

## Research Article

# Study of Railway Track Irregularity Standard Deviation Time Series Based on Data Mining and Linear Model

Jia Chaolong,<sup>1</sup> Xu Weixiang,<sup>2</sup> Wei Lili,<sup>3</sup> and Wang Hanning<sup>1</sup>

<sup>1</sup> State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>3</sup> Chongqing Public Security Bureau, Chongqing 401147, China

Correspondence should be addressed to Xu Weixiang; [wxxu@bjtu.edu.cn](mailto:wxxu@bjtu.edu.cn)

Received 13 July 2013; Revised 1 September 2013; Accepted 2 September 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Jia Chaolong 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.

Good track geometry state ensures the safe operation of the railway passenger service and freight service. Railway transportation plays an important role in the Chinese economic and social development. This paper studies track irregularity standard deviation time series data and focuses on the characteristics and trend changes of track state by applying clustering analysis. Linear recursive model and linear-ARMA model based on wavelet decomposition reconstruction are proposed, and all they offer supports for the safe management of railway transportation.

## 1. Introduction

Temporal data and temporal data mining which reflect the dynamic nature of data are one of the focuses of academic community in recent years. Time series is an important temporal data. Time series similarity has been widely used in speech recognition. Euclidean distance and dynamic time warping are two classic methods. Euclidean distance is most frequently used in the time series, but Euclidean distance is a very brittle distance measure [1]. There is an obvious defect of Euclidean distance: sometimes when the sequence is very similar but the distance is great. Then DTW is proposed. DTW is an algorithm for measuring the similarity between two sequences which may vary in time or speed. A well-known application has been automatic speech recognition [2–4], to cope with different speaking speeds. Since Agrawal et al. first proposed overall matching algorithm of time series similarity search in 1993 [5], more and more scholars began to focus on temporal data mining study.

In the past four decades, scholars have proposed a variety of classic time series forecast methods, including Autoregressive Model (AR) [6], Moving Average Model (MA) [7], Autoregressive Moving Average Model (ARMA) [8–10], and Autoregressive Integrated Moving Average (ARIMA) [11].

There are a variety of methods and models in the forecast area, such as determining function method, statistical regression analysis, time series analysis, Markov model [12], state-space model [13], Bayesian forecasting model [14], as well as a variety of methods combined with theories, techniques, and methods, such as the hybrid of fuzzy theory and linear regression [15], the hybrid ARIMA and support vector machines model [16], the hybrid of time series forecast using neural networks, fuzzy logic, fractal theory [17], and the hybrid Markov model and neural networks [18]. At the same time, the number of subjects is divided into univariate and multivariate. In time series analysis and forecasting methods, the most commonly used method is based on the time domain and frequency domain. Linear and stochastic linear models and nonlinear models are two main types of study models. Random process cannot be expressed by definite function. Typically, there are mainly two types of methods to analyze random process: one is a probabilistic method, and the other is the analysis method, and the two methods are often used simultaneously in practical study. Each of AR model, MA model, ARMA model, the Markov Forecast [19], and Kalman filter model [20, 21], can be used to study the stochastic process. The characteristics of the first three models are linear forecast model and are relatively simple in terms of elements taken into consideration.

In the recent 20 years, fuzzy system is one of the most frequently used methods [22–25]. Fuzzy system has been used for a variety of optimization algorithms. In many studies, the design of fuzzy system forecast has been proposed.

In the study of time series forecast, the forecast accuracy is considered as top priority in the selection of forecast methods. The neural network is the most representative time series forecasting method, which has drawn more and more attention. In the past ten years, the neural network model is used in the study of time series forecast. The discovery of the neural network is considered to be a competitor to a variety of traditional time series models [26–28]. Because of flexible computing framework and general approximation, artificial neural network [29] is widely used in the field of predictive analytics and has higher precision.

However, an important and difficult task is faced by decision makers in many areas to improve the accuracy of the time series forecast. Using a hybrid model by combining several models has become a common practice to improve forecast accuracy. This study field has been significantly expanded [30–34].

The study has shown that forecasts of certain hybrid models are often better than those obtained from a single model forecast. Hybrid model forecast is considered to be a more accurate forecast. Its main problem can be described as follows: suppose there are  $n$  kinds of forecasts, such as  $\hat{y}_1(t), \hat{y}_2(t), \dots, \hat{y}_n(t)$ . The general form of this hybrid forecasting model can be defined as weighted sums of  $\hat{y}_i(t)$ , and the sum of all weights stack up is 1. The biggest difficulty is to determine the weights of every single forecast.

According to the principles recognized by the scientific community, simple theories are more reliable than complex ones under the same circumstances. The best scientific theory should be the simplest. railway transportation is different from automobile traffic. In automobile traffic, the driver is the major safety factor [35, 36], but in Railway transportation, track irregularity state is the decisive factor. This paper tries to look for a simple, reliable model to analyze the track irregularity change trend and to explore knowledge.

The paper is organized as follows. Section 2 introduces forecast models of time series. Data mining of track irregularity time series is described in Section 3. Section 4 presents analysis and forecast of track irregularity time series based on linear recursive forecast model. Section 5 proposes linear-ARMA model based on wavelet decomposition reconstruction. Finally, Section 6 concludes the paper with a summary.

## 2. Methods and Processes of Time Series Data Study

**2.1. Forecast Methods of Time Series Data.** Typically, there are two methods to forecast time series data: qualitative forecast and quantitative forecast. Qualitative forecast method focuses on the forecast of the nature of the development, trends, direction, and major turning points in the road of development and mainly depends on human experience and analytical ability. Quantitative forecast focuses on the analysis of quantitative aspects of development, attaches importance

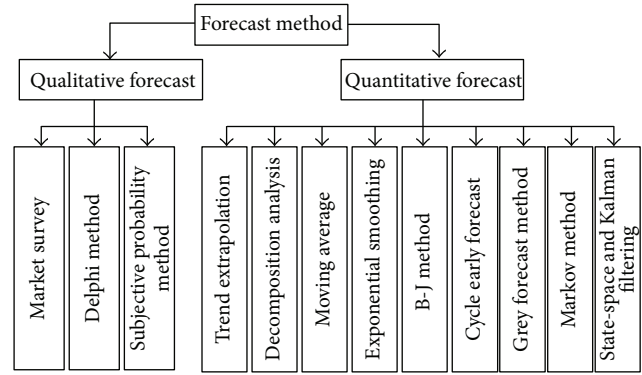


FIGURE 1: Main forecast methods of time series.

to the quantitative description of the degree of development and changes, is based more on historical statistics, and is less affected by subjective factors. Qualitative forecast and quantitative forecast are not mutually exclusive but can be complementary to each other and should be combined correctly in the actual forecast process.

Two types of time series forecasting methods and their methods are shown in Figure 1.

**2.2. Study Processes of Time Series Data.** The study processes of time series data include three steps: data acquisition, correlation analysis, and model identification. Specific study processes are shown in Figure 2.

