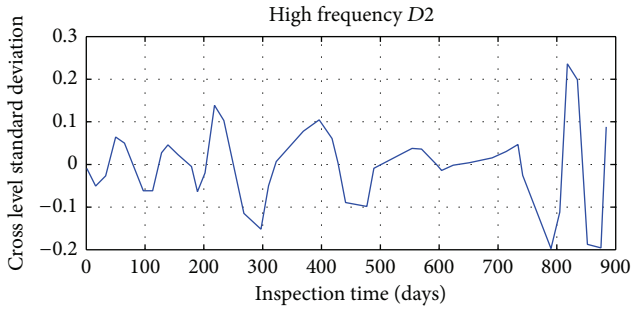


FIGURE 18: The second layer detail waveform signal of track irregularity (HF).

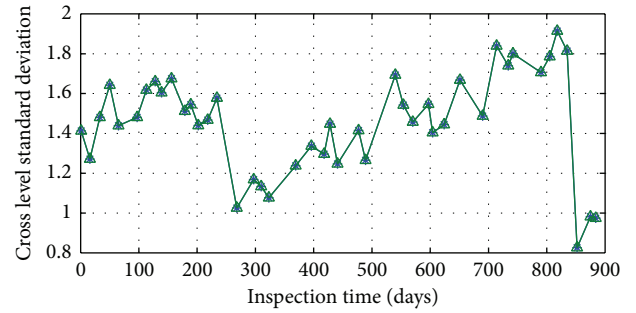


FIGURE 21: Comparison between original data sequence and reconstructed data sequence.

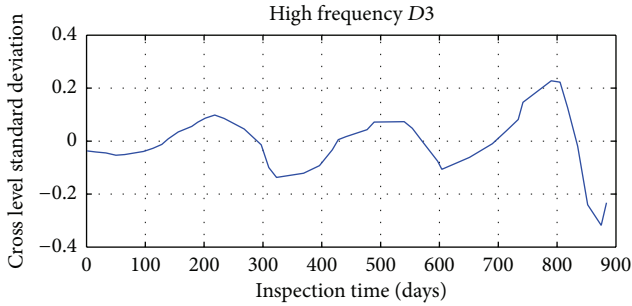


FIGURE 19: The third layer detail waveform signal of track irregularity (HF).

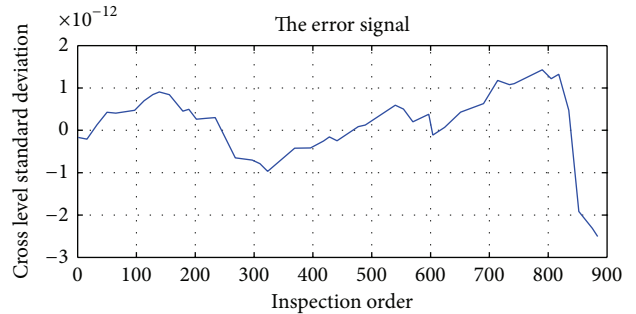


FIGURE 22: Data error.

approximate sequence waveform signal (LF) of layer 3 is represented by A3.

The results of specific decomposition are shown in Figures 17, 18, 19, and 20.

All layers waveform signals after decomposition are reconstructed with a weight of 1, and the reconstruction formula is as follows:

$$s = ca3 + cd3 + cd2 + cd1. \quad (12)$$

Reconstruction results are shown in Figure 21.

Error analysis is shown in Figure 22.

Error analysis showed that when reconstruction is used by weight of 1, the order of error will be 10^{-12} , which is basically negligible.

It is very important to study the wavelet decomposition-reconstruction of track irregularity data. After wavelet decomposition, track irregularity time series data can be transformed into multifeature smooth sequence from non-stationary characteristics, which is an effective data preprocessing method in time series modeling with the premise for a smooth sequence. By wavelet decomposition, further clarification can be done to the characteristics of data changes and thus can provide a basis for classification, clustering, and pattern recognition. Meanwhile, by modeling and analysis on data at each layer, respectively, optimal fit and predictive models can be obtained, and then we can carry out weighted calculation to models of all layers and then get fit and

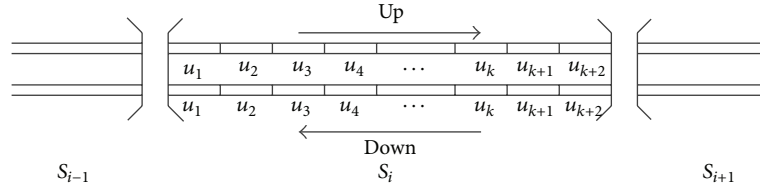


FIGURE 23: Selection of study data.

