Physica A 467 (2017) 419–431

Contents lists available at ScienceDirect

Physica A

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

How long will the traffic flow time series keep efficacious to forecast the future?

PengCheng Yuan [a,*], XuXun Lin [b]

[a] Business School, University of Shanghai for Science and Technology, Shanghai 200093, China
[b] Business School of Changzhou University, Changzhou 213164, China

HIGHLIGHTS

- The traffic flow time series memory property is analyzed using R/S analysis method.
- Both of the freeway and the groundway traffic flow time series have long-term memory.
- This study can provide theoretical basis for ITS data acquisition and forecasting.
- It would be useful but rarely involved in previous researches.

ARTICLE INFO

Article history:
Received 26 March 2016
Received in revised form 7 July 2016
Available online 14 October 2016

Keywords:
Traffic flow time series
Long memory
Modified rescaled range analysis
Maximum Lyapunov exponent

ABSTRACT

This paper investigate how long will the historical traffic flow time series keep efficacious to forecast the future. In this frame, we collect the traffic flow time series data with different granularity at first. Then, using the modified rescaled range analysis method, we analyze the long memory property of the traffic flow time series by computing the Hurst exponent. We calculate the long-term memory cycle and test its significance. We also compare it with the maximum Lyapunov exponent method result. Our results show that both of the freeway traffic flow time series and the ground way traffic flow time series demonstrate positively correlated trend (have long-term memory property), both of their memory cycle are about 30 h. We think this study is useful for the short-term or long-term traffic flow prediction and management.

1. Introduction

The growth in urban traffic congestion has been recognized as a serious problem in all large metropolitan areas in the country, with significant effect on the economy, travel behavior, land use and a cause of discomfort for millions of motorists [1,2]. In order to improve the traffic environment performance, people have tried lots of ways, but the most widely used way is to construct many roads. However, in fact, the construction of urban roads alone cannot solve the urban traffic congestion problem effectively [3–5]. The Intelligent Traffic Systems (ITS) may provide a compensation way to relieve this problem. Traffic guidance system is a part of ITS, it can predict the variation trend of traffic flow, provide real traffic information (travel time, traffic flow state etc.) for travelers through collecting and analyzing the efficacious traffic data. So it is important to improve the performance of traffic system. But at the same time a new problem subsequently produced, that is, how long the traffic flow time series data keeps efficacious to forecast the future? Or in other words, whether we can estimate the

* Corresponding author.
E-mail addresses: danis_cx@126.com (P. Yuan), linxuxun@hotmail.com (X. Lin).

http://dx.doi.org/10.1016/j.physa.2016.10.020
0378-4371/

traffic flow variation trends of time $t_0 + t$ depending on the existing traffic flow data t_0? If it is yes, then we can say the traffic flow time series data is efficacious in time t, or it has t-long memory.

Although humans have mastered a lot of knowledge about traffic flow, in fact, so far, humans do not know exactly how it evolved. Is it orderly or chaotic, linear or nonlinear, predictable or unpredictable? Unfortunately, till now there is no clear answers for these problems. Since 1990s, nonlinear dynamics theory, chaos theory, fractal theory and other nonlinear theories are widely used to discover the traffic flow evolution mechanism, and it provides a new perspective for the traffic flow studies [6]. With the help of fractal theory and other statistical methods, we can deeply study the evolution process of the traffic flow in the following areas: (1) Whether the traffic flow state has the characteristic of continuity? (2) Whether the traffic flow time series data has the same or similar statistical characteristics at different time scales? (3) Whether the traffic flow state can be predicted, how long of the prediction period is reliable? Obviously, we cannot take advantage of the existing traffic flow time series data to predict the future traffic flow variation trends (because there is no any correlation between the data of t_0 and $t_0 + t$) if the traffic flow time series data is completely random (there is no memory of the data). But, if it is not completely random (which means the data of t_0 is in correlation with $t_0 + t$, or the time series data has t-long memory), then we can use the correlationship to infer the variation trend of data $t_0 + t$ using data t_0. Now, a problem arises, whether the traffic flow time series indeed displays the memory property (above problem (1))? If it is, whether the memory cycle of them is similar, and how long it keeps (above problem (2))? All of these questions are core issues of this paper. So, in fact, the purpose and contribution of this paper is to take an empirical study for the freeway and ground way travel time series using fractal theory. This study can provide reliable theoretical basis for the intelligent transportation system data acquisition and traffic flow forecasting.

Fractal geometry is first founded by Mandelbrot in 1967, and soon it is rapidly developed into an emerging branch of mathematics. It is a powerful theoretical tool in processing the irregular problem of the science and engineering. Now, it has been applied to almost all fields of the natural sciences (for example human heartbeats [7], image compression algorithms [8] and so on), and actually plays the connection role for various fields in modern science. In recent years, it has been discovered that many human actions such as the traffic flow [4,9] and the stock market [10] display fractal properties. At the beginning of the traffic flow study, researchers often use statistical methods to analyze the mathematical relationship among various parameters. For example, the Greenshields model, Greenberg model, GHR model etc. [11]. Nowadays, many researchers tend to use chaos/fractal etc. nonlinear statistical methods to study the traffic flow [9,12–14]. Dendrinos firstly introduce the chaos concept to the traffic flow studies [15]. Then, by an improved car-following model, Low and Addison discovered different chaotic properties of the traffic flow [16]. Safonov studied the chaotic properties of traffic flow through a delayed difference equation model, he finds the chaos occurs mainly in the medium-density conditions of traffic flow, and there is multifractal attractor structure [17]. Nagatani studied the dynamical behavior of vehicular traffic through a sequence of traffic lights positioned self-similarly on a high way, where all traffic lights turn on and off simultaneously with cycle time, they found the vehicle exhibits the complex chaotic motion behavior under certain conditions with varying cycle time [18–20]. Shang applied multifractal modeling techniques to analyze the traffic flow time series data collected from the Beijing Yuquanying [1], the results show that chaotic properties of traffic flow really exist, and perhaps the Holder exponent may be used as an indicator to predict the presence of the traffic congestion?

Previous researches show the chaos/fractal properties of the traffic flow, however, till now there are few studies on exploiting these properties in traffic engineering and practice. In traffic engineering, especially in ITS, the chaos/fractal properties are so important for us to analyze the time-dependent variation characteristics of traffic flow, to forecast the short-time traffic flow, traffic state, travel time and so on. All of these studies' premise in the traffic flow time series is large-timescale auto covariance and we know the variation cycle clearly. But, much to our pity, till now there is few literature to exploit these questions. So in this paper we study the self-similar fractal property of traffic flow time series, especially analyze the variation cycle of it. We think this research can provide a basic theory evidence for the traffic engineers, and it is meaningful for traffic guidance and control.