## 3. Data Mining of Track Irregularity Time Series

**3.1. Theoretical Analysis of Data Mining.** Data mining is the process of mining interested information from databases, data warehouses, or other data repositories. It is an iterative process, whose study steps are shown in Figure 3.

The main method, purpose, and contents of time series data mining include: time series segmentation studying underlying mechanisms change of time series and representation at high level; similarity search looking for similar sequences; clustering analysis on similarity measure, clustering algorithms and results, and gathering similar time sequence variation into one class; classification and sequence analysis on time series and the time points in the entire time series; anomaly detection which finds the abnormality of sequence, points, and mode; analysis on the law and trends of time series changing over time, forecast of the future data and trends based on the historical and current data; using graphics technology, virtual reality technology, and data mining technology to display complex time series in a way which is easy, intuitive, and graphical for people to understand so that we can realize visualized and practical study of time series data.

**3.2. Time Series Similarity.** Time series similarity is the basis of time series analysis and time series data mining. The similarity is achieved by calculating the distance between the time series.

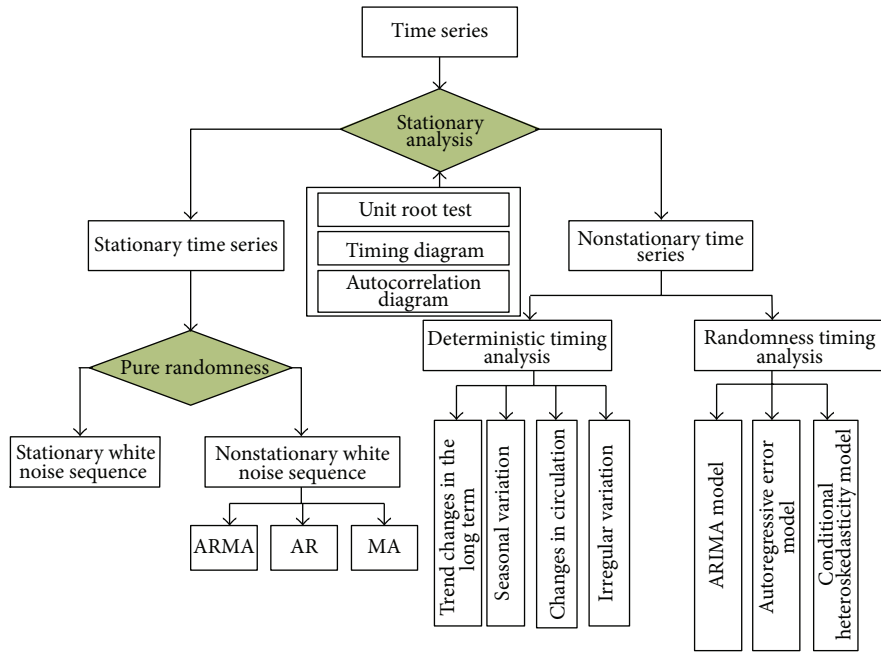


FIGURE 2: Classification of time series analysis.

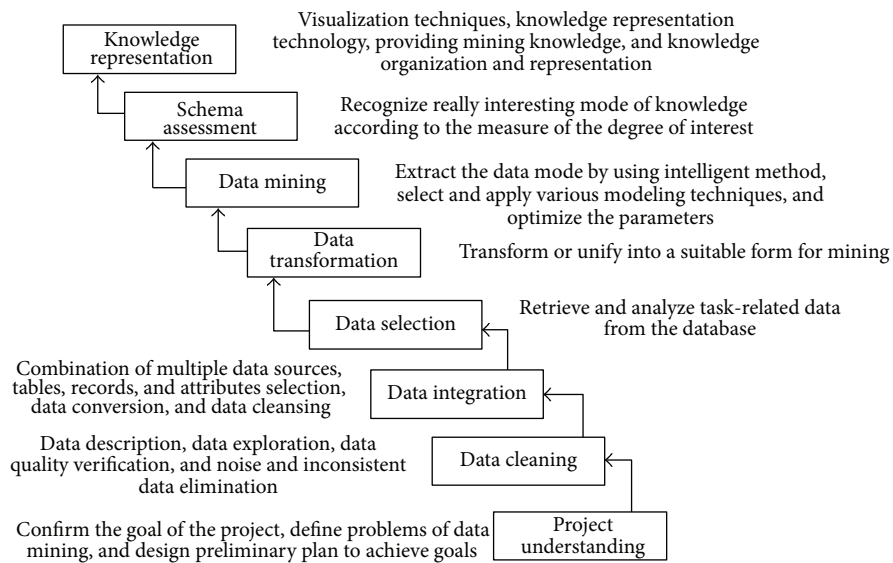


FIGURE 3: Steps of data mining process.

The definition of distance needs to meet the following four properties:

- (1)  $d_{ij} = 0 \Leftrightarrow x_i = x_j$ ,
- (2)  $\forall x_i, x_j, d_{ij} \geq 0$ ,
- (3)  $\forall x_i, x_j, d_{ij} = d_{ji}$ ,
- (4)  $\forall x_i, x_j, x_k, d_{ij} \leq d_{ik} + d_{kj}$ .

$d_{ij}$  is the distance between  $x_i$  and  $x_j$  in a group of objects  $D = \{x_1, x_2, \dots, x_n\}$  in  $m$ -dimensional space.

Minkowski distance is the common formula to calculate distance, and the expression is

$$d_{ij} = \sum_{k=1}^m (|x_{ik} - y_{jk}|^p)^{1/p}, \quad (p > 0). \quad (1)$$

Many distances are formed by changing the parameters of the Minkowski distance. Distance formula commonly used includes Manhattan distance, Euclidean distance, Chebyshev distance, custom distance, Mahalanobis distance, Minkowski distance, variance weighted distance, Canberra Distance,

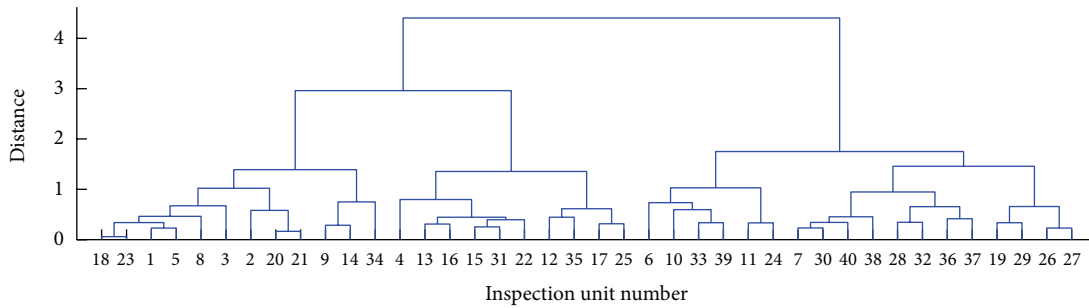


FIGURE 4: Visual cluster tree.

