Physica A 492 (2018) 639–650

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A better understanding of long-range temporal dependence of traffic flow time series

Shuo Feng [a], Xingmin Wang [c], Haowei Sun [a], Yi Zhang [a,b], Li Li [a,*]

[a] *Department of Automation, Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, China*
[b] *Tsinghua – Berkeley Shenzhen Institute (TBSI), Tower C2, Nanshan Intelligence Park 1001, Xueyuan Blvd., Nanshan District, Shenzhen 518055, China*
[c] *Department of Automotive Engineering, Tsinghua University, Beijing 100084, China*

HIGHLIGHTS

- Both the daily and long-range temporal dependence exert considerable influence on the traffic flow series.
- The daily temporal dependence creates crossover phenomenon when estimating the Hurst.
- PCA-based method turns out to be a better method to extract the daily temporal dependence.

ARTICLE INFO

Article history:
Received 1 June 2017
Received in revised form 22 August 2017
Available online 21 November 2017

Keywords:
Long-range temporal dependence
Daily temporal dependence
Traffic flow time series

ABSTRACT

Long-range temporal dependence is an important research perspective for modelling of traffic flow time series. Various methods have been proposed to depict the long-range temporal dependence, including autocorrelation function analysis, spectral analysis and fractal analysis. However, few researches have studied the daily temporal dependence (i.e. the similarity between different daily traffic flow time series), which can help us better understand the long-range temporal dependence, such as the origin of crossover phenomenon. Moreover, considering both types of dependence contributes to establishing more accurate model and depicting the properties of traffic flow time series. In this paper, we study the properties of daily temporal dependence by simple average method and Principal Component Analysis (PCA) based method. Meanwhile, we also study the long-range temporal dependence by Detrended Fluctuation Analysis (DFA) and Multifractal Detrended Fluctuation Analysis (MFDFA). The results show that both the daily and long-range temporal dependence exert considerable influence on the traffic flow series. The DFA results reveal that the daily temporal dependence creates crossover phenomenon when estimating the Hurst exponent which depicts the long-range temporal dependence. Furthermore, through the comparison of the DFA test, PCA-based method turns out to be a better method to extract the daily temporal dependence especially when the difference between days is significant.

* Corresponding author.
E-mail addresses: s-feng14@mails.tsinghua.edu.cn (S. Feng), wxm14@mails.tsinghua.edu.cn (X. Wang), sun-hw15@mails.tsinghua.edu.cn (H. Sun), zhyi@tsinghua.edu.cn (Y. Zhang), li-li@mail.tsinghua.edu.cn (L. Li).

https://doi.org/10.1016/j.physa.2017.10.006
0378-4371/

1. Introduction

Traffic flow time series analysis is a very important part of most Intelligent Transportation Systems (ITS) research and applications [1]. As pointed out in [2], modelling of traffic flow time series is very crucial to traffic flow prediction, missing traffic data imputation, traffic data compression and abnormal traffic data detection.

Traffic flow time series is so complicated that a significant amount of work has been proposed to describe different perspectives of traffic flow time series. For example, Williams et al. [3] assumed that weekly-periodic exists in traffic flow time series and can be extracted by seasonal autoregressive integrated moving average (ARIMA) process. Vlahogianni et al. [4] proposed a multilayer strategy which identifies patterns of traffic based on their structure and evolution in time. Chen et al. [5] assumed that traffic flow time series is affected by internal and external factors which can be modelled respectively. In [6], it assumed intra-day trend exists and can be extracted by simple average method, PCA-based method and wavelet based methods. Li et al. [7] utilized the information of temporal and spatial dependence of traffic flow to improve the efficiency of missing data imputing methods. Ran et al. [8] proposed a tensor based method modelling traffic flow time series into high-dimensional matrices. Cheng et al. [9] proposed a modelled traffic flow time series by chaos theory and promoted the performance for short-term traffic flow prediction. Feng et al. [10,11] extracted the daily trend of traffic flow time series by PCA, Kronecker Product and tensor based methods and improved the performance of data compression.

In all the research perspectives mentioned above, long-range temporal dependence in traffic flow time series receive special attentions. Generally, temporal dependence relates to the rate of decay of statistical dependence of two points with increasing time interval. A phenomenon is usually considered to have long-range dependence if the dependence decays more slowly than an exponential decay, typically a power-like decay. Long-range temporal dependence has been found in various fields. For example, Bunde et al. [12] found that the persistence, characterized by the correlation of temperature variations separated by days, has the long-range temporal dependence. Plerou et al. [13] found the long-range temporal dependence in financial time series. Peng et al. [14] found long-range temporal dependence in intron-containing genes and in nontranscribed regulatory DNA sequences. In contrast, some other time series (e.g. the output of a Markov process) do not behave long-range temporally dependence. Various methods have been proposed to depict the long-range temporal dependence, such as autocorrelation function analysis, spectral analysis and fractal analysis.

In [15], autocorrelation function analysis is proposed to distinguish between short- and long-range correlated dependence. The autocorrelation function of a time series can be expressed as

$$C(s) = \frac{E[(x_i - \bar{x})(x_{i+s} - \bar{x})]}{\sigma^2}, \tag{1}$$

where E means the mathematical expectation, $\bar{x}$ is the average of x and σ is the variance. A time series is defined to be short-range correlated if the autocorrelation function declines exponentially and long-range correlated if the autocorrelation function declines as a power-law, $C(s) \propto s^{-\gamma}$, with a correlation exponent $0 < \gamma < 1$. As shown in many studies, traffic flow time series is long-range correlated to a large extent. For example, Stathopoulos et al. [16] utilized partial autocorrelation function to model the long-range temporal dependence of traffic flow time series to promote the short-term prediction.

Spectral analysis is another tool to identify long-range temporal dependence of time series in frequency domain. As shown in [17], we can apply spectral analysis techniques (Fourier transform) and then calculate the power spectrum $S(f)$ of the time series as a function of the frequency to obtain scaling behaviour of temporal dependence. For long-range correlated data characterized by the correlation exponent $0 < \gamma < 1$, we have

$$S(f) \propto f^{-\beta} \quad with \quad \beta = 1 - \gamma, \tag{2}$$

which can be derived from the Wiener–Khinchin theorem, as discussed in [18]. In [19], it is observed that the fluctuation of a traffic current on an expressway obeys the $1/f$ law for low spectral frequencies. We say a time series obeys $1/f$ law when its power spectral density is proportional to the inverse of frequency. Nassab et al. [20] studied the $1/f$ law for traffic flow with open boundaries and additional connection sites.