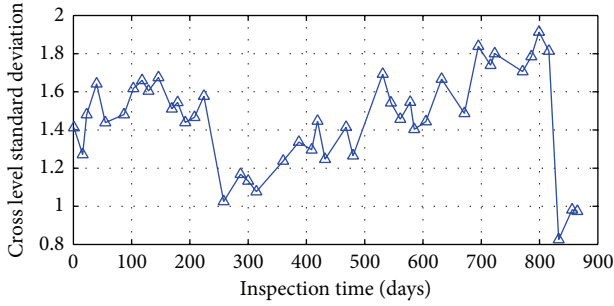


FIGURE 24: Cross level status trends at unit section of K449+200-K449+225.

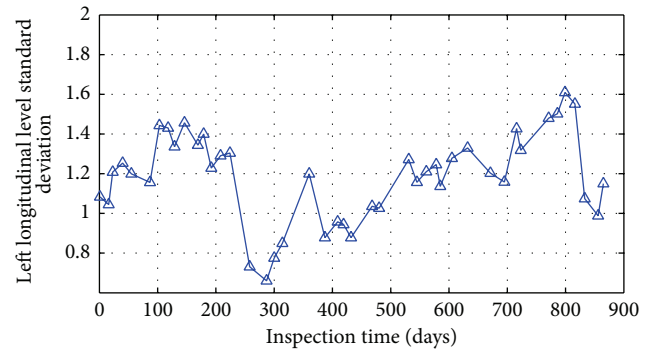


FIGURE 25: Left longitudinal level status trends at unit section of K449+675-K449+700.

predicted values of the original track irregularity time series data.

7. Change Mode of Unit Section

It is less meaningful to study track state changes of a fixed inspection point; based on the tools and interval of data collection, it is of great significance to study the state changes of the overall length of certain sections.

Track Irregularity inspection data appears near zero mean, positive and negative phases alternatively. There is a strong stochastic changing characteristic of each measuring point in track irregularity state inspection process. Character of track irregularity state in a single measuring point position showed that track irregularity track geometry data fluctuate on the standard values, but this variable is a random process, with the direction and the size changing from time to time, and the real trend of track state changes cannot be reflected. Therefore, irregularity size change in a single direction and magnitude of a single track geometry measurement points should not be seen as the basis in the study. The distribution deviating from the normal value and the rate of development of the unit section should be used to measure changes of track irregularity values.

In summary, to study the features of a certain length of section track irregularity state changes, the standard deviation of track irregularity inspection data can be used as the object in study.

Take the 44 times' inspection data of the cross level and longitudinal track irregularity, Beijing-Kowloon line K449+000-K450+000 section, in 884 days, between February 20, 2008, and July 23, 2010, as the study data. The section is divided into 40 unit sections, and each data contain 100 measure points, and the track state changes of each unit

section are studied. The way of data selection is shown in Figure 23.

S_{i-1} , S_i , and S_{i+1} are division of unit section in study in figure, 1km is selected as the length of unit section in this paper, and u_1, u_2, \dots are unit sections divided in the study.

According to the data selection methods above, track irregularity standard deviation data of some unit sections is shown in Figures 24 and 25.

There are two types of changing modes presented in the standard deviation curve in the changing process over time from the analysis of Figures 24 and 25.

(1) *Jump*. In the adjacent inspection time $t_{i-n}, \dots, t_i, t_{i+1}, \dots, t_{i+m}$, $sd_{i-n}, \dots, sd_i, sd_{i+1}, \dots, sd_{i+m}$ are track irregularity standard deviations corresponding to them, when

$$\begin{aligned}
 |sd_{i+1} - sd_i| &\gg |sd_i - sd_{i-1}|, \\
 |sd_{i+1} - sd_i| &\gg |sd_{i-1} - sd_{i-2}|, \\
 &\vdots \\
 |sd_{i+1} - sd_i| &\gg |sd_{i+2} - sd_{i+1}|, \\
 |sd_{i+1} - sd_i| &\gg |sd_{i+3} - sd_{i+2}|, \\
 &\vdots
 \end{aligned} \tag{13}$$

This phenomenon of standard deviation curves is considered to be showing jump change in the adjacent time t_i, t_{i+1} . The reason for the jump changes in the time t_i is that track state degradation reaches a critical value of maintenance, and the maintenance operation is imminent; the track condition