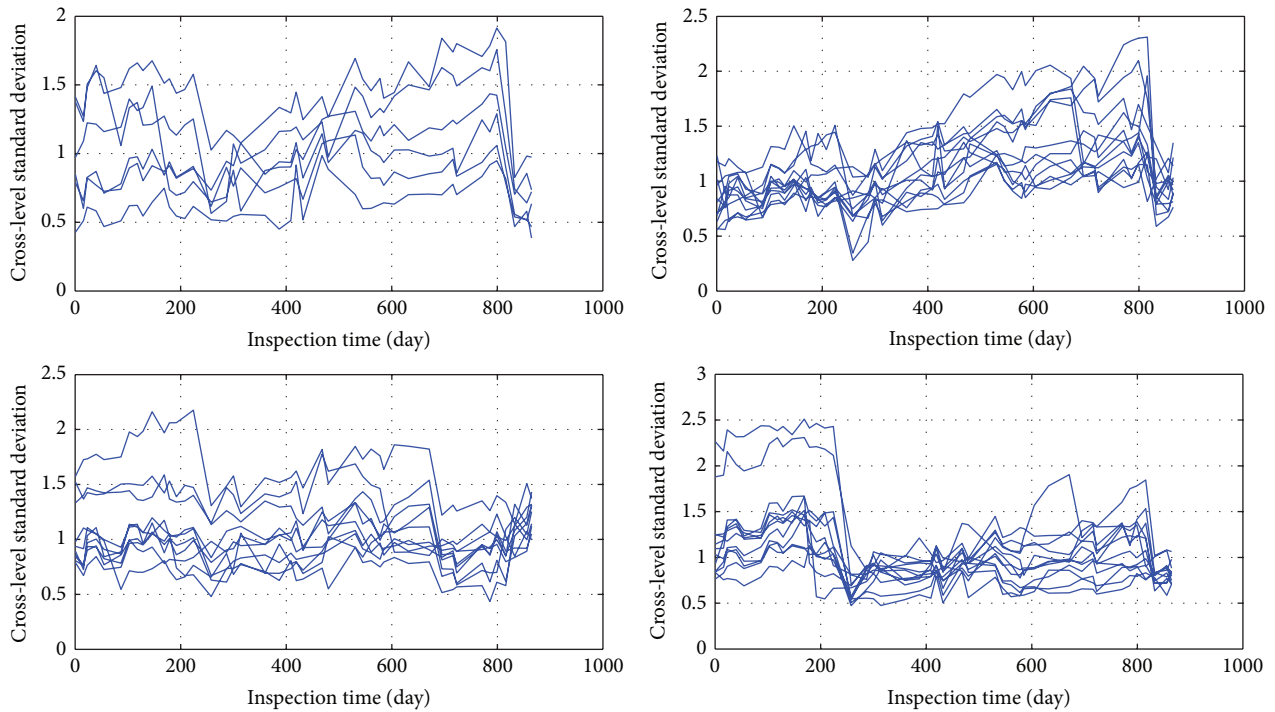


FIGURE 5: Clustering results.

dynamic time warping distance, the cross-correlation distance, KL distance, and so forth.

**3.3. Clustering Analysis of Track Irregularity.** Standard deviation between the data is used to reflect the degree of variation. When the two sets of data are under circumstances of the same units and similar means, the greater the standard deviation is, the greater the degree of variation between the data will be. When the data around the mean of the distribution is more discrete, representation of the mean is weaker. On the contrary, the smaller the standard deviation is, the little the variance between the data is indicated, and the distribution of data around the mean is more intensive, and representation of the mean is better. Therefore, standard deviation of track irregularity data is selected in the study of discrete distribution of track irregularity data over time to evaluate the development trends of track irregularity. In this study, track irregularity data is provided by State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University.

We take K449-K450 km section, Beijing-Kowloon line, with 44th cross-level inspection data as study data. The section is divided into 40 units, and the standard deviation of each units is calculated; then we get 40 cross-level standard deviation time-series data. This standard deviation data is the object of cluster analysis. In this paper,  $k$ -mean algorithm is used in clustering algorithm, cross correlation distance is used in the distance between objects in the matrix, and the minimum variance algorithm is used in the connection between the variables. Clustering results are shown in Figures 4 and 5.

The above track irregularity standard deviation time series which have similar clustering methods, changing trends, and characteristics are clustered into one category.

## 4. Linear Recursive Model

**4.1. Core Ideas.** Because of the inertia characteristics of the track state changes, track state has a memory effect. The latest

track state and the nearest previous state shares similarity, and the inspection state of the adjacent time points has a similar trend. From the macroperspective, the track state presents nonlinear changes throughout the whole life cycle of the track [37, 38], but from the microperspective, in a short time, track state changes at adjacent time will be close to linear features.

Based on the above assumptions, this paper proposes linear model of track state changes as follows:

$$s_{i,j} = k_i t_{i,j} + b_i. \quad (2)$$

In the formula,  $s_{ij}$  is the track state of the  $j$ th day between the  $i$ th to the  $(i+1)$ th inspection;  $k_i$  is slope value of track state linear change between the  $i$ th to the  $(i+1)$ th inspection;  $b_i$  is intercept;  $t_{i,j}$  is the  $j$ th day between the  $i$ th to the  $(i+1)$ th inspection.

According to the least square method, the model employs the vector form. Least squares estimation of the model parameters accords to the following conditions:

$$\hat{\alpha} = (B^T B)^{-1} B^T Y. \quad (3)$$

In the formula:

$$\hat{\alpha} = \begin{pmatrix} k_i \\ b_i \end{pmatrix},$$

$$Y = \begin{pmatrix} s_{i+1,0} \\ s_{i+2,0} \\ \vdots \\ s_{i+n,0} \end{pmatrix}, \quad (4)$$

$$B = \begin{pmatrix} t_{i+1} & 1 \\ t_{i+2} & 1 \\ \vdots & \vdots \\ t_{i+n} & 1 \end{pmatrix}.$$

In the formula,  $t_{i+n}$  is the length of time between the  $i$ th to the  $(i+1)$ th inspection;  $t_{i+n} = T_{i+n} - T_{i+n-1}$ ,  $T_{i+n}$  is the time of state value  $s_{i+n,0}$  at the  $(i+n)$ th inspection;  $s_{i+n,0}$  is state value of the  $(i+n)$ th inspection.

The form of recursive forecast of the model is as follows:

$$\begin{aligned} s_{i+m,j} &= k_{i+m} t_{i+m,j} + b_{i+m}, \\ s_{i+m+1,j} &= k_{i+m+1} t_{i+m+1,j} + b_{i+m+1}, \\ s_{i+m+2,j} &= k_{i+m+2} t_{i+m+2,j} + b_{i+m+2} \\ &\vdots \\ s_{i+T,j} &= k_{i+T} t_{i+T,j} + b_{i+T}. \end{aligned} \quad (5)$$

In the formula,  $m > n$ , and  $T > m + n$ .