A fractal series refers to a series that can be characterized by a scaling law with a fractal. Hurst exponent, introduced by Hurst [21], is widely used to analyse a fractal series. There are many methods to estimate the Hurst exponent [22] during which fluctuation analysis (FA) is one of the most commonly used [14]. As shown in [23], traffic flow time series is non-stationary, so Detrended Fluctuation Analysis (DFA) is usually used to analyse the traffic flow time series [24]. Compared to FA, DFA removes a local polynomial trend of time series to eliminate the influence of non-stationary. More details about the difference between DFA and FA can be found in [25]. There are many studies on long-range temporal dependence of traffic flow time series by Hurst exponent. Shang et al. [26] applied the generalized Hurst exponent to traffic congestion warning. Li et al. [22] proposed a new traffic flow model to explain the crossover phenomena of Hurst exponent.

There are internal relations between the methods mentioned above. As proven in [27], the relationship among Hurst exponent, spectral analysis and autocorrelation function analysis for mono-fractal time series is

$$2H = 1 + \beta = 2 - \gamma, \tag{3}$$

where H is Hurst exponent, γ is correlation exponent and β is calculated by Eq. (2). Mono-fractal time series refers to a series that follows a scaling law with a single fractal exponent. Therefore, all these methods depict the long-range correlated temporal dependence for mono-fractal time series.

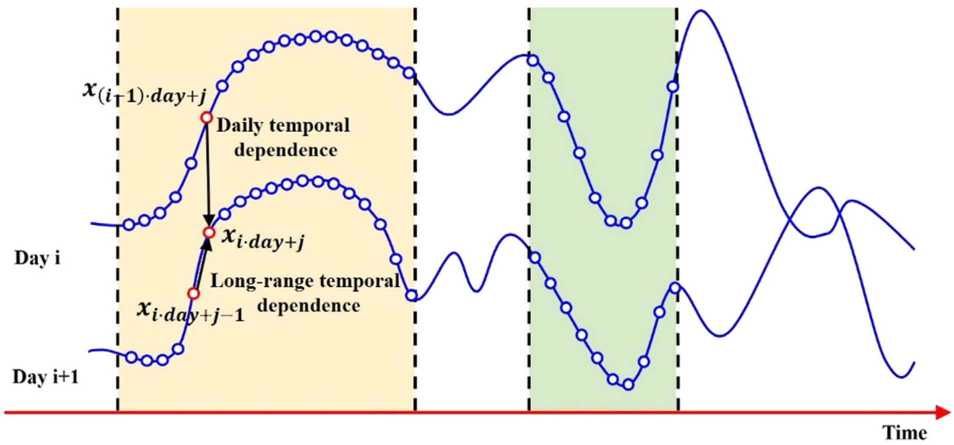


Fig. 1. An illustration of daily dependence of traffic flow time series. The traffic flow point, $x_{i\cdot day+j}$, where *day* represents the number of the points of single day, is influenced by the last moment point $x_{i\cdot day+j-1}$ and the last day point $x_{(i-1)\cdot day+j}$, which correspond to the long-range temporal dependence and the daily-temporal dependence respectively.

Although previous researches aimed at enhancing our understanding of traffic flow time series from the perspective of long-range temporal dependence, few of them have studied the daily temporal dependence. Daily temporal dependence is proposed to depict the relationship between different daily traffic flow time series. As pointed in [2], we can say two daily time series collected from the same site have daily dependence if for example, the large value of Time Series 1 (Day i in Fig. 1) indicates that the value of Time Series 2 (Day $i+1$ in Fig. 1) tends to be large (or small), too. In detail, as shown in Fig. 1, the traffic flow point, $x_{i\cdot day+j}$, where *day* represents the number of the points of single day, is influenced by both the last time point $x_{i\cdot day+j-1}$ and the last day point $x_{(i-1)\cdot day+j}$, which correspond to the long-range and daily temporal dependence respectively.

There are some approaches which implicitly or explicitly utilize the daily temporal dependence of traffic flow time series, especially in traffic flow prediction research. One typical approach is to use weekly seasonal differences to improve prediction performance [28,29]. It was asserted that a 1-week lagged first seasonal difference applied to the traffic flow series yields a stationary weekly transformation. Another approach is to retrieve the common daily-periodic trend which is shared by the consecutive days. There are various definitions for this type of common daily-periodic trend, including weekly-average/monthly-average trend [30,31], online regression [32], principle component analysis [11,33], cluster analysis and wavelet methods [34], etc.

However, there are few studies considering the co-existence of the long-range temporal dependence and the daily temporal dependence of the traffic flow series. In this paper, we aim to study the properties of daily temporal dependence and comparison between the two different dependence of traffic flow time series. Specifically, we study the daily temporal dependence by retrieving daily trend based on simple average method and PCA method. Meanwhile, we study the long-range temporal dependence by DFA and Multifractal Detrended Fluctuation Analysis (MFDFA) [35]. Then we compare the two kinds of temporal dependence by the crossover phenomenon and the correlations of different series, where linear regression is used to measure the effects of the two types of dependences. The flowchart of the research in this paper is shown in Fig. 2.

By numerical experiments using data from the publicly accessible PeMS dataset, we find that both the daily temporal dependence and the long-range dependence exert considerable influence on the traffic flow series. Besides, we also conduct different DFA tests by utilizing different methods of removing the daily temporal dependence. The result shows that the crossover phenomenon can be eliminated by removing the daily temporal dependence and PCA turns out to be a more effective way than simply removing the average. Finally, the MFDFA also leads to the same conclusion.

The rest of this paper is organized as follows. Section 2 explains the two kinds of models for temporal dependence of traffic flow time series. Data description and the experimental tests are presented in Section 3. Finally, the conclusions are summarized in Section 4.

2. Two kinds of models for temporal dependence of traffic flow time series

2.1. Models for long-range temporal dependence

Hurst exponent is commonly used to characterize the long-range temporal dependence. There are many methods to estimate the Hurst exponent, such as autocorrelation function analysis, spectral analysis and fluctuation analysis (FA), all of which assume the data to be stationary. However, as pointed out in [23], traffic flow time series are generally non-stationary for both mean and variance of the series.