In practice, we often use traffic flow velocity, average vehicle speed and average vehicle time occupancy as three types of parameters to characterize the traffic flow. So in this paper, we study the large-timescale auto covariance and variation cycle of traffic flow time series data mainly depends on these three types of parameters. This paper is organized as flows: after the introduction, Section 2 introduces the modified R/S analyzation method and two methods for calculating the variation cycle. Section 3 introduces the sample data collection. In Section 4, we analyze the Hurst exponent, test significance of the results and calculate the variation cycle. The empirical results are also shown in Section 4. Last, some concluding remarks and discussion of future research are discussed in Section 5.

2. Methodology

2.1. Modified rescaled range analysis

If the system we studied is completely random and its limit distribution is normal, standard statistical methods can be used effectively. But, if the system is a nonlinear system which is some what between completely random and determined, these standard statistical methods are invalid. The classic rescaled range analysis (R/S) (proposed by Hurst 1951) is a non-parametric statistical method which is widely used to analyze these types of systems. The most advantage of this method is that the robustness of the results will not be affected whether the time series data follows normal distribution or not, so we do not need to know the distribution of the time series data when we are using this method. Now, classic rescaled range analysis is widely used for detecting the long memory in financial markets [21], stock market [22], traffic and transportation [5,13,14]. The calculating procedure of R/S analysis can be described as follows [23].

Step 1: Divided the time series $\{R_t\}$ (with length N) into A (A is taken by the integer part of N/n) equal sub-sequence with length n ($n \geq 3$), each sub-sequence is denoted as D_a ($a = 1, 2, 3 \ldots A$), and the elements in each sub-sequence are marked as $R_{k,a}$;

Step 2: Calculate mean value of each subsequence;

$$e_a = (n-1)\sum_{k=1}^{n} R_{k,a}. \tag{1}$$

Step 3: Calculate the cumulative deviation (deviate from the mean value of the subsequences) of each subsequence D_a;

$$X_{k,a} = \sum_{i=1}^{k}\left(R_{i,a} - e_a\right). \tag{2}$$

Step 4: Calculate the range of each subsequence D_a;

$$R_a = max\left\{X_{k,a}\right\} - min\left\{X_{k,a}\right\} \quad k = 1, 2, 3 \ldots n \tag{3}$$

Step 5: Calculate the standard deviation of each subsequence D_a;

$$S_a = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(R_{i,a} - e_a\right)^2}. \tag{4}$$

Step 6: Calculate the rescaled range of the subsequence D_a;

$$(R/S)_a = R_a/S_a. \tag{5}$$

Step 7: Now, we get a rescaled range sequence $\left\{(R/S)_1, (R/S)_2, \ldots, (R/S)_A\right\}$. Calculate mean value of this sequence;

$$(R/S)_n = (1/A)\sum_{a=1}^{A}(R/S)_a. \tag{6}$$

Step 8: Increase the sub-sequence length n to a larger value, repeat the above steps until $n = N/2$. Then we can get the Hurst value based on the regression equation as shown by (7).

$$\log (R/S)_n = \log C + H \log n. \tag{7}$$

$\log n$ is independent variable, $\log (R/S)_n$ is dependent variable. Using the ordinary least squares estimation method, we can get the estimation value of the Hurst exponent H.

Here, the Hurst exponent H can be obtained by observing the slope of log–log plot of R/S –- n using the method of ordinary least squares. For the time series data of traffic flow, if $0.5 < H < 1.0$, it displays long memory property, the form of the memory is positive. However, if $H = 0.5$, the time series data is completely random, there is no memory property.

Despite classic R/S analysis can be used to analyze the short-term and long-term dependence of the time series, it cannot distinguish them [24]. So, if a time series displays strong short-term correlation, the result often shows some deviation by the classic R/S analysis method, it inclines to get long-term correlation. In order to compensate for the deficiency, [24] presents a modified R/S analysis method. By introducing the sample covariance,the modified R/S analyzing result can filter the short-term memory. The modified $(R/S)_n$ statistic can be got as

$$(R/S)_n = (1/A)\sum_{a=1}^{A}\left[R_a/\sigma_a(q)\right] \tag{8}$$

where $\sigma_a(q)$ is

$$\sigma_a(q) = \frac{1}{n}\sum_{k=1}^{n}\left(R_{k,a} - e_a\right)^2 + \frac{2}{n}\sum_{j=1}^{q}\omega_j(q) \times \left[\left(R_{k,a} - e_a\right)\left(R_{k-j,a} - e_a\right)\right] = \sigma_a^2 + 2\sum_{j=1}^{q}\omega_j(q)\,\gamma_j. \tag{9}$$

When $q < n$, $\omega_j(q) = 1 - \frac{k}{q+1}$, the σ_a^2 is the sample variance and the γ_j is the sample jth order auto covariance. The optimal value of q is given according to the following rule

$$q^* = int\left\{(3N/2)^{1/3}\left[2\rho/\left(1-\rho^2\right)\right]^{2/3}\right\} \tag{10}$$

where ρ is the 1th order auto covariance parameters of sequence $\{R_t\}_{t=1}^{N}$.

2.2. Calculate the time series cycle length

The modified R/S method can be used to measure the cycle length of the time series. As mentioned above, the time series has memory property when $0.5 < H < 1.0$, it means the time series is sensitive by initial value. But, in general, the memory length is bounded. If the time series cycle length is m, then in all n which is greater than m, the memory property would disappear, and the scatter plot of $\log(R/S)_n$ and $\log n$ will be a straight line with slope about 0.5. So, we can get the cycle by finding the inflection point of the scatter plot of $\log(R/S)_n$ and $\log n$. While, it is worth noting that the cycle is a statistic concept, it means average cycle length of the time series. Based on the modified R/S analysis, Peters presents a statistic for calculating the cycle length $V_n(q)$ [23].

$$V_n(q) = \frac{1}{\sqrt{n}}(R/S)_n. \tag{11}$$

When the statistic of modified R/S and the $\sqrt{n}$ increase simultaneously, the scatter plot of V_n and $\log n$ will be distributed in a same horizontal line. If the statistic of modified R/S is increased faster than $\sqrt{n}$, the scatter plots occur in upward trend, otherwise the scatter plots occur in downward trend. The inflection point is the cycle point.