**4.2. Model Analysis.** Three time intervals data of the first four historical inspection data and the second to fourth inspection data in cross-level standard deviation time series data are considered as an input; the slope and intercept of track

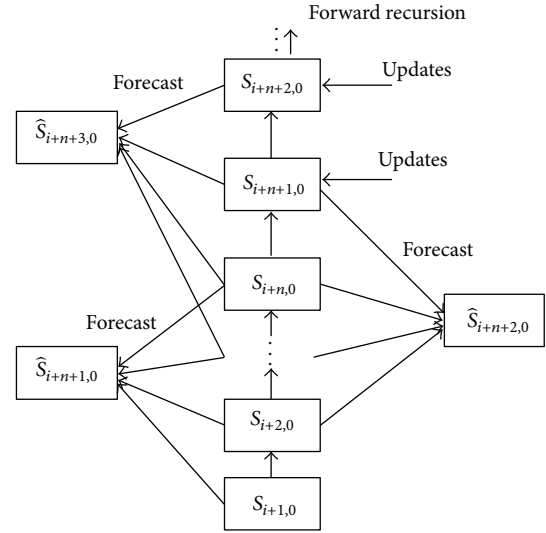


FIGURE 6: Forecast process of linear recursive model.

state linear changes in time interval between the fourth and fifth inspection are considered as output. When the forecast finishes, the fifth forecast data is updated by the fifth standard deviation of the actual inspection data, then three-time interval between the second and the fifth inspection data and the third to fifth inspection data are used to forecast the slope and intercept of track state linear changes in time interval between the fifth and sixth inspections. As the data is updated gradually, the model pushes and forecasts towards ahead.

Meanwhile, the linear forecast model proposed in this paper does not have to consider equal interval requirements of data samples like general time series models. In the track irregularity time series data, especially data of track inspection car, due to maintenance work, passenger and cargo plan changes, and the track inspection car scheduling and other reasons, inspection data is unequal interval, namely, in the model  $t_{i+n,0} \neq \text{const}$ . Linear model has the following advantages: simple model based on a linear model and least squares solution; without regard to unequal interval characteristics. Length of unequal interval is seen as a parameter directly in the model.

Forecast process of linear recursive model is shown in Figure 6.

**4.3. Residual Correction.** In order to further improve the forecast accuracy of the model, the residual needs to be corrected. Model residuals sequence data is a cyclical variation with concussive time series. In this paper, the Fourier transforming idea is used on residuals analysis.

Model residuals sequence data is as follows:

$$e_{i+k} = \hat{s}_{i+k} - s_{i+k}. \quad (6)$$

The Fourier transforming residuals expression  $e(t_i)$  is

$$\hat{e}_{i+k} = \frac{1}{2} h_0 + \sum_{j=1}^m [h_j \cos(\omega_j \cdot T_{i+k}) + l_j \sin(\omega_j \cdot T_{i+k})]. \quad (7)$$



In the formula,

$$\begin{aligned}
 i &= 1, 2, \dots, n, \\
 h_0 &= \frac{1}{n} \sum_{k=1}^n |e_{i+k}|, \\
 h_j &= \frac{2}{n} \sum_{k=1}^n e_{i+k} \cdot \cos(\omega_j \cdot T_{i+k}), \\
 l_j &= \frac{2}{n} \sum_{k=1}^n e_{i+k} \cdot \sin(\omega_j \cdot T_{i+k}), \\
 \omega_j &= \frac{2j\pi}{\sum_{k=1}^n t_{i+k}}.
 \end{aligned} \tag{8}$$

By looking for the appropriate value of  $m$ , fitting errors between model residuals  $e_{i+k}$  and periodic sequence  $\hat{e}_{i+k}$  achieved through Fourier transforming can get to a minimum value; namely,

$$\sum_{k=1}^n \frac{(\hat{e}_{i+k} - e_{i+k})^2}{e_{i+k}} = \min. \tag{9}$$

After getting the best value of  $m$ , the final forecast model is

$$\begin{aligned}
 s_{i+k} &= \hat{s}_{i+k} + \hat{e}_{i+k} \\
 &= \hat{s}_{i+k} + \frac{1}{2}h_0 \\
 &\quad + \sum_{j=1}^m [h_j \cos(\omega \cdot T_{i+k}) + l_j \sin(\omega \cdot T_{i+k})], \\
 i &= 1, 2, \dots, n.
 \end{aligned} \tag{10}$$

**4.4. A Case Study.** Change trend and periodic characteristic of track irregularity standard deviation data can be classified by cluster analysis. Since the cycle of track maintenance work is determined by track irregularity change, so it is uncertain. Thus, data containing multiple cycle cannot be used for the study. The data within a period that best represents the development trend of track irregularity characteristics is selected as data object. Change trends of track irregularities within each period can be forecasted by studying the law of track irregularity standard deviation changes in this cycle. So, 25 of track cross-level standard deviation data of Beijing-Kowloon line, K449 + 800 – K449 + 825 mileage unit, from November 13, 2008 to May 18, 2010, are taken as the study object to forecast 5–25th standard deviation data of track cross-level.

By solving the least squares estimation, we get recursion  $\{(k_i, b_i)\}$  as follows:  $\{(k_i, b_i)\} = \{(0.004012877, 1.052496244), (0.004880568, 1.072606649), (-0.000814172, 1.257042149), (0.003246744, 1.247075988), (-0.003644461, 1.423421705), (0.001636605, 1.30965592), (-0.001246396, 1.374786532), (0.009376521, 1.168171032), (0.001440352, 1.442494856), (-0.006255225, 1.647403617), (0.002275014, 1.502381251),$

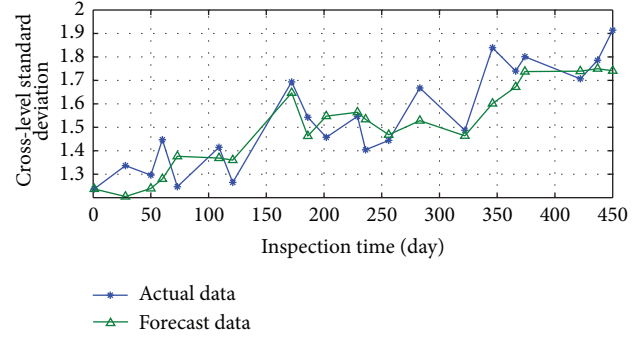


FIGURE 7: Forecast curve and the actual curve of cross-level standard deviation (K449 + 800 to K449 + 825 unit segment).

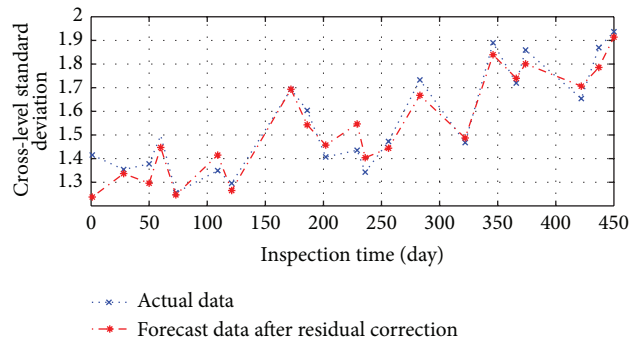


FIGURE 8: Forecast curve after residual revised and the actual curve of cross-level standard deviation.