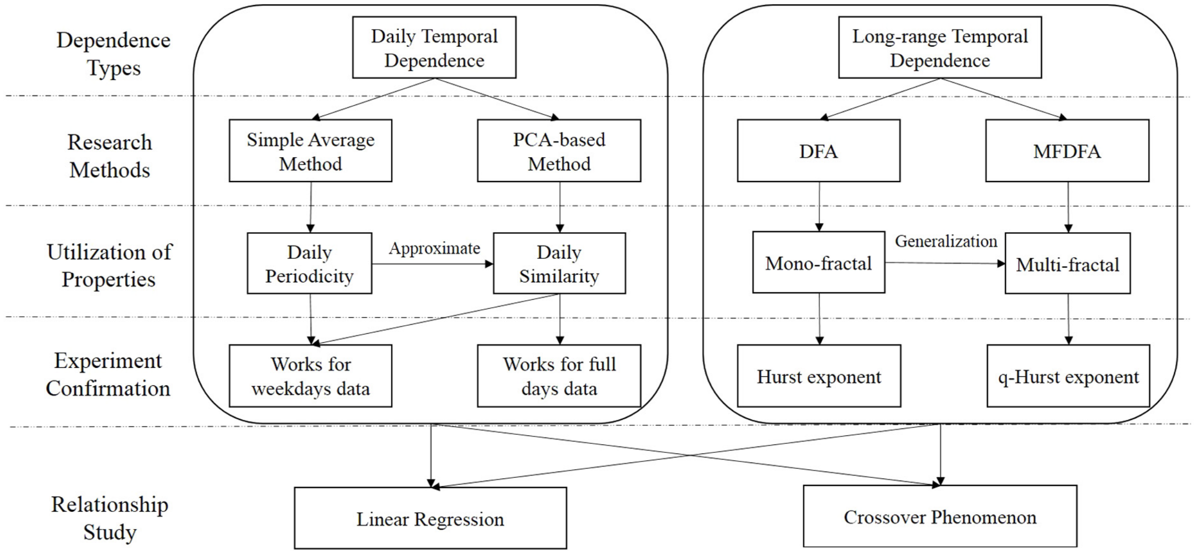


Fig. 2. The flowchart of the research in this paper.

Therefore, the Detrended Fluctuation Analysis (DFA) is used to analyse long-range temporal dependence of traffic flow time series in this paper. DFA, originally introduced by Peng et al. [14], has been established as an important method to detect long-range dependence in non-stationary time series. The method is based on random walk theory and basically represents a linear or higher order detrending version of FA.

The DFA method consists of four steps [36]:

(a) Determine the profile:

$$y(i) = \sum_{k=1}^{i} x_k - < x > \quad i = 1, \ldots, N. \tag{4}$$

(b) Cut the profile y(i) into N non-overlapping segments of equal length s. To take full use of the all data, the same procedure is repeated starting from the end of the series. Thus, $2N_s$ segments are obtained altogether.Y_s (i) as the difference:

$$Y(i) = y(i) - y^{fit}(i) \tag{5}$$

y^{fit} is the fitting polynomial in the νth segment. Generally, linear, quadratic, cubic or higher order can be used to the fitting procedure (DFA1, DFA2, DFA3, and higher order DFA). These methods differ in their capability of eliminating trends in data.

(c) Calculate the variance for each of the $2N_s$ segments

$$F_s^2(\nu) =< Y_s^2(i) >= \frac{1}{s}\sum_{i=1}^{s} Y^2[(\nu-1)\cdot s + i] \tag{6}$$

of the detrended time series $Y_s(i)$ by averaging over all data points i in the νth segments.

(d) At last, we average all segments and take the square root to obtain the DFA fluctuation function

$$F(s) = \sqrt{\frac{1}{2N_s}\sum_{\nu=1}^{2N_s} F^2(\nu)}. \tag{7}$$

We are interested in the s-dependence of $F^{(n)}(s)$. If the data are long-range power-law correlated, the fluctuation functions $F^{(n)}(s)$ increase by a power-law:

$$F^{(n)}(s) \propto s^{\alpha}, \tag{8}$$

where $\alpha \approx H$ is the Hurst exponent.

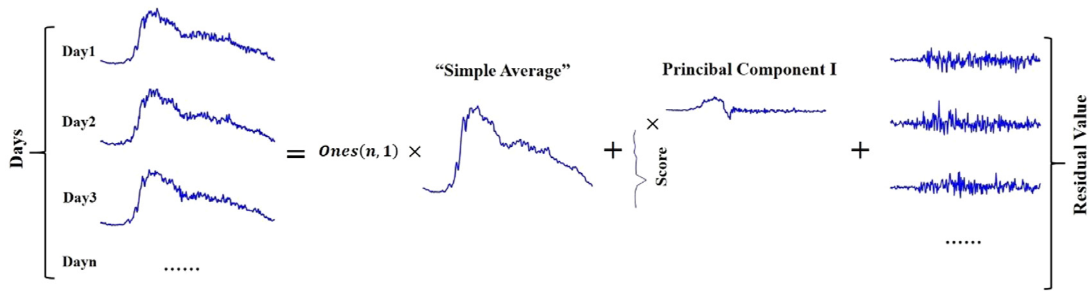


Fig. 3. An illustration of the PCA method. The traffic flow series is stored in an N-row matrix, where each row stores a single day's traffic flow data.

However, many records do not exhibit a simple mono-fractal scaling behaviour, which can be accounted for by a single scaling exponent. Multifractal Detrended Fluctuation Analysis (MFDFA), the expansion of the DFA, has been successfully applied to different fields. The difference between DFA and MFDFA derives form equation Eq. (7). In MFDFA, it changes to

$$F(s) = \sqrt[q]{\frac{1}{2N_s}\sum_{\nu=1}^{2N_s} F^q(\nu)}. \tag{9}$$

Then the scaling law can be expressed as

$$F_q(s) \propto s^{h(q)}. \tag{10}$$

The $h(q)$ is called generalized Hurst exponent which is independent only when the time series is mono-fractal. In general, if small and large fluctuations scale differently, there will be a significant dependence of $h(q)$ on q.

2.2. Models for daily temporal dependence

In this paper, we retrieve daily trend by simple average method and PCA-based method to study the influence of daily temporal dependence.

2.2.1. Simple average method

Suppose that the traffic time series sampled in N consecutive work days can be written as a series of one-dimensional vectors

$$\boldsymbol{x_1} = [x_{1,1}, x_{2,1}, \ldots, x_{n,1}]^T, \ldots, \boldsymbol{x_N} = [x_{1,N}, x_{2,N}, \ldots, x_{n,N}]^T. \tag{11}$$

If the sample time interval is 5 min, then we have $n = 288$. The simple average daily trend over the last N working days can be calculated as

$$\boldsymbol{x_{Average}} = [\frac{1}{N}\sum_{j=1}^{N} x_{1,j}, \ldots, \frac{1}{N}\sum_{j=1}^{N} x_{n,j}]. \tag{12}$$

2.2.2. PCA-based method

PCA is a type of eigenvector-based multivariate analysis that identifies common patterns in data and highlights the similarities as well as differences [37]. It will keep the main features of the traffic flow time series and discard the insignificant ones. Different from the simple average trend, the PCA-based daily trend can vary between different days. In factor, before extracting the principal components, the PCA method have removed the simple average in the first step.

Suppose that the PCA-based daily trend $\bar{Y}_{PCA}(t)$ is written as

$$\bar{Y}_{PCA} = [\bar{Y}_{N-L+1}(t), \ldots, \bar{Y}_N(t)]. \tag{13}$$

As shown in [37] and [10], PCA analysis determines a latent feature space with lower dimensions to approximately represent $Y_{n,L}Y_{n,L}^T$ as $Y_{n,L}Y_{n,L}^T \approx \tilde{Y}_{n,L}\tilde{Y}_{n,L}^T$, where $\tilde{Y}_{n,L} = [\bar{Y}_{N-L+1}(t)^T, \ldots, \bar{Y}_N^T]$. Here, $Y_{n,L}(t) = [Y_{N-L+t}(t)^T, \ldots, Y_N(t)^T]$ consists of L columns of the empirical traffic flow series from the last L days. $\tilde{Y}_{n,L}$ will be reconstructed from the first few eigenvalues/vectors of the covariance matrix $Y_{n,L}Y_{n,L}^T$.