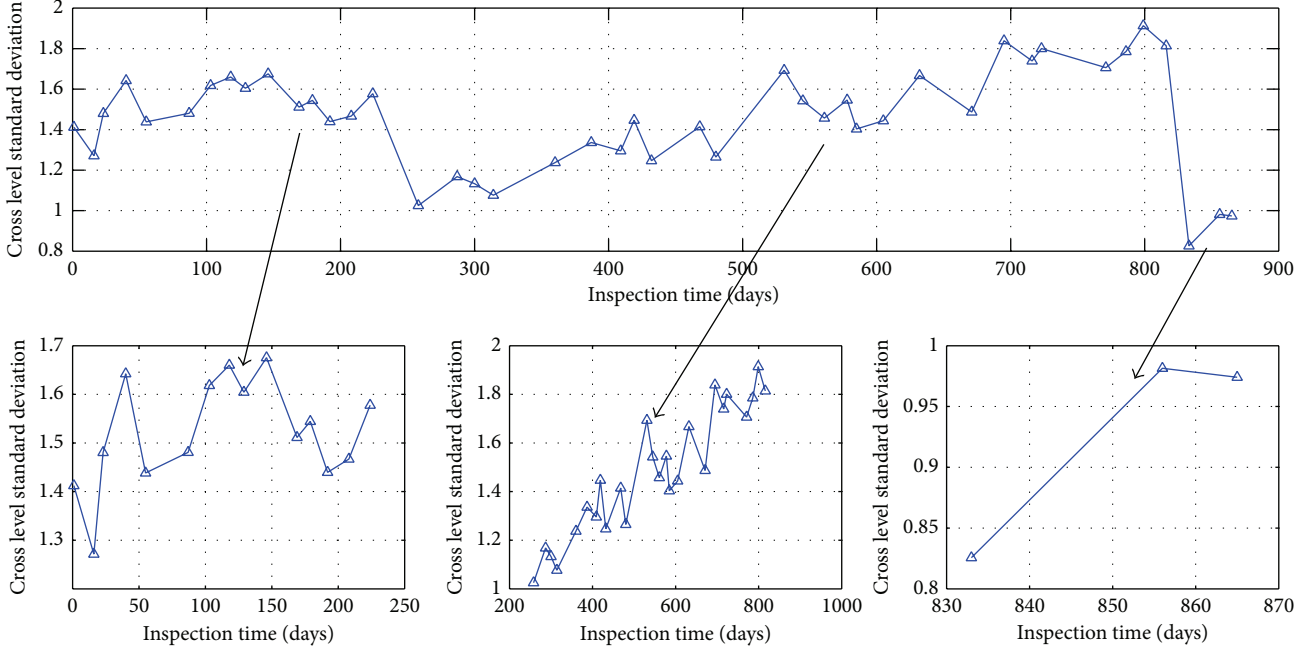


FIGURE 26: Cycle decomposition of cross level trend at unit section of K449+800–K449+825.

is significantly becoming better in time t_{i+1} , showing that the track has undergone maintenance operations. This jump change is the demarcation point of track status cycle change.

(2) *Gradual Variation.* In the adjacent inspection time $t_{i-n}, \dots, t_i, t_{i+1}, \dots, t_{i+m}$, $sd_{i-n}, \dots, sd_i, sd_{i+1}, \dots, sd_{i+m}$ are track irregularity standard deviations corresponding to them, when

$$\begin{aligned}
 |sd_{i+1} - sd_i| &\approx |sd_i - sd_{i-1}|, \\
 |sd_{i+1} - sd_i| &\approx |sd_{i-1} - sd_{i-2}|, \\
 &\vdots \\
 |sd_{i+1} - sd_i| &\approx |sd_{i+2} - sd_{i+1}|, \\
 |sd_{i+1} - sd_i| &\approx |sd_{i+3} - sd_{i+2}|, \\
 &\vdots
 \end{aligned} \tag{14}$$

This phenomenon of standard deviation curves is considered to be showing gradual changes in the adjacent time t_i, t_{i+1} . The reason for the gradual change is that the track changes in a steady state at the moment of t_i and the adjacent time, indicating that track changes are in a maintenance cycle currently.

It can be considered that the changes of cross level standard deviation and left longitudinal level standard deviation show a periodic growth pattern through the curve geometric features in Figures 24 and 25. Take the changes of cross level standard deviation state at K449+800–K449+825 unit section as the example; the changing trend of track irregularity state characters in 884 days is divided by the two jump models at

268th days and 835th days into three cycles. Among this, it is a complete changing cycle between the 268th days and the 835th days. The cycle is shown in Figure 26 periodically.