$(-0.002533806, 1.551368672), (-0.000951367, 1.486815078), (0.007406862, 1.328357308), (-0.002678251, 1.568176542), (0.0066245, 1.442194457), (0.002152384, 1.629003922), (-0.001214053, 1.747250648), (-0.001427253, 1.807423072), (0.000826365, 1.737111301), (0.004952398, 1.676475089)\}$ .

When the parameters  $\{(k_i, b_i)\}$  are substituted into the short time track state changes linearly formula, the daily track state changes between two inspection time points can be estimated, and the forecast curve and the actual curve are shown in Figure 7.

Forecast results after residual correction are shown in Figure 8.

If we put several indicators together, we can fully measure the forecast accuracy of the model. Here, we use mean square error (MSE) and mean absolute percentage error (MAPE).

MSE is the expectation value of square of difference between the estimate value and true value of the parameter, and it indicates the degree of changes of data. The smaller the MSE value is, the more accurate the forecast model is in describing the experimental data. The expression of MSE is as follows:

$$\text{MSE} = \sqrt{\frac{\sum \varepsilon^2}{n}}. \tag{11}$$

TABLE 1: Analysis of accuracy of the cross level standard deviation.

Units	MSE	MAPE	MSE after residual correction	MAPE after residual correction
K449 + 000 – K449 + 025	0.11	12%	0.08	7.4%
K449 + 100 – K449 + 125	0.14	9.8%	0.07	5.0%
K449 + 200 – K449 + 225	0.09	6.4%	0.07	4.2%
K449 + 325 – K449 + 350	0.11	7.6%	0.06	3.8%
K449 + 425 – K449 + 450	0.08	6.0%	0.06	4.0%
K449 + 475 – K449 + 500	0.12	12%	0.07	7.5%
K449 + 575 – K449 + 600	0.10	7.7%	0.07	5.6%
K449 + 675 – K449 + 700	0.09	4.9%	0.05	2.3%
K449 + 800 – K449 + 825	0.10	5.5%	0.07	3.6%
K449 + 875 – K449 + 900	0.07	4.4%	0.04	2.0%
K449 + 975 – K450 + 000	0.09	7.4%	0.05	3.7%

Expression of MAPE is as follows:

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{z_t - \widehat{z}_t}{z_t} \right|. \quad (12)$$

Forecast accuracy of MAPE can be divided into four indicators: high-precision forecast (MAPE < 10%), good forecast (10% < MAPE < 20%), feasible forecast (20% < MAPE < 50%), and error forecast (MAPE > 50%).

Forecast accuracy of the model in some units segment is shown in Table 1.

Through the comparison of forecast accuracy indicators MSE and MAPE, we can find that the values of MSE and MAPE of the model after residuals correction are significantly reduced, and the forecast accuracy is generally improved by 30–40%, which belongs to high-precision forecast.

## 5. Linear-ARMA Model Based on Wavelet Decomposition Reconstruction

Forecast models after residual correction generally enjoy higher forecast accuracy than the original forecast models, but this also increases the computation and complexity of the model. Therefore, this paper will propose linear-ARMA model based on wavelet decomposition reconstruction.

**5.1. Wavelet Transform.** The wavelet transform [39–45] is a new field developing rapidly in applied mathematics and engineering. It is a new branch of mathematics and perfectly combines functional analysis, Fourier analysis, sample transfer analysis, and numerical analysis. It is based on certain special functions; it converts data process or data sequence into series in order to find the similar spectrum characteristics and finally achieves data processing. The wavelet transform is the local transformation of space (time) and frequency and can extract information effectively from the signal and do multiscale detailed analysis to functions or signals by dilation and translation and other computing functions.

As the name implies, “wavelet” means a waveform with a small area, limited length, and zero mean value. “Small” refers to the decay of the wavelet; while the “wave” refers to its volatility, its amplitude shocks in alternate positive and

negative forms. Compared to the Fourier transform, wavelet transform is the localized analysis of time (spatial) frequency; it eventually reaches breakdown of time at a high frequency; and subdivision of frequency at the low frequency, can automatically adapt to the requirements of time-frequency signal analysis, and can focus on any detail of the signal, solving the difficulties of Fourier transforming. Thus, it becomes a major breakthrough in the scientific method after the Fourier transforming. So, the wavelet transform is even called “mathematical microscope.”

In summary, the method that divides functions into a series of simple basic functions is of theoretical and practical significance. In this paper, Daubechies wavelet is used in track irregularity time series data decomposition. Daubechies is the general name of a series of binary wavelet proposed by the French scholar Daubechies, which can do multiscale wavelet decomposition to signal.

**5.2. Core Ideas.** According to the idea of wavelet decomposition and reconstruction, the standard deviation of track irregularity sequence data waveform signal is decomposed into detail waveform signal ( $D1, D2, D3$ ) and approximate waveform signal ( $A3$ ), in which detail waveform signal is stationary series with zero mean. We can use the random linear model to study it. We use ARMA model in this paper. Approximate waveform signals are generally nonstationary, smooth sequence curves. According to its smooth characteristics, the linear recursive model is used in trend analysis. The modeling idea is shown in Figure 9.

In the figure,  $A3_i$  is the actual value and  $\widehat{A3}_i, \widehat{A3}_{i+1}$  are forecast values. According to the geometric relationship, the values  $\widehat{A3}_i, \widehat{A3}_{i+1}$  are as follows:

$$\begin{aligned} \widehat{A3}_i &= A3_{i-1} + (A3_{i-1} - A3_{i-2}), \\ \widehat{A3}_{i+1} &= A3_i + (A3_i - A3_{i-1}). \end{aligned} \quad (13)$$

Finally, the two partial results are combined, and all the forecast sequence data are added up by weight 1. The formula is shown as follows:

$$\widehat{S} = \widehat{D1} + \widehat{D2} + \widehat{D3} + \widehat{A3}. \quad (14)$$

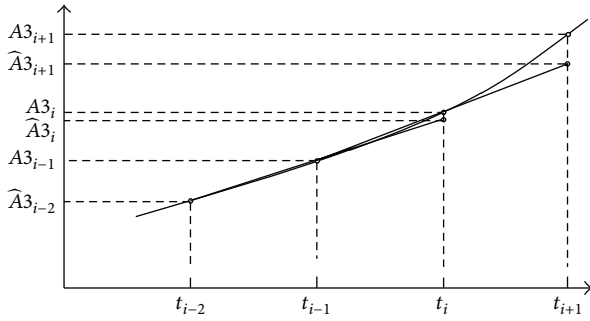


FIGURE 9: Linear recursive model of approximate waveform signal (LF).