2.3. Maximum Lyapunov exponent

Lyapunov exponent is an effective indicator for measuring the sensitive dependence on initial conditions of the system. It can reflect the divergence or convergence degree of two points in the system evolved by time. A negative Lyapunov exponent shows how long the system can be restored to its original state when it is disturbed, a positive Lyapunov exponent measures how fast the system is diverging.

Before calculating the maximum Lyapunov exponent, we need to reconstruct the phase space. For the time series $\{R_t\}$ with length $N : R_1, R_2 \ldots R_N$, if we want to embed it in the m dimensional Euclidean space, we need select a time delay τ at first. Then, from the start data R_1, we select the rest data series by the time delay τ until we get all of the m dates. At last we can get the first phase point $X_1 = \left(R_1, R_{1+\tau}, \ldots, R_{1+(m-1)\tau}\right)$. Repeat above process from the second data of $\{R_t\} : R_2, R_3 \ldots R_N$, and we can get the phase point $X_2 = \left(R_2, R_{2+\tau}, \ldots, R_{2+(m-1)\tau}\right)$, $X_3 = \left(R_3, R_{3+\tau} \ldots, R_{3+(m-1)\tau}\right), \ldots, X_{N_m} = \left(R_{N_m}, R_{N_m+\tau}, \ldots, R_N\right)$, where $N_m = N - (m-1)\tau$. The new phase space $\left\{X_1, X_2, \ldots, X_{N_m}\right\}$ is the reconstructed phase space. In this paper we let $m = 4$, the value of $\tau = 7$ h, the value of m is got by our experience, the value of τ can be got by the relationship [25] $\tau = int\ (T/m)$, T is the memory cycle which will be analyzed in Section 4.

The calculating procedure of maximum Lyapunov exponent can be described as follows.

Step 1: Reconstruct the traffic flow time series to an m dimensional phase space;

Step 2: Based on the initial phase point X_{t_0}, choose a new point as the endpoint which keeps the distance at least one period to point X_{t_0}, generate an initial vector $\overrightarrow{V_0}$. Then, calculate the vector length of $\overrightarrow{V_0}$, marked by L_0;

Step 3: After an evolution time τ_0, the initial vector $\overrightarrow{V_0}$ evolves to a new vector $\overrightarrow{V_1}$, its corresponding starting point can be marked as $X_{t_0+\tau_0}$, calculate its phase length L_1. The phase length becomes L_1 from L_0 in the time interval τ_0. Using λ_1 represents the index growth rate of the phase length, then we have

$$L_1 = L_0 e^{\lambda_1 \tau_0} \quad \text{or} \quad \lambda_1 = \frac{1}{\tau_0}\log(L_1/L_0). \tag{12}$$

Step 4: Let the $X_{t_0+\tau_0}$ be a new start point, choose a new vector $\overrightarrow{V_1'}$, let $\overrightarrow{V_1'}$ be the new initial vector, we can get the exponential growth rate λ_2 by the same method above.

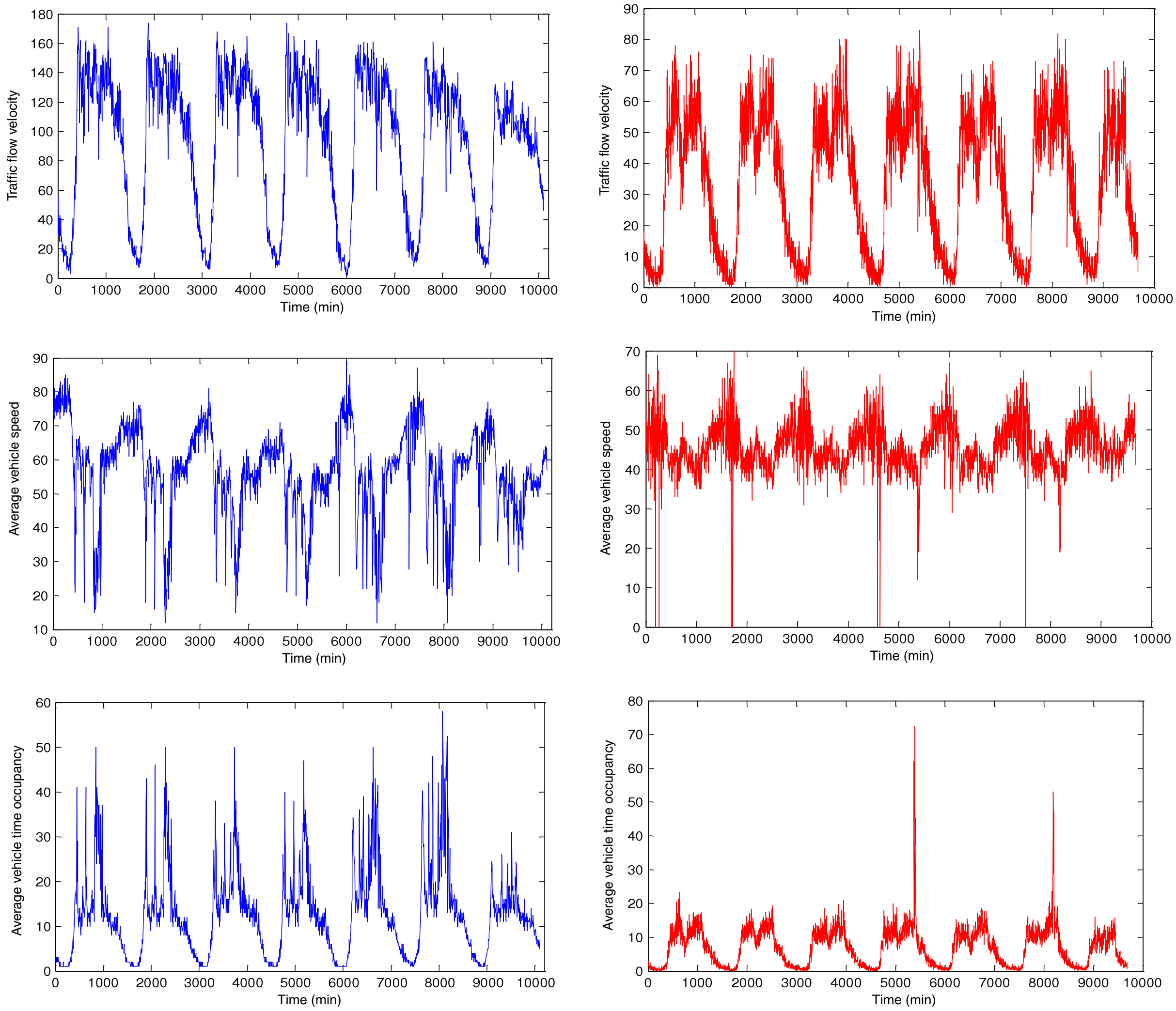
$$\lambda_2 = \frac{1}{\tau_0}\log(L_2/L_1). \tag{13}$$