Cyclical characteristics result from the operation of railway maintenance. Due to regular track irregularity maintenance and repair work, as well as the timely remediation repair in cases of overrun, the trend curve of track irregularity often shows a number of jumping phenomena. Also, since this maintenance and repair work are conducted at specific locations, for example, they are operated within tens of meters, a few hundred meters, and even several kilometers and the difference of irregularities and targets in remediation operations, for example, one or a number of remediation to cross level and longitudinal level, thus in the specific object section study, the cyclical nature of the state reflected by each single irregularity inspection will be different.

8. Conclusions

The characteristics of track irregularity data are systematically analyzed in this paper. Targeted on the problems of data quality, data offset correction algorithm is proposed based on trends similarity, as well as the outlier identification and noise cancellation algorithms based on the abnormal degree, so as to do treatment on data. Next, the paper proposes track irregularity time series decomposition and reconstruction by using the wavelet decomposition and reconstruction approach. Finally, since the data of track geometry irregularity reflect dynamic changing characteristics of the track state, as a result, through the research on pattern features of track irregularity standard deviation data series of the section, the changing trends of data is discovered and described. The model proposed in this paper is a general model and

model can be used in most cases. The results can provide a theoretical basis for subsequent track condition predictions.

Conflict of Interests

The authors declare that they have no financial and personal relationships with other people or organizations that can inappropriately influence their work; there is no professional or other personal interests of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in, or the review of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant no. 61272029) and National Key Technology R&D Program (Grant no. 2009BAG12A10).