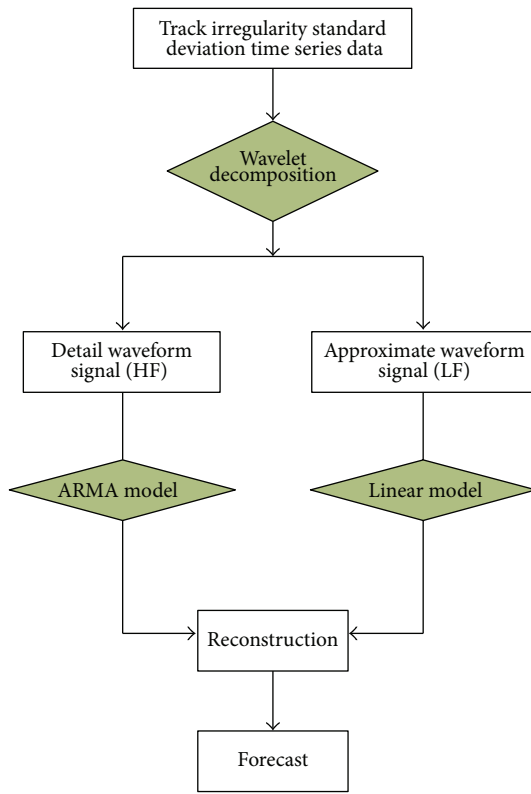


FIGURE 10: Modeling process of linear-ARMA model based on wavelet decomposition reconstruction.

In the formula,  $\widehat{D}1$ ,  $\widehat{D}2$ , and  $\widehat{D}3$  are high-frequency detail signal sequence forecasted by the ARMA model,  $\widehat{A}3$  is sequence of the low-frequency approximation by linear model forecast, and  $\widehat{S}$  is forecast value of the final state.

After decomposition-reconstruction process, the final forecast result appears, and the modeling process is shown in Figure 10.

5.3. Case Study. The Daubechies wavelet is selected to decompose the track irregularity standard deviation time series data signal; with the decomposition depth being 3, we use Mallat tower algorithm to do decomposition and

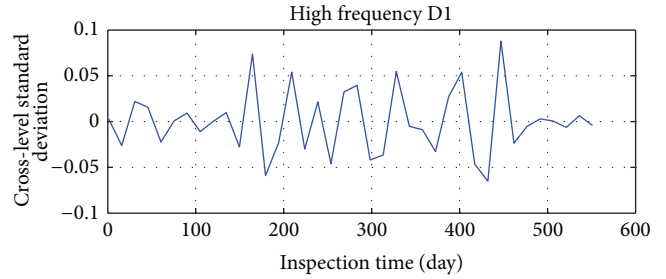


FIGURE 11: Level 1 detail waveform signal (HF) of cross-level standard deviation.

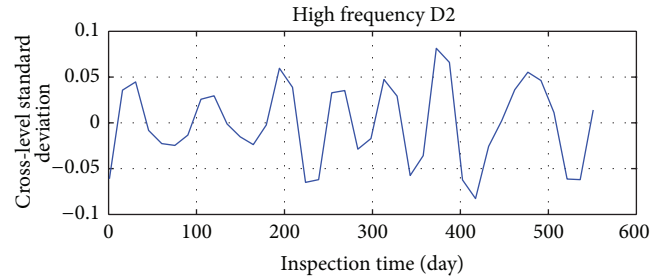


FIGURE 12: Level 2 detail waveform signal (HF) of cross-level standard deviation.

reconstruction of track irregularity standard deviation time series.

K449-450 km section, the Beijing-Kowloon line, upgoing, at 44th cross-level inspection data is selected as the study data. The section is divided into 40 units, and the standard deviation of each unit is calculated, and we get 40 cross-level standard deviation time-series data. After equal time intervals conversion of standard deviation time-series, we use linear-ARMA model based on wavelet decomposition and reconstruction for analysis.

Take K449 + 800 – K449 + 825 unit sections, from November 13, 2008, to May 18, 2010, with 38 cross-level standard deviation data as study data. The Linear-ARMA model process based on wavelet decomposition and reconstruction is as follows.

- (1) Wavelet decomposition of track cross-level standard deviation data: wavelet decomposition process of cross-level standard deviation time series is the process that divides cross-level standard deviation time series data into high frequency detail waveform signal and low frequency approximate waveform signal. The decomposed waveform signals are shown in Figures 11, 12, 13, and 14.
- (2) ARMA model forecast of high-frequency signals: high-frequency signal is stationary time series. We use ARMA model to forecast, and the results are shown in Figures 15, 16, and 17.
- (3) Low-frequency signal linear model forecast: low-frequency signal is a smooth curve, and its linear forecast result is shown in Figure 18.



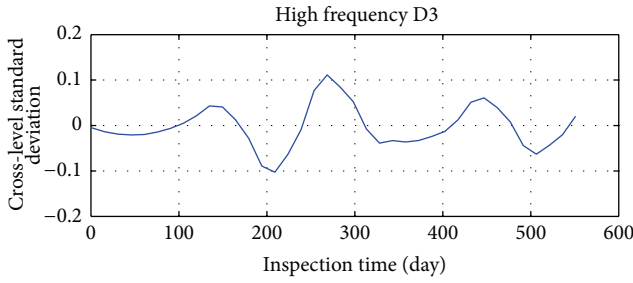


FIGURE 13: Level 3 detail waveform signal (HF) of cross-level standard deviation.

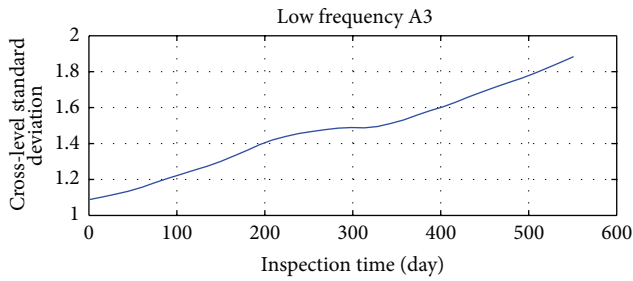


FIGURE 14: Layer 3 approximate series waveform signal (LF) of cross-level standard deviation.

(4) Model reconstruction and accuracy analysis: all forecast sequence data is added up by weight 1; comparison of forecast values and original values are shown in Figure 19.

Forecast error of the model is shown in Figure 20.

5.4. Accuracy Analysis. MSE and MAPE are used to measure accuracy of the model, and the forecast the accuracy of some units is shown in Table 2.

According to the forecast results of MSE, MAPE value in Table 2, the model has higher forecast accuracy, and so there is no need to correct the residual like linear recursive model. This has shown that linear-ARMA model based on wavelet decomposition reconstruction is an effective way to forecast the trend of track state changes.

## 6. Conclusions

In this paper, data mining and time series theory are used to study track irregularity standard deviation time series data. The main purpose of this study is to forecast track irregularity state in future. By using clustering analysis theory in data mining, different patterns and characteristics of track irregularity change can be found. Through a systematic study of time series data classification and time series forecast model, this paper puts forward linear recursive models and linear-ARMA model based on wavelet decomposition reconstruction to forecast the changing trends of track irregularity standard deviation time series. Simulation results show that the models have higher accuracy.