Repeat the process until the end of the point set, then generate an exponential growth sequence: $\lambda_1, \lambda_2, \ldots, \lambda_{N_0}$. Calculate the mean value of this sequence, it will be the estimated value of the largest Lyapunov exponent.

$$\lambda = \frac{1}{N_0}\sum_{k=1}^{N_0}\lambda_k = \frac{1}{N_0}\sum_{k=1}^{N_0}\frac{1}{\tau_0}\log(L_k/L_{k-1}). \tag{14}$$

Step 5: Increase the embedding dimension m, repeat the steps above until the estimated value maintains balance with the m, the calculation result is the maximum Lyapunov exponent.

The maximum Lyapunov exponent is significant for us: first, its reciprocal reflects the prediction boundary of the system; second, if the traffic flow time series is chaotic, then there at least exists one positive Lyapunov exponent which reflects the divergence speed of the system.



(a) Left: freeway traffic flow time series.

(b) Right: ground way traffic flow time series.

Fig. 1. Traffic flow time series data.

3. Data

There are various parameters that can be used to characterize traffic flow, however, as mentioned in Ref. [1], not all of them can be viewed as interesting, meaningful or even easy to detect for the kind of fractal data analysis. In this paper, we collect two types of traffic flow time series data: freeway and ground way. Each of the them includes three types parameters: velocity, average vehicles speed and average vehicles time occupancy. These are collected from the Shanghai north and south freeway in 2008/4/28–2008/5/19 (about 21 days), downloaded from the freeway performance measurement project run by shanghai series intelligent system company. The raw data of them are collected in every 5 min for each lane by the loop detector located on the freeway. The ground way traffic flow time series data is observed on the Changzhou China in 2008/9/23–2008/10/8 (about 16 days), downloaded from the traffic intelligent transportation system control center run by Changzhou traffic police detachment. The raw data of them are collected in every 3 min for each lane by the loop detector located on the ground way. Fig. 1 shows the summary of the data collection.

4. Empirical results

We use the modified R/S method to analyze the memory property of freeway traffic flow and the ground way traffic flow time series. At first, we calculate the Hurst exponent value, the cycle length of them, and test the significant of the Hurst exponent. We choose the interval of 10 min (freeway)/15 min (ground way) to half of the time series for freeway $(10 < n < N/2)$ and ground way $15 < n < N/2$, calculate the rescaled range sequence. Then, we calculate the statistic

Table 1
The regression and test results of Hurst exponent (freeway 5 min data).

Project		$10 < n < 381$			$381 < n < 3034$			$10 < n < 3034$		
		Velocity	Speed	Occupancy	Velocity	Speed	Occupancy	velocity	speed	occupancy
Regression results	Constant	−0.231	−0.132	0.019	−0.435	−0.879	−0.332	0.132	0.215	0.099
	Coefficient	0.721	0.732	0.718	0.521	0.434	0.503	0.625	0.564	0.602
	Test Statistics (R^2)	0.978	0.997	0.956	0.987	0.976	0.990	0.967	0.921	0.923
Hurst value	Estimated Value	0.721	0.732	0.719	0.522	0.443	0.534	0.625	0.564	0.602
	Significance test (t-test)	8.98	7.91	7.77	8.09	8.84	8.34	1.58	1.21	1.93

value of $V(q)$. The time interval of maximum value of $V(q)$ is the cycle length. By the ordinary least squares regression, we can get the Hurst exponent value.

4.1. Empirical results of freeway traffic flow data

Using the modified R/S analysis method, we analyze the freeway traffic flow time series data with 5 min granularity (include traffic flow velocity, average vehicle speed and average vehicle occupancy). The scatter plots of the $\log(R/S)_n$ and $\log n$, $V(q)$ and $\log n$ are shown in Fig. 2.

We know from the right of Fig. 2(a) that the statistical V value of the traffic flow velocity time series reaches the maximum when n equals 381, the V value decreases when $n > 381$. So, we can determine that $n = 381$ is the inflexion point of the traffic flow velocity time series. Because the traffic flow velocity time series data granularity size is 5 min, we can get the traffic flow velocity time series cycle by $381 \times 5 = 1905$ min (about 31 h). We show the results of estimated parameters for traffic flow velocity time series in the left of Fig. 2(a), where we can know that the Hurst value equals 0.7215 when $n \leq 381$ and 0.5212 when $n > 381$. It obviously shows that there is a long-term memory of the 5 min freeway traffic flow velocity time series. But, this long-term memory property disappears after 1905 min ($n > 381$).

We know from the left of Fig. 2(b) that the statistical V value displays inflex when $n = 354$. Similarly, we can determine the cycle of the average vehicles speed time series by $354 \times 5 = 1770$ min (about 29 h). We show the results of estimated parameters for average vehicle speed time series in the left of Fig. 2(b), where we can know that the Hurst value equals 0.7321 when $n \leq 354$ and 0.4343 when $n > 354$. It obviously shows that there is a long-term memory of the 5 min freeway average vehicles speed time series.