References

- [1] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, Oakland, Calif, USA, 1976.
- [2] P. S. P. Cowpertwait and A. V. Metcalfe, *Introductory Time Series with R*, Springer, New York, NY, USA, 2009.
- [3] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.
- [4] B. Parmentier and J. R. Eastman, "Land transitions from multivariate time series: using seasonal trend analysis and segmentation to detect land-cover changes," *International Journal of Remote Sensing*, vol. 35, no. 2, pp. 671–692, 2014.
- [5] M. A. Nishi, C. F. Ahmed, M. Samiullah, and B.-S. Jeong, "Effective periodic pattern mining in time series databases," *Expert Systems with Applications*, vol. 40, no. 8, pp. 3015–3027, 2013.
- [6] T. Vondra and J. Sedivy, "Maximizing utilization in private IaaS clouds with heterogenous load through time series forecasting," *International Journal on Advances in Systems and Measurements*, vol. 6, no. 1-2, pp. 149–165, 2013.
- [7] A. Prasad, S. Manna, A. Mahato, and V. K. Singh, "The generalized continuous wavelet transform associated with the fractional Fourier transform," *Journal of Computational and Applied Mathematics*, vol. 259, part B, pp. 660–671, 2014.
- [8] S. Jin, L. Su, and A. Ullah, "Robustify financial time series forecasting with bagging," *Econometric Reviews*, vol. 33, no. 5-6, pp. 575–605, 2014.
- [9] L. Meligkotsidou and I. D. Vrontos, "Detecting structural breaks in multivariate financial time series: evidence from hedge fund investment strategies," *Journal of Statistical Computation and Simulation*, vol. 84, no. 5, pp. 1115–1135, 2014.
- [10] S. Alvisi, N. Ansaloni, and M. Franchini, "Generation of synthetic water demand time series at different temporal and spatial aggregation levels," *Urban Water Journal*, vol. 11, no. 4, pp. 297–310, 2014.
- [11] J.-N. Pan, P. Kung, A. Bretholt, and J.-X. Lu, "Prediction of energy's environmental impact using a three-variable time series model," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1031–1040, 2014.
- [12] S. M. Shahandashti and B. Ashuri, "Forecasting engineering news-record construction cost index using multivariate time series models," *Journal of Construction Engineering and Management*, vol. 139, no. 9, pp. 1237–1243, 2013.
- [13] M. Markus, H. V. Knapp, A. Flegel, S. McConkey, and W. O. Thomas, "Episodic change analysis of the annual flood peak time series for a flood insurance study," *Journal of Hydrologic Engineering*, vol. 18, no. 1, pp. 85–91, 2013.
- [14] D. Das and A. Maitra, "Time series prediction of rain attenuation from rain rate measurement using synthetic storm technique for a tropical location," *AEU—International Journal of Electronics and Communications*, vol. 68, no. 1, pp. 33–36, 2014.
- [15] P. E. Williams and W. D. Pesnell, "Time-series analysis of supergranule characteristics at solar minimum," *Solar Physics*, vol. 289, no. 4, pp. 1101–1113, 2014.
- [16] R. Palivonaite and M. Ragulskis, "Short-term time series algebraic forecasting with internal smoothing," *Neurocomputing*, vol. 127, pp. 161–171, 2014.
- [17] J. A. Ferreira, R. H. Loschi, and M. A. Costa, "Detecting changes in time series: a product partition model with across-cluster correlation," *Signal Processing*, vol. 96, pp. 212–227, 2014.
- [18] T. J. Glezakos, T. A. Tsiligiridis, and C. P. Yialouris, "Piecewise evolutionary segmentation for feature extraction in time series models," *Neural Computing and Applications*, vol. 24, no. 2, pp. 243–257, 2014.
- [19] P. R. A. Firmino, P. S. G. de Mattos Neto, and T. A. E. Ferreira, "Correcting and combining time series forecasters," *Neural Networks*, vol. 50, pp. 1–11, 2014.
- [20] J. F. Murphy, J. H. Winterbottom, S. Orton, G. L. Simpson, E. M. Shilland, and A. G. Hildrew, "Evidence of recovery from acidification in the macroinvertebrate assemblages of UK fresh waters: a 20-year time series," *Ecological Indicators*, vol. 37, pp. 330–340, 2014.
- [21] M. K. P. So and R. S. W. Chung, "Dynamic seasonality in time series," *Computational Statistics & Data Analysis*, vol. 70, pp. 212–226, 2014.
- [22] X. Meng, B. Cui, L. Jia, Y. Qin, and J. Xu, "Networked timetable stability improvement based on a bilevel optimization programming model," *Mathematical Problems in Engineering*, vol. 2014, Article ID 290937, 10 pages, 2014.
- [23] X. Meng, B. Cui, and L. Jia, "Line planning in emergencies for railway networks," *Kybernetes*, vol. 43, no. 1, pp. 40–52, 2014.
- [24] X. Meng and B. Cui, "Train repathing in emergencies based on fuzzy linear programming," *The Scientific World Journal*, vol. 2014, Article ID 598968, 10 pages, 2014.
- [25] X. Meng and L. Jia, "Train timetable stability evaluation based on analysis of interior and exterior factors information entropy," *Applied Mathematics & Information Sciences*, vol. 8, no. 3, pp. 1319–1325, 2014.
- [26] C. Jia, W. Xu, L. Wei, and H. Wang, "Study of railway track irregularity standard deviation time series based on data mining and linear model," *Mathematical Problems in Engineering*, vol. 2013, Article ID 486738, 12 pages, 2013.
- [27] J. Chaolong, X. Weixiang, W. Futian, and W. Hanning, "Track irregularity time series analysis and trend forecasting," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 387857, 15 pages, 2012.
- [28] A. A. Alugongo, "Parametric excitation and wavelet transform analysis of a ground vehicle propeller shaft," *Journal of Vibration and Control*, vol. 20, no. 2, pp. 280–289, 2014.
- [29] R. R. Sahay and A. Srivastava, "Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and

- neural network,” *Water Resources Management*, vol. 28, no. 2, pp. 301–317, 2014.
- [30] P. Rajmic and Z. Prusa, “Discrete wavelet transform of finite signals: detailed study of the algorithm,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 12, no. 1, Article ID 1450001, 38 pages, 2014.
- [31] C. A. Naik and P. Kundu, “Power quality index based on discrete wavelet transform,” *International Journal of Electrical Power & Energy Systems*, vol. 53, pp. 994–1002, 2013.
- [32] L. Pahlavan, C. Kassapoglou, and Z. Gürdal, “Spectral formulation of finite element methods using Daubechies compactly-supported wavelets for elastic wave propagation simulation,” *Wave Motion*, vol. 50, no. 3, pp. 558–578, 2013.
- [33] K. L. V. Iyer, X. Lu, Y. Usama, V. Ramakrishnan, and N. C. Kar, “A twofold Daubechies-wavelet-based module for fault detection and voltage regulation in SEIGs for distributed wind power generation,” *IEEE Transactions on Industrial Electronics*, vol. 60, no. 4, pp. 1638–1651, 2013.
- [34] O. Jemai, M. Zaied, C. Ben Amar, and M. A. Alimi, “Pyramidal hybrid approach: wavelet network with OLS algorithm-based image classification,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 9, no. 1, pp. 111–130, 2011.