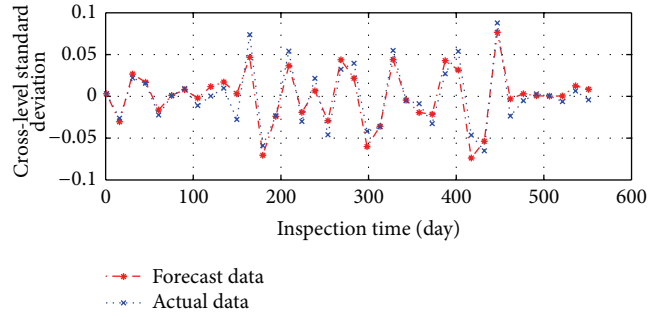


FIGURE 15: Level 1 forecast value of detail waveform signal (HF) and original data cross-level standard deviation.

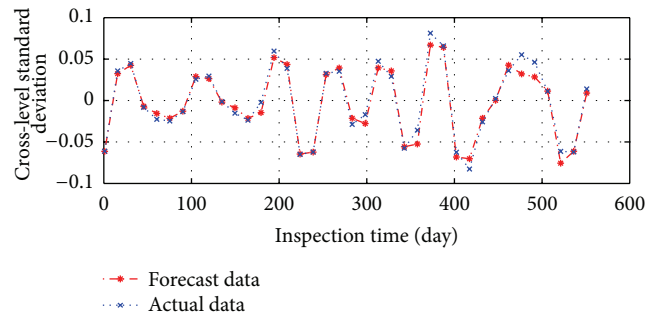


FIGURE 16: Level 2 forecast value of detail waveform signal (HF) and original data cross-level standard deviation.

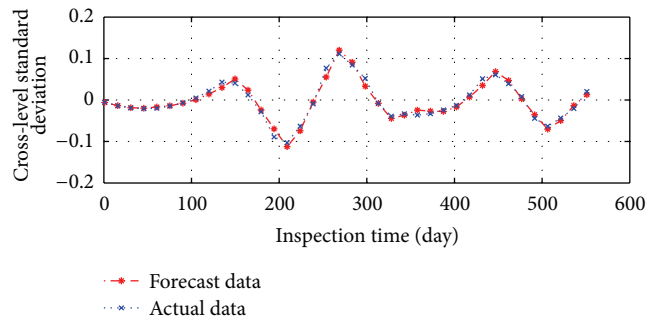


FIGURE 17: Level 3 forecast value of detail waveform signal (HF) and original data cross-level standard deviation.

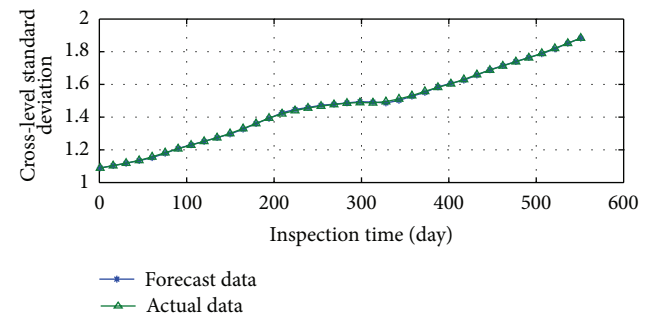


FIGURE 18: Layer 3 forecast value of approximation waveform signal (LF) and original data of cross-level standard deviation.

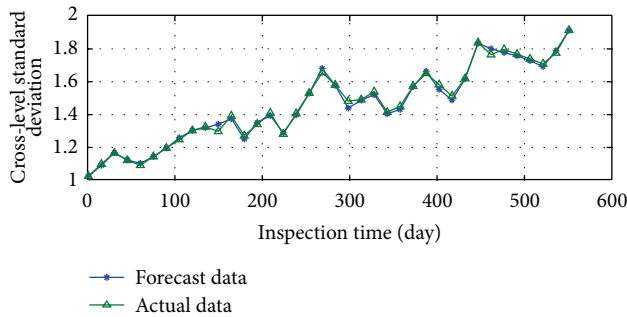


FIGURE 19: Comparison of forecast values and original values.

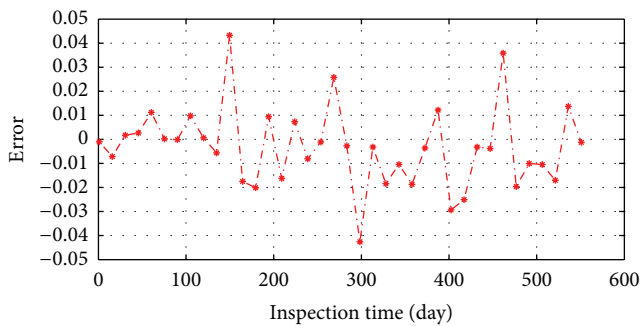


FIGURE 20: Error analysis.

TABLE 2: Model accuracy.

Units	MSE	MAPE
K449 + 000 – K449 + 025	0.03	3.5%
K449 + 100 – K449 + 125	0.02	2.0%
K449 + 200 – K449 + 225	0.03	2.1%
K449 + 325 – K449 + 350	0.02	2.0%
K449 + 425 – K449 + 450	0.02	1.8%
K449 + 475 – K449 + 500	0.03	2.8%
K449 + 575 – K449 + 600	0.03	1.9%
K449 + 675 – K449 + 700	0.03	1.1%
K449 + 800 – K449 + 825	0.02	0.8%
K449 + 875 – K449 + 900	0.02	1.3%
K449 + 975 – K450 + 000	0.02	1.7%

The change of railway track state is a complex process. It is affected by various aspects of the situation. Although it is an extremely difficult task to explore laws of its development and changes, the significance of the study is far reaching. Because track state is inspected by section, it is not carried out for each fixed measuring points, although there is mileage data information of fixed measuring points to be studied in the inspection data. However, it is unavoidable that there is mileage deviation between the actual measuring point's mileage data information and measuring point's mileage data information in inspection data, and the mileage deviation will be a few or dozens of meters. It is the basic idea in the study of track state changes that puts the data after mileage relative calibration as the subject or takes the section as a whole as

the subject. Due to track maintenance and repair cycle, we only study track state changes in a cycle. However, track state change trend between each of the maintenance and repair cycle is also worth to be studied.

## 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), National Key Technology R&D Program (Grant no. 2009BAG12A10), independent subject of State Key Laboratory of Rail Traffic Control and Safety, and Beijing Jiaotong University (Contract no. RCS2009ZT007) and partially supported by the MOE key Laboratory for Transportation Complex Systems Theory and Technology School of Traffic and Transportation, Beijing Jiaotong University.