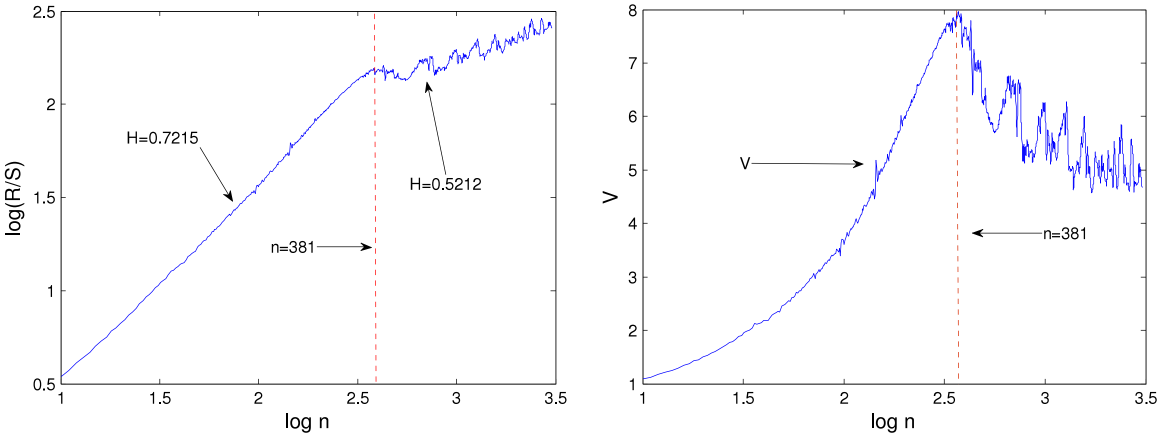
We can know from the right of Fig. 2(c) that the statistical V value displays inflex when $n = 384$, we can determine the cycle of the average vehicle time occupancy time series by $384 \times 5 = 1920$ minute (about 32 h). We show the results of estimated parameters for average vehicles time occupancy time series in the left of Fig. 2(c), where we can know that the Hurst value equals 0.7189 when $n \leq 384$ and 0.5034 when $n > 384$. It obviously shows that there is a long-term memory of the 5 min granularity freeway average vehicle time occupancy time series.

In summary, there is a long-term memory of 5 min freeway traffic flow time series (traffic flow velocity, average vehicles speed and average vehicles time occupancy), the memory cycle is about 30 h. In order to assure the reliability of the results, the regression results are examined. The test results are shown in Table 1. It indicates that the first and second stages of the Hurst value show significant, but it is not significant for the all stages. So we think the Hurst value analyzed above is reliable.

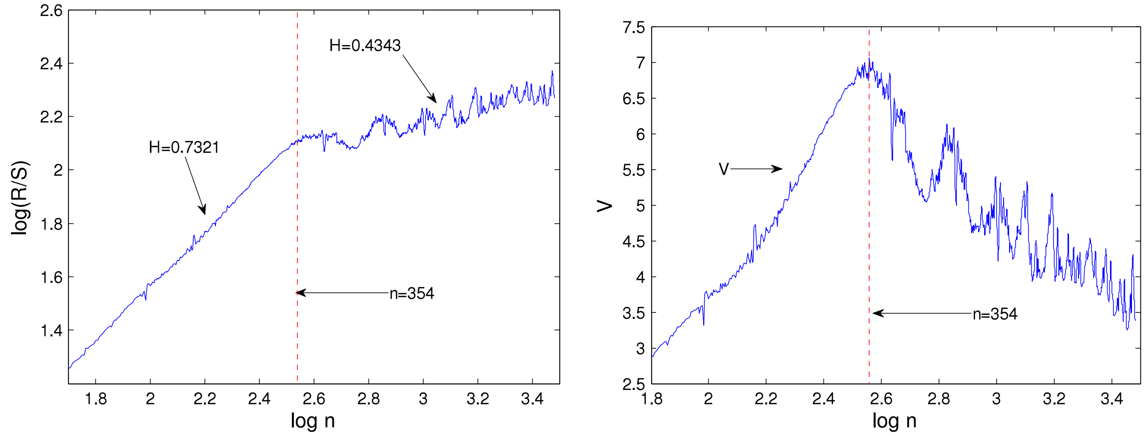
Fig. 3 provides us with the relationship between Hurst values and sequence interval n, which is derived by the following way. We conduct a regression analysis every time the value of n increases, with the regression starting point of 10 min and the increment of 5 min each, until a half of sequence length. We can see that the Hurst value of traffic flow velocity reaches the maximum with $n = 381$ (1905 min), before which the Hurst value increases with the increase of n, while the trend becomes contrary after n is above 381 (1905 min). The aforementioned feature indicates that the long-term memory gradually increases until $n = 381$ (1905 min); however, it decreases gradually after $n = 381$ (1905 min) and nearly disappears after $n = 2500$. In the same way we can know the change trends of the relationship between the Hurst and the other parameters (the average vehicles speed and average vehicles time occupation), which are $n = 354$ and $n = 384$ respectively. These conclusions are consistent with foregoing analysis results.

In Fig. 4(a), we depict the relationship between Hurst values and sequence interval n using the data of 25 min freeway traffic flow velocity. The results are extremely similar with that of 5 min. Both of them show almost identical Hurst values. The cycling period n is 76. The actual cycling period can be demonstrated to be 1900 (76×25) minutes(approximately 31 h). So it can be illustrated that the 5 min and 25 min freeway traffic flow velocity time series have the same long-term memory characteristics. In Fig. 4(b) and (c), the freeway traffic flow average vehicles speed and average vehicles time occupation of 25 min are analyzed respectively. In Fig. 4(b), the Hurst value and cycling period of average vehicles speed of 25 min are revealed to be 0.6893 and 75(1875 min), only have 0.0428 and 105 min differences with that of 5 min. In Fig. 4(c), these two values are 0.7208 and 76(1900 min), the differences are only 0.0109 and 20 min compared with that of 5 min.

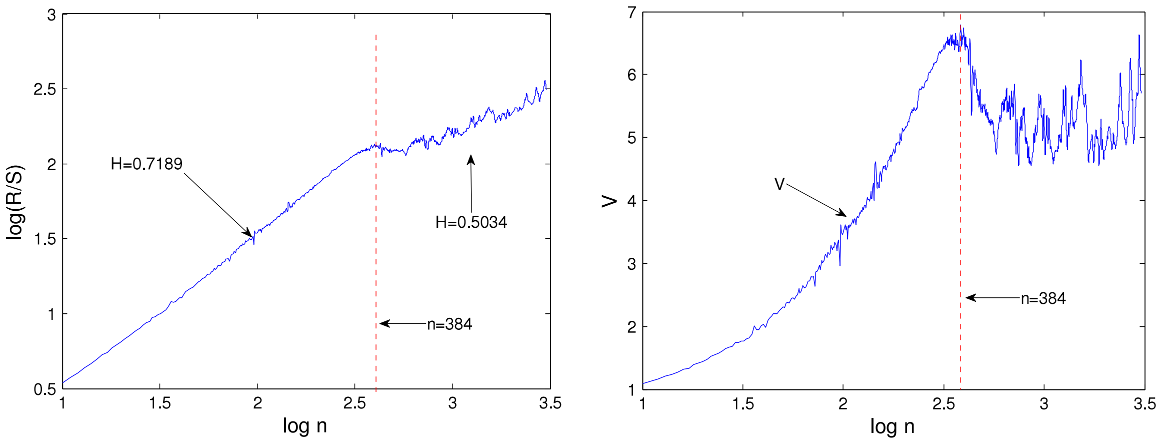
According to above analysis, we know that the freeway traffic flow time series of 25 min have long-term memory too, their long-term characteristic is similar to that of 5 min. However, the estimated value of Hurst index gradually decreases as sampling frequency increases. It means increasing sampling frequency will result in the increase of data noise. At the same



(a) Traffic flow velocity.