Fig. 3 is an illustration of the PCA method. The traffic flow series is stored in an N-row matrix, where each row is a single day's traffic flow data. Then the average of each row is removed and the principal components are extracted. We consider the residual value only affected by the long-range temporal dependence (the daily temporal dependence has been removed by the simple average and principal component I).

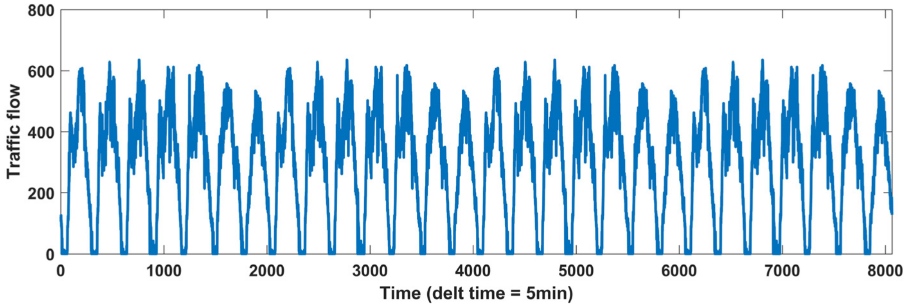


Fig. 4. An illustration of raw data for full days by detector (ID 311974).

Table 1
The time series used in the experiments.

Series type	Data source	Data preprocess
Series type	Data source	Data preprocess
Series I	Full days data	Original series
Series II	Weekdays data	Original series
Series III	Full days data	Remove simple average trend
Series IV	Weekdays data	Remove simple average trend
Series V	Full days data	Remove PCA trend
Series VI	Weekdays data	Remove PCA trend

3. Numerical experiment

3.1. Data description and experiment design

All the data used in this paper are extracted from the publicly accessible PeMS dataset [38]. The traffic flow time series are recorded at 1000 freeway traffic flow stations during August 1, 2011 to August 31, 2011. The sampled interval of the raw data is 5 min. In this paper, we choose two types of series to study the influence of the diversity in different days, which are full days data and weekdays data. Compared with weekdays data, there are much more diversity in full days data because of the difference between weekdays and holidays. Fig. 4 is an illustration of raw data for full days by detector (ID 311974). Note that though most results of the experiments in this paper are based on the data collected by detector (ID 311974), we have also tested the data from other detectors and the results are quite similar.

Here is a brief introduction of the design of the numerical experiment. First, we compare the long-range temporal dependence and the daily temporal dependence by calculating respective correlations. As pointed in Fig. 1, the correlation between series $X_1 = \{x_{i \cdot day+j}, \forall i, j\}$ and $X_2 = \{x_{(i-1) \cdot day+j}, \forall i, j\}$ derives from daily temporal dependence while the correlation between series X_1 and $X_3 = \{x_{i \cdot day+j-1}, \forall i, j\}$ derives from long-range temporal dependence. The correlations of these series can be calculated to compare the influence of two temporal dependences. Moreover, to further study on the relationship of two dependence, we calculate the correlation between series X_1 and a linear regression series

$$X_4 = \{\beta_0 + \beta_1 \cdot x_{i \cdot day+j-1} + \beta_2 \cdot x_{(i-1) \cdot day+j}, \forall i, j\}, \tag{14}$$

where β_0 is the offset, β_1 and β_2 reflect the weights of long-range and daily temporal dependence. The values of β_0, β_1 and β_2 can be obtained by utilizing the least squared method.

Second, we analyse the two kinds of temporal dependence by applying the DFA method to six different traffic flow time series listed in Table 1. To compare the two kinds of dependence, the crossover phenomenon of the Hurst exponent is studied.

Finally, the MFDFA method is also applied to look further into the properties of the series before and after removing the daily temporal dependence.

3.2. Correlations comparison

The results of the correlation coefficient and the regression parameters are shown in Tables 2 and 3. The specific distribution of the relative residual is also can be observed in Fig. 5. As we can see, β_1 and β_2 of the regression equation are close, indicating that the effects of long-range and daily temporal dependence are to some extent at the same level. Besides, the correlation coefficients reveal that the linear combination series have a better performance than only considering one of the factors despite the linear regression is the easiest way to put the two factors together.

This fundamental tests reveal that daily temporal dependence and the long-range dependence are equally important and should be taken into consideration at the same time when analysing the traffic flow data.

Table 2
Linear combination coefficients.

β_0	β_1	β_2	R^2
1.2052	0.4089	0.5930	0.9941

Table 3
Comparison of correlation coefficients for different series.

	Linear combination	Long-range dependence	Daily dependence
R^2	0.9941	0.9827	0.9818

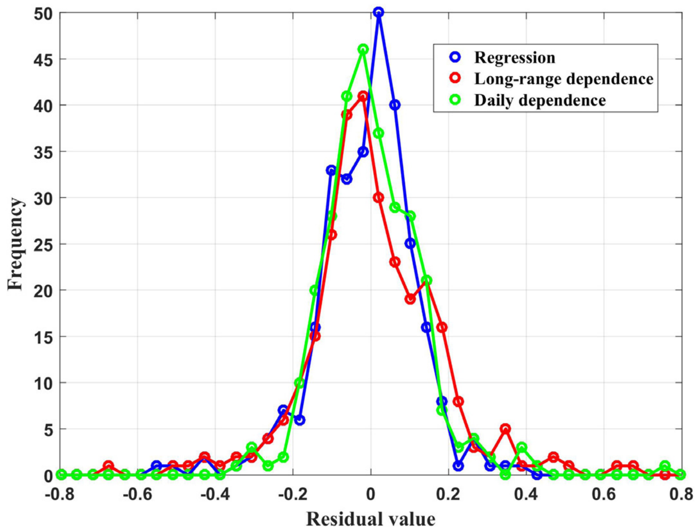


Fig. 5. Distribution of the relative residual.