## References

- [1] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.
- [2] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 23, no. 1, pp. 67–72, 1975.
- [3] J. B. Kruskal and M. Liberman, *The Symmetric Time Warping Algorithm: From Continuous to Discrete, Time Warps*, Addison-Wesley, Reading, Mass, USA, 1983.
- [4] C. Myers, L. R. Rabiner, and A. E. Rosenberg, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 6, pp. 623–635, 1980.
- [5] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," in *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, pp. 69–84, Chicago, USA, 1993.
- [6] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London A*, vol. 226, pp. 267–298, 1927.
- [7] G. Walker, "On periodicity in series of related terms," *Proceedings of the Royal Society of London A*, vol. 131, no. 818, pp. 518–532, 1931.
- [8] E. J. Hannan, *Multiple Time Series. Wiley Series in Probability and Mathematical Statistics*, John Wiley and Sons, New York, NY, USA, 1970.
- [9] P. Whittle, *Prediction and Regulation by Linear Least-Square Methods*, University of Minnesota Press, Minneapolis, Minn, USA, 1983.

- [10] E. J. Hannan, M. Deistler, and John Wiley and Sons, *Statistical Theory of Linear Systems. Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, New York, NY, USA, 1988.
- [11] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, NC, USA, 1970.
- [12] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.
- [13] J. Durbin and S. Koopman, *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, UK, 2001.
- [14] P. J. Harrison and C. F. Stevens, "Bayesian forecasting," *Journal of the Royal Statistical Society B*, vol. 38, no. 3, pp. 205–247, 1976.
- [15] P.-T. Chang and E. S. Lee, "Fuzzy linear regression with spreads unrestricted in sign," *Computers and Mathematics with Applications*, vol. 28, no. 4, pp. 61–70, 1994.
- [16] P. Pai and C. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.
- [17] O. Castillo and P. Melin, "Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1395–1408, 2002.
- [18] Y. H. Tay, M. Khalid, R. Yusof, and C. Viard-Gaudin, "Offline cursive handwriting recognition system based on hybrid Markov model and neural networks," in *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation*, vol. 3, pp. 1190–1195, 2003.
- [19] R. Krzysztofowicz, "A Bayesian Markov model of the flood forecast process," *Water Resources Research*, vol. 19, no. 6, pp. 1455–1465, 1983.
- [20] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [21] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Journal of Basic Engineering*, vol. 83, no. 1, pp. 95–107, 1961.
- [22] I. Sugiarto and S. Natarajan, "Parameter estimation using least square method for MIMO Takagi-Sugeno neuro-fuzzy in time series forecasting," *Jurnal Teknik Elektro*, vol. 7, no. 2, pp. 82–87, 2007.
- [23] M. Zounemat-Kermani and M. Teshnehlab, "Using adaptive neuro-fuzzy inference system for hydrological time series prediction," *Applied Soft Computing Journal*, vol. 8, no. 2, pp. 928–936, 2008.
- [24] H. J. Rong, N. Sundararajan, G. B. Huang, and P. Saratchandran, "Sequential Adaptive Fuzzy Inference System (SAFIS) for nonlinear system identification and prediction," *Fuzzy Sets and Systems*, vol. 157, no. 9, pp. 1260–1275, 2006.
- [25] S. Paul and S. Kumar, "Subsethood-product fuzzy neural inference system (SuPFuNIS)," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 578–599, 2002.
- [26] S. D. Balkin and J. K. Ord, "Automatic neural network modeling for univariate time series," *International Journal of Forecasting*, vol. 16, no. 4, pp. 509–515, 2000.
- [27] Y. Chen, B. Yang, J. Dong, and A. Abraham, "Time-series forecasting using flexible neural tree model," *Information Sciences*, vol. 174, no. 3–4, pp. 219–235, 2005.
- [28] F. Giordano, M. La Rocca, and C. Perna, "Forecasting nonlinear time series with neural network sieve bootstrap," *Computational Statistics and Data Analysis*, vol. 51, no. 8, pp. 3871–3884, 2007.
- [29] R. Adhikari and R. K. Agrawal, "Forecasting strong seasonal time series with artificial neural networks," *Journal of Scientific & Industrial Research*, vol. 71, no. 10, pp. 657–666, 2012.
- [30] G. Armano, M. Marchesi, and A. Murru, "A hybrid genetic-neural architecture for stock indexes forecasting," *Information Sciences*, vol. 170, no. 1, pp. 3–33, 2005.
- [31] S. M. Chen and N. Y. Chung, "Forecasting enrollments using high-order fuzzy time series and genetic algorithms," *International Journal of Intelligent Systems*, vol. 21, no. 5, pp. 485–501, 2006.
- [32] Z. J. Zhou and C. H. Hu, "An effective hybrid approach based on grey and ARMA for forecasting gyro drift," *Chaos, Solitons and Fractals*, vol. 35, no. 3, pp. 525–529, 2008.
- [33] L. Yu, S. Wang, and K. K. Lai, "A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates," *Computers and Operations Research*, vol. 32, no. 10, pp. 2523–2541, 2005.
- [34] H. Kim and K. Shin, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets," *Applied Soft Computing Journal*, vol. 7, no. 2, pp. 569–576, 2007.
- [35] H. Guo, W. Wang, W. Guo, X. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [36] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [37] C. Jia, W. Xu, F. Wang, and H. Wang, "Track irregularity time series analysis and trend forecasting," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 387857, 15 pages, 2012.
- [38] H. Wang, W. Xu, F. Wang, and C. Jia, "A cloud-computing-based data placement strategy in high-speed railway," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 396387, 15 pages, 2012.
- [39] C. Cattani and A. Kudreyko, "Harmonic wavelet method towards solution of the Fredholm type integral equations of the second kind," *Applied Mathematics and Computation*, vol. 215, no. 12, pp. 4164–4171, 2010.
- [40] S. V. Narasimhan, M. Harish, A. R. Haripriya, and N. Basumallick, "Discrete cosine harmonic wavelet transform and its application to signal compression and subband spectral estimation using modified group delay," *Signal, Image and Video Processing*, vol. 3, no. 1, pp. 85–99, 2009.
- [41] H. A. Yousef, M. E. Elkhatab, and O. A. Sebakhy, "Wavelet network-based motion control of DC motors," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1522–1527, 2010.
- [42] M. Seonguk and K. Min, "Detection of combustion start in the controlled auto ignition engine by wavelet transform of the engine block vibration signal," *Measurement Science and Technology*, vol. 19, no. 8, Article ID 085407, pp. 6686–6696, 2008.
- [43] H. Janicke, M. Bottinger, U. Mikolajewicz, and G. Scheuermann, "Visual exploration of climate variability changes using wavelet analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1375–1382, 2009.
- [44] N. Kazantzis, K. T. Chong, J. H. Park, and A. G. Parlos, "Control-relevant discretization of nonlinear systems with time-delay using Taylor-Lie series," *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, vol. 127, no. 1, pp. 153–159, 2005.

- [45] E. Hossaini-asl and M. Shahbazian, "Nonlinear dynamic system control using wavelet neural network based on sampling theory," in *Proceedings IEEE International Conference on Systems, Man and Cybernetics (SMC '09)*, pp. 4502–4507, October 2009.





# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