(b) Average vehicle speed.



(c) Average vehicle time occupancy.

Fig. 2. The Hurst exponent and cycle of freeway traffic flow time series (5 min data).

time, another conclusion can be drawn from the above analysis that there exists scale-free property about freeway traffic flow time series. In order to verify the conclusions above, we use another method to estimate the cycling period of the freeway traffic flow time series. If these two results are similar, we think the conclusions above are reliable. As mentioned in 2.3, if we know the maximum Lyapunov exponent value L, then we can get the cycle by $T = 1/L$. By the method of 2.3

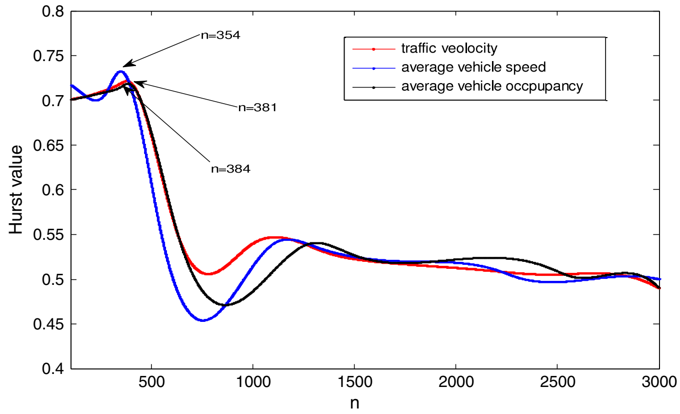


Fig. 3. The Hurst value of freeway traffic flow.

Table 2
The regression and test results of Hurst exponent (groundway 3 min data).

Project		$10 < n < 636$			$636 < n < 3732$			$10 < n < 3732$		
		Velocity	Speed	Occupancy	Velocity	Speed	Occupancy	Velocity	Speed	Occupancy
Regression results	Constant	−0.782	−0.981	−1.651	−0.651	−0.876	−0.272	0.124	0.187	−0.453
	Coefficient	0.745	0.680	0.752	0.201	0.512	0.243	0.437	0.441	0.501
	Test Statistics (R^2)	0.967	0.978	0.934	0.895	0.923	0.956	0.911	0.879	0.859
Hurst value	Estimated Value	0.745	0.680	0.752	0.201	0.512	0.243	0.437	0.441	0.501
	Significance test (t-test)	6.131	7.982	9.140	8.801	9.263	11.22	1.171	0.993	1.688

we can get the value of freeway traffic flow time series $L_v = 0.00248$, $L_s = 0.00271$ and $L_o = 0.0025$. then we can get: $T_v = 1/L_v = 403$, $T_s = 1/L_s = 370$, and $T_o = 1/L_o = 400$. These results are comparatively approximate with the modified R/S analysis results.

4.2. Empirical results of ground way traffic data

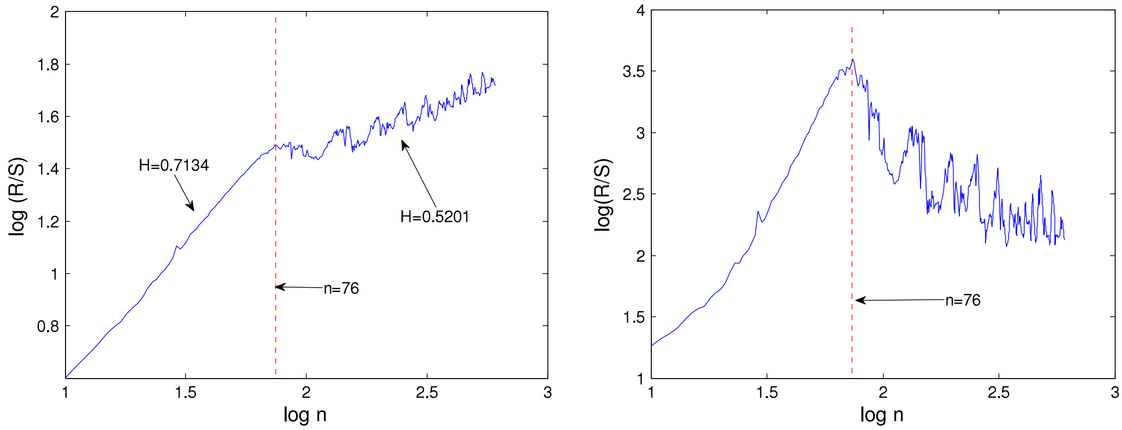
We also analyze the ground way traffic flow time series data with 3 min by the modified R/S method. Let $15 < n < N/2$ we get the scatter plot of the $\log (R/S)_n$ and $\log n$, $V_n\ (q)$ and $\log n$, which is shown in Fig. 5.

We can know from the right of Fig. 5(a) that the statistical V value of the ground way traffic flow velocity time series reaches the maximum when n equals 630, the V value decreases when $n > 630$. Therefore, we can determine that $n = 630$ is the inflexion point of the ground way traffic flow velocity time series. Because the data granularity size is 3 min, so we can get the ground way traffic flow velocity time series cycle by $630 \times 3 = 1890$ min (about 31 h). We show the results of estimated parameters for traffic flow velocity time series in the left of Fig. 5(a), where we can know that the Hurst value equals 0.7451 when $n < 630$ and 0.2015 when $n > 630$. It obviously shows that there is a long-term memory of the 3 min granularity size ground way traffic flow velocity time series. But, after 1890 min ($n > 630$), this long-term memory property disappears, and it displays anti-persistent property.