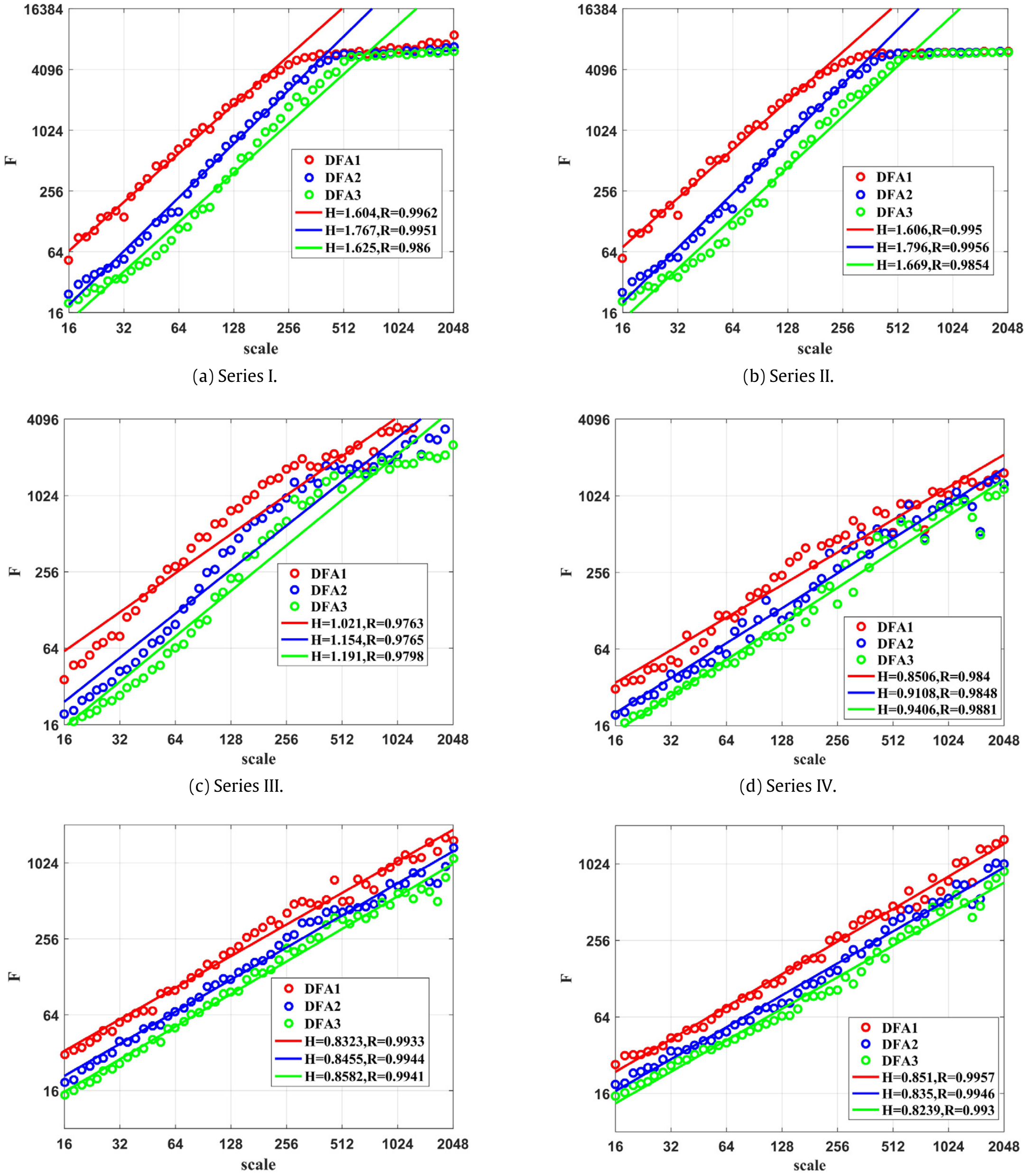
3.3. DFA result

Fig. 6 is the log–log figure of Detrended Fluctuation Analysis (DFA) for different traffic flow time series. The details of the series used in experiments are listed in Table 1. DFA1, DFA2 and DFA3 respectively represent the linear, quadratic and cubic fitting polynomial in Step (b) of detrended procedure, as shown in Eq. (5). In this paper, all the three types of fitting polynomial are used to eliminate the deviation of detrended procedure.

As Fig. 6 shows, F increases with scale by an approximate power-law and the slope of the fitting straight line is often used to estimate the Hurst exponents by previous studies. However, the crossover phenomena of Hurst exponent in Fig. 6(a) and (b) undermines the linear fitting procedure. The DFA methods fail in this situation to get correct Hurst exponent. Some previous researches have studied the crossover phenomena and proposed different explanations. For example, Li et al. [22] proposed a mixed-fractal traffic flow model to explain the crossover phenomena, which claimed that the Hurst exponent for the mainline traffic flow appears crossover phenomena if it consists of several self-similar components with different Hurst exponents. Hu et al. [39] pointed that trends such as linear trends and periodic trends affect the DFA and create the crossover phenomenon. This indicates that daily temporal dependence of traffic flow time series causes the crossover phenomena. If the daily temporal dependence is removed by daily trend, the crossover phenomena is supposed to be eliminated as well.

Fig. 6(a) and (b) confirm the assumption that the daily temporal dependence of traffic flow time series causes the crossover phenomena. Comparing the figures of the same column, it can be discovered that the crossover phenomenon can be eliminated by removing the daily temporal trend. Therefore, we can draw the conclusion that affected by the daily temporal dependence, there is crossover phenomenon in the log–log figure when analysing the long-range temporal dependence by the DFA method.

Furthermore, we compare the two methods of extracting the daily temporal dependence by comparing Fig. 6(c)-(f). Generally, we find that the PCA-based method is more effective for removing the daily temporal dependence in the sense that the correlation coefficients of PCA-based method are obviously greater than those of simple average method. This is because different from the simple average trend, the PCA-based daily trend can vary between different days. Fig. 7 is an illustration of PCA-based daily trend for different 3 days of all week data (left column) and weekdays data (right column). The daily temporal dependence of different days removed by PCA-based method is quite similar consistent with Fig. 1. One thing to notice is that before removing the PC1 of the series, the PCA-based method has removed the average. Besides, comparing



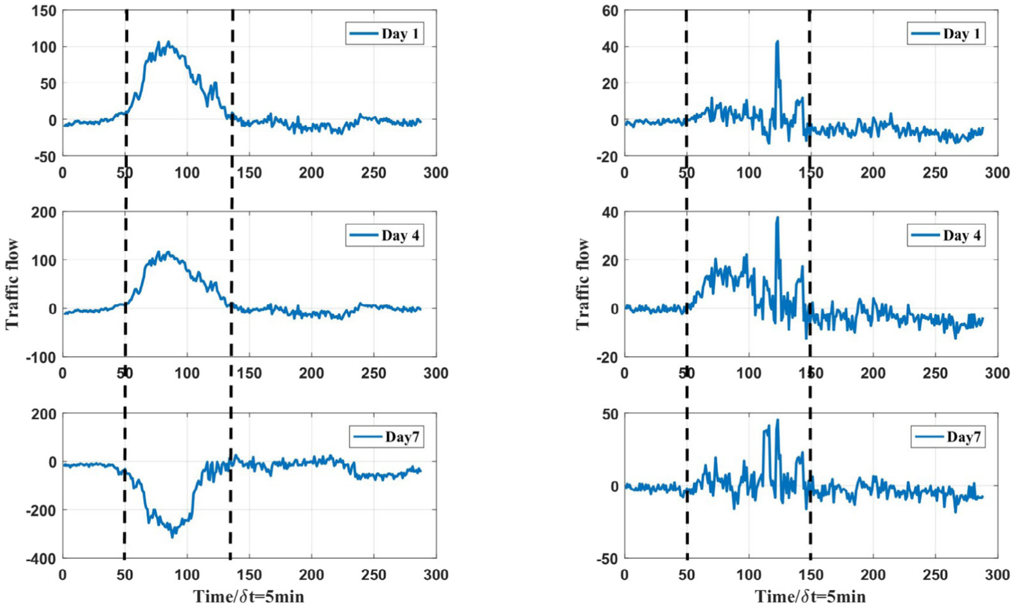
(a) Series I. (b) Series II. (c) Series III. (d) Series IV. (e) Series V. (f) Series VI.

Fig. 6. Log–log figures of the DFA. X-axis is $log(s)$ and y-axis is $logF$ of Eq. (7). The units of the scale is the number of the data points, of which the time interval is 5 min. The slope of the fitting line is so-called Hurst exponent. The details of the series used in experiments are listed in Table 1.

Fig. 6(c) and (d), the weekdays series has a larger correlation coefficient than that of full days series though both are removed the daily temporal dependence. This is because the difference between days is less significant for weekdays data so that the residual series of weekdays series removed the simple average has less offsets than full days series as shown in Fig. 8.

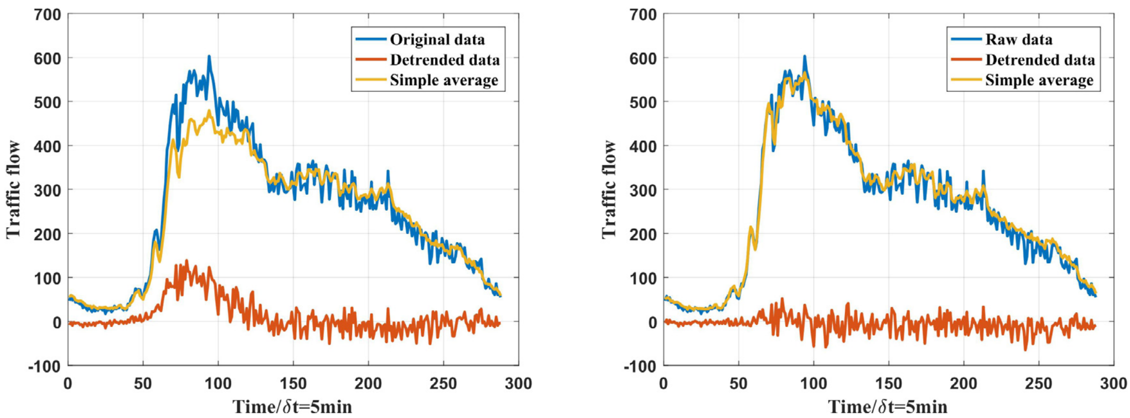
3.4. MFDFA result

We also apply MFDFA to study the series listed in Table 1 by changing the q in DFA given by Eq. (9). It can be seen that the crossover phenomenon is similar to that in DFA. Furthermore, the generalized Hurst exponent can be estimated by calculating the slope of the log–log figures. Fig. 9 gives the $Hq - q$ figures of the different series.



(a) PC1 of the full days data. (b) PC1 of the weekdays data.

Fig. 7. An illustration of PCA-based daily trend by the first PCs (PC1) for different 3 days. Notice that the curves denote the PC1 of the PCA method which do not include the simple average trend.

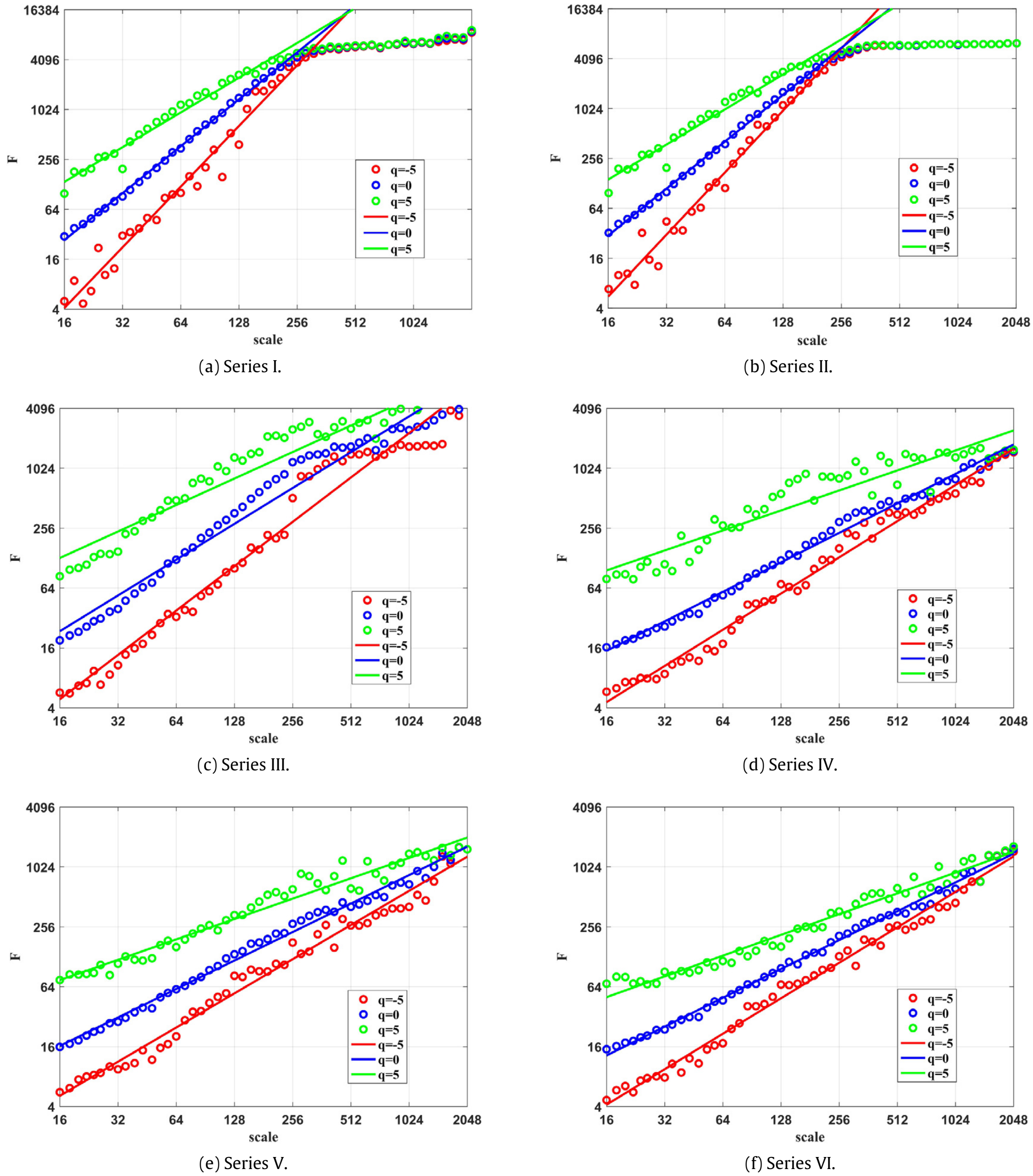


(a). (b).