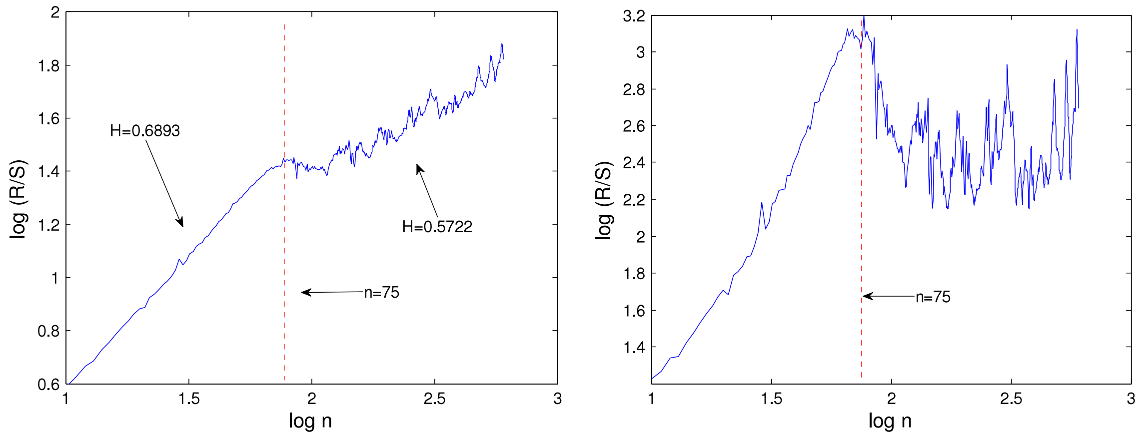
We can know from the right of Fig. 5(b) that the statistical V value displays inflex when $n = 636$, similarly we can determine the cycle of the average vehicle speed time series by $636 \times 3 = 1908$ min (about 31 h). We show the results of estimated parameters for average vehicle speed time series in the left of Fig. 5(b), where we can know that the Hurst value equals 0.6866 when $n \leq 636$ and 0.5125 when $n > 636$. It obviously shows that there is a long-term memory of the 5 min granularity size freeway average vehicle speed time series. But, this long-term memory property disappears after 1908 min ($n > 636$).

We can know from Fig. 5(c) that the statistical V value displays inflex when $n = 633$, similarly we can determine the cycle of the average vehicle time occupancy time series by $633 \times 3 = 1908$ min (about 31 h). We show the results of estimated parameters for average vehicle time occupancy time series in the left of Fig. 5(c), where we can know that the Hurst value equals 0.7522 when $n \leq 633$ and 0.2435 when $n > 633$. It obviously shows that there is a long-term memory of the 3 min granularity size freeway average vehicle time occupancy time series. But, after 1899 min ($n > 633$), this long-term memory property disappears, and it displays anti-persistent property.

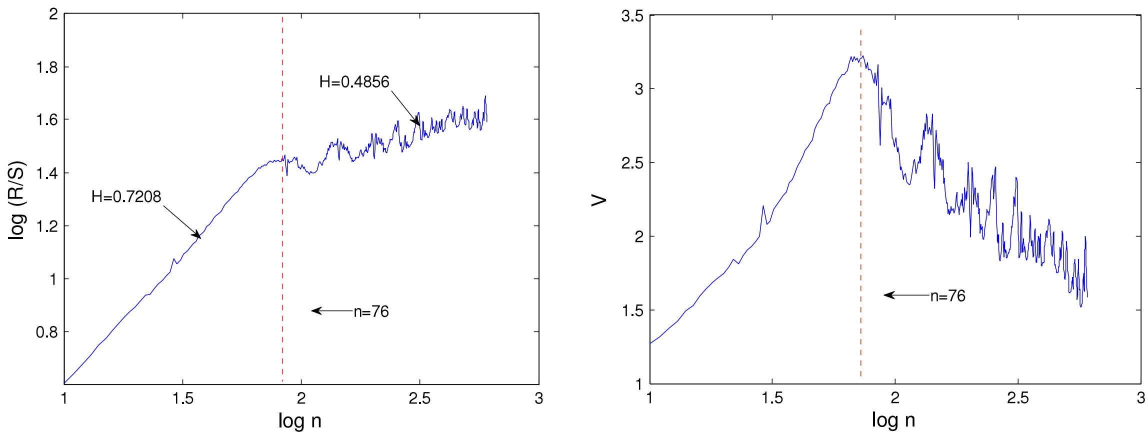
In summary, there is a long-term memory of 3 min groundway traffic flow time series (traffic flow velocity, average vehicles speed and average vehicles time occupancy), the memory cycle is about 31 h. In order to assure the reliability of the results, the regression results are examined. The test results are shown in Table 2. It indicates that the first and second



(a) Traffic flow velocity.



(b) Average vehicle speed.

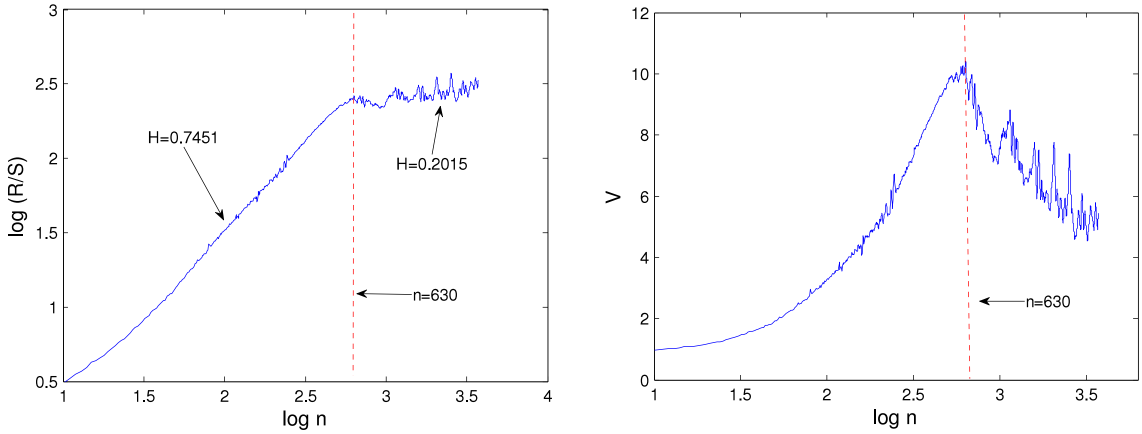


(c) Average vehicle time occupancy.