Fig. 8. An illustration of the simple average method to remove the daily temporal dependence for full days data (a) and weekdays data (b).

Besides the crossover phenomenon, the original data of the traffic flow series turns out to behave clearly as a multifractal series, as shown by the red line in Fig. 10. This multifractal property may derive from long-range temporal dependence or daily temporal dependence, which can be distinguished by removing one of them. The blue lines, which present the series removed the simple average, are quite different between the full day series and the weekdays series, indicating that the simple average method is more effective for weekdays series. In contrast, the green lines, which has been eliminated the daily temporal dependence by PCA method, are quite similar for the full day series and the weekdays series.

Comparing the red lines with the green lines, it can be observed that the multifractal property of the series has been considerably weakened as the slope becomes flat after removing the daily temporal dependence. Such a result reveals that the multifractal behaviour of the original series is more likely to derive from the daily temporal dependence.



(a) Series I.

(b) Series II.

(c) Series III.

(d) Series IV.

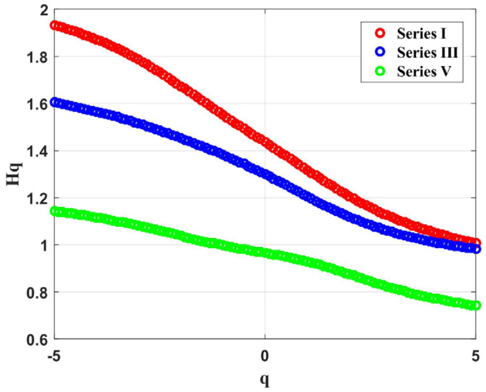
(e) Series V.

(f) Series VI.

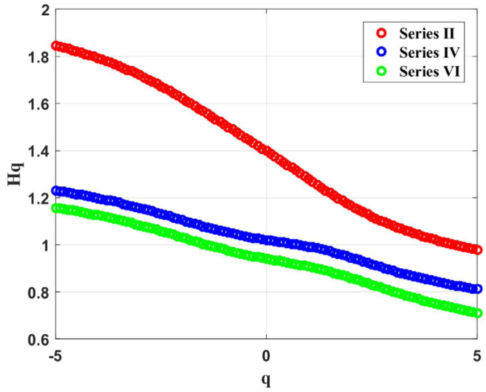
Fig. 9. Log–log figures of MFDFA by changing the q given by Eq. (9). The units of the scale is the number of the data points, of which the time interval is 5 min. The details of the series uesd in experiments are listed in Table 1.

4. Conclusion

In this paper, we focus on the long-range temporal dependence and the daily temporal dependence of the traffic flow series. First, we compare the long-range temporal dependence and the daily temporal dependence. Results show that both these two temporal dependence are equally important and should be taken into consideration at the same time. Second, the Detrended Fluctuation Analysis (DFA) is applied to compare the original series and the series removing the daily temporal dependence by different methods. Results reveal that the daily temporal dependence creates crossover phenomenon when estimate the Hurst exponent. Furthermore, through the comparison of the DFA test, PCA-based method turns out to be a better method to extract the daily temporal dependence especially when the difference between days is significant. Finally, we also conduct MFDFA to confirm the result from another perspective by changing the order q.



(a) Full days series.



(b) Weekdays series.

Fig. 10. Hq-q figure. The generalized Hurst exponents vary with q for the full days series (a) and the weekdays series (b). The red, blue and the green curves respectively stand for the original series, series removed the daily temporal dependence by simple average method and the PCA-based method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This better understanding of the temporal dependence of the traffic flow series can help us to establish more accurate model and promote the performance of predicting, compressing, etc.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China under Grant 61673233, Beijing Municipal Science and Technology Program under Grant D15110900280000.

References

[1] M.G.K. Vlahogianni, I. Eleni, J.C. Golias, Short-term traffic forecasting: Where we are and where were going, Transportation Research Part C: Emerging Technologies 43 (2014) 3–19. http://dx.doi.org/10.1016/j.trc.2014.01.005.
[2] L. Li, X. Su, Y. Zhang, Y. Lin, Z. Li, Trend modeling for traffic time series analysis: An integrated study, IEEE Transactions on Intelligent Transportation Systems 16 (6) (2015) 3430–3439. http://dx.doi.org/10.1109/TITS.2015.2457240.
[3] B.M. Williams, L.A. Hoel, Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results, Journal of Transportation Engineering 129 (2003) 664–672. http://dx.doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664).
[4] E.I. Vlahogianni, M.G. Karlaftis, J.C. Golias, Temporal evolution of short-term urban traffic flow: A nonlinear dynamics approach, Computer-aided Civil and Infrastructure Engineering 23 (7) (2008) 536–548. http://dx.doi.org/10.1111/j.1467-8667.2008.00554.x.
[5] Y.-D. Chen, L. Li, Y. Zhang, J.-M. Hu, X.-X. Jin, Fluctuations in urban traffic networks, Modern Physics Letters B 22 (02) (2008) 101–115. http://dx.doi.org/10.1142/S0217984908014468.
[6] C. Chen, Y. Wang, L. Li, J. Hu, Z. Zhang, The retrieval of intra-day trend and its influence on traffic prediction, Transportation Research Part C: Emerging Technologies 22 (2012) 103–118. http://dx.doi.org/10.1016/j.trc.2011.12.006.
[7] L. Li, Y. Li, Z. Li, Efficient missing data imputing for traffic flow by considering temporal and spatial dependence, Transportation Research Part C: Emerging Technologies 34 (2013) 108–120. http://dx.doi.org/10.1016/j.trc.2013.05.008.
[8] B. Ran, H. Tan, Y. Wu, P.J. Jin, Tensor based missing traffic data completion with spatial–temporal correlation, Physica A: Statistical Mechanics and its Applications 446 (2016) 54–63. http://dx.doi.org/10.1016/j.physa.2015.09.105.
[9] A. Cheng, X. Jiang, Y. Li, C. Zhang, H. Zhu, Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method, Physica A: Statistical Mechanics and its Applications 466 (2017) 422–434. http://dx.doi.org/10.1016/j.physa.2016.09.041.
[10] S. Feng, Y. Zhang, L. Li, A comparison study for traffic flow data compression, in: 2016 12th World Congress on Intelligent Control and Automation (WCICA), IEEE, 2016, pp. 977–982. http://dx.doi.org/10.1109/WCICA.2016.7578453.
[11] F. Shuo, K. Ruimin, X. Wang, Y. Zhang, L. Li, Traffic flow data compression considering burst components, IET Intelligent Transport Systems. http://dx.doi.org/10.1049/iet-its.2016.0328.
[12] E. Koscielny-Bunde, A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, H.-J. Schellnhuber, Indication of a universal persistence law governing atmospheric variability, Physical Review Letters 81 (3) (1998) 729. http://dx.doi.org/10.1103/PhysRevLett.81.729.
[13] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Physical Review Letters 83 (7) (1999) 1471. http://dx.doi.org/10.1103/PhysRevLett.83.1471.
[14] C.K. Peng, S.V. Buldyrev, et al., Long-range correlations in nucleotide sequences, Nature 356 (6365) (1992) 168.
[15] J. Wu, S. Wei, Time Series Analysis, Hunan Science and Technology Press, ChangSha, 1989.
[16] A. Stathopoulos, M.G. Karlaftis, A multivariate state space approach for urban traffic flow modeling and prediction, Transportation Research Part C: Emerging Technologies 11 (2) (2003) 121–135. http://dx.doi.org/10.1016/S0968-090X(03)00004-4.
[17] G. Hunt, Random fourier transforms, Transactions of the American Mathematical Society 71 (1) (1951) 38–69. http://dx.doi.org/10.2307/1990858.
[18] G. Rangarajan, M. Ding, Integrated approach to the assessment of long range correlation in time series data, Physical Review E 61 (5) (2000) 4991. http://dx.doi.org/10.1103/PhysRevE.61.4991.
[19] T. Musha, H. Higuchi, The 1/f fluctuation of a traffic current on an expressway, Japanese Journal of Applied Physics 15 (7) (1976) 1271.
[20] K. Nassab, M. Schreckenberg, S. Ouaskit, A. Boulmakoul, 1/f noise in a cellular automaton model for traffic flow with open boundaries and additional connection sites, Physica A: Statistical Mechanics and its Applications 354 (2005) 597–605.

[21] H.E. Hurst, Long-term storage capacity of reservoirs, Trans. Amer. Soc. Civil Eng. 116 (1951) 770–808.
[22] L. Li, Z. Li, Y. Zhang, Y. Chen, A mixed-fractal traffic flow model whose hurst exponent appears crossover, in: 2012 Fifth International Joint Conference on Computational Sciences and Optimization (CSO), 2012, pp. 443–447. http://dx.doi.org/10.1109/CSO.2012.103.
[23] M.R. Perati, K. Raghavendra, H.R. Koppula, M.R. Doodipala, R. Dasari, Self-similar behavior of highway road traffic and performance analysis at toll plazas, Journal of Transportation Engineering 138 (10) (2012) 1233–1238.
[24] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of dna nucleotides, Physical Review E 49 (2) (1994) 1685. http://dx.doi.org/10.1103/PhysRevE.49.1685.
[25] M. Höll, H. Kantz, Y. Zhou, Detrended fluctuation analysis and the difference between external drifts and intrinsic diffusionlike nonstationarity, Physical Review E 94 (4) (2016) 042201. http://dx.doi.org/10.1103/PhysRevE.94.042201.
[26] P. Shang, Y. Lu, S. Kama, The application of hölder exponent to traffic congestion warning, Physica A: Statistical Mechanics and its Applications 370 (2) (2006) 769–776. http://dx.doi.org/10.1016/j.physa.2006.02.032.
[27] C.-K. Peng, S. Buldyrev, A. Goldberger, S. Havlin, M. Simons, H. Stanley, Finite-size effects on long-range correlations: Implications for analyzing dna sequences, Physical Review E 47 (5) (1993) 3730. http://dx.doi.org/10.1103/PhysRevE.47.3730.
[28] I. Okutani, Y.J. Stephanedes, Dynamic prediction of traffic volume through kalman filtering theory, Transportation Research Part B: Methodological 18 (1) (1984) 1–11. http://dx.doi.org/10.1016/0191-2615(84)90002-X.
[29] B.M. Williams, L.A. Hoel, Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results, Journal of Transportation Engineering 129 (6) (2003) 664–672. http://dx.doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)#sthash.0oeAWrF2.dpuf.
[30] X. Ning, Z. Li, J. Song, A practical research on dynamic traffic flow data profile, in: Proceedings. the IEEE 5th International Conference on Intelligent Transportation Systems, 2002, IEEE, 2002, pp. 791–794. http://dx.doi.org/10.1109/ITSC.2002.1041320.
[31] L. Zhiheng, Y. Shengchao, T. Ye, L. Li, Z. Zhiqiang, J. Yan, Urban traffic flow volume modeling for beijing using a mixed-flow model, Journal of Transportation Systems Engineering and Information Technology 8 (3) (2008) 111–114. http://dx.doi.org/10.1016/S1570-6672(08)60029-X.
[32] Y. Kamarianakis, H.O. Gao, P. Prastacos, Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions, Transportation Research Part C: Emerging Technologies 18 (5) (2010) 821–840. http://dx.doi.org/10.1016/j.trc.2009.11.001.
[33] L. Qu, L. Li, Y. Zhang, J. Hu, Ppca-based missing data imputation for traffic flow volume: A systematical approach, IEEE Transactions on Intelligent Transportation Systems 10 (3) (2009) 512–522. http://dx.doi.org/10.1109/TITS.2009.2026312.
[34] D. Boto-Giralda, F.J. Díaz-Pernas, D. González-Ortega, J.F. Díez-Higuera, M. Antón-Rodríguez, M. Martínez-Zarzuela, I. Torre-Díez, Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks, Computer-aided Civil and Infrastructure Engineering 25 (7) (2010) 530–545. http://dx.doi.org/10.1111/j.1467-8667.2010.00668.x.
[35] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H.E. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, Physica A: Statistical Mechanics and its Applications 316 (1) (2002) 87–114. http://dx.doi.org/10.1016/S0378-4371(02)01383-3.
[36] J.W. Kantelhardt, E. Koscielny-Bunde, H.H. Rego, S. Havlin, A. Bunde, Detecting long-range correlations with detrended fluctuation analysis, Physica A: Statistical Mechanics and its Applications 295 (3) (2001) 441–454. http://dx.doi.org/10.1016/S0378-4371(01)00144-3.
[37] I. Jolliffe, Principal Component Analysis, Wiley Online Library, 2002. http://dx.doi.org/10.1002/9781118445112.stat06472.
[38] Performance measurement system (pems), http://pems.dot.ca.gov/.
[39] K. Hu, P.C. Ivanov, Z. Chen, P. Carpena, H.E. Stanley, Effect of trends on detrended fluctuation analysis, Physical Review E 64 (1) (2001) 011114. http://dx.doi.org/10.1103/PhysRevE.64.011114.