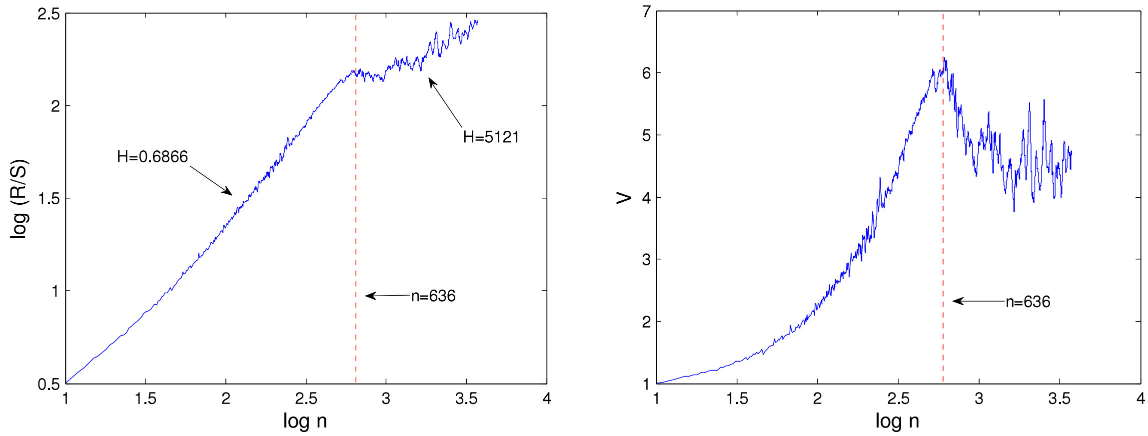
Fig. 4. The Hurst exponents and cycle of freeway traffic flow time series (25 min data).

stages of the Hurst value show significance, but it is not significant for the all stages. So we think the Hurst value analyzed above is reliable.

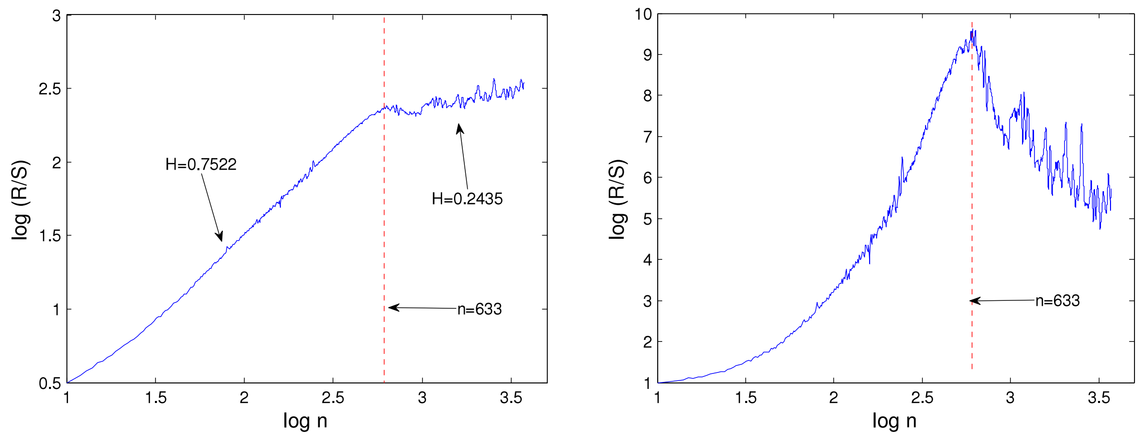
Fig. 6 provides us with the relationship between Hurst values and sequence interval n, which is derived by the following way. We conduct a regression analysis every time the value of n increases, with the regression starting point of 3 min and the increment of 3 min each, until a half of sequence length. We can see that the Hurst value of traffic flow velocity



(a) Traffic flow velocity.



(b) Average vehicle speed.



(c) Average vehicle time occupancy.

Fig. 5. The Hurst exponents and cycle of ground way traffic flow time series (3 min data).

reaches the maximum with $n = 630$ (1890 min), before which the Hurst value increases with the increase of n, while the trend becomes contrary after n is above 630 (1890 min). The aforementioned feature indicates that the long-term memory gradually increases until $n = 630$ (1890 min); however, it decreases gradually after $n = 630$ (1890 min) and nearly disappears after $n = 3000$. In the same way we can know the change trends of the relationship between the Hurst and

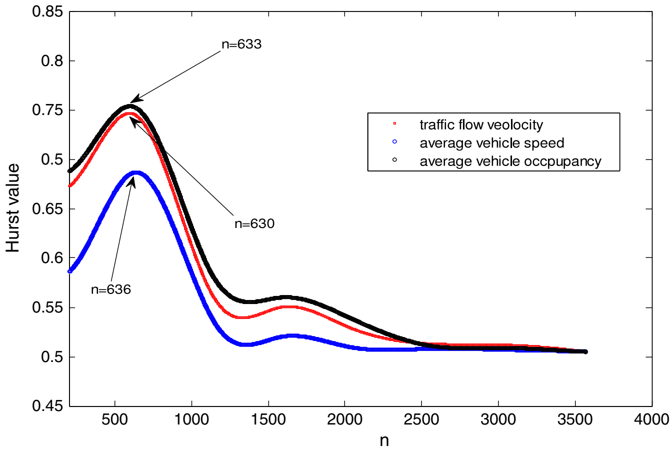


Fig. 6. The Hurst value of ground way traffic flow.

the other parameters (the average vehicles speed and average vehicles time occupation), which are $n = 636$ and $n = 633$ respectively. These conclusions are consistent with foregoing analysis results.

In Fig. 7(a), we depict the relationship between Hurst values and sequence interval n using the data of 15 min ground way traffic flow velocity. The results are extremely similar with that of 3 min. Both of them show almost identical Hurst values. The cycling period n is 126. The actual cycling period can be demonstrated to be 1890 (126×15) min (approximately 31 h). So it can be illustrated that the 3 min and 15 min ground way traffic flow velocity time series have the same long-term memory characteristics. In Fig. 7(b) and (c), the ground way traffic flow average vehicles speed and average vehicles time occupation of 15 min are analyzed respectively. In Fig. 7(b), the Hurst value and cycling period of average vehicles speed of 15 min are revealed to be 0.7015 and 127 (1905 min), only have 0.0193 and 3 min differences with that of 3 min. In Fig. 4(c), these two values are 0.7601 and 126 (1890 min), the differences are only 0.0079 and 9 min compared with that of 3 min.

According to above analysis, we know that the ground way traffic flow time series of 15 min have long-term memory too, their long-term characteristic is similar to that of 3 min. At the same time, another conclusion can be drawn from the above analysis that there exists scale-free property about ground way traffic flow time series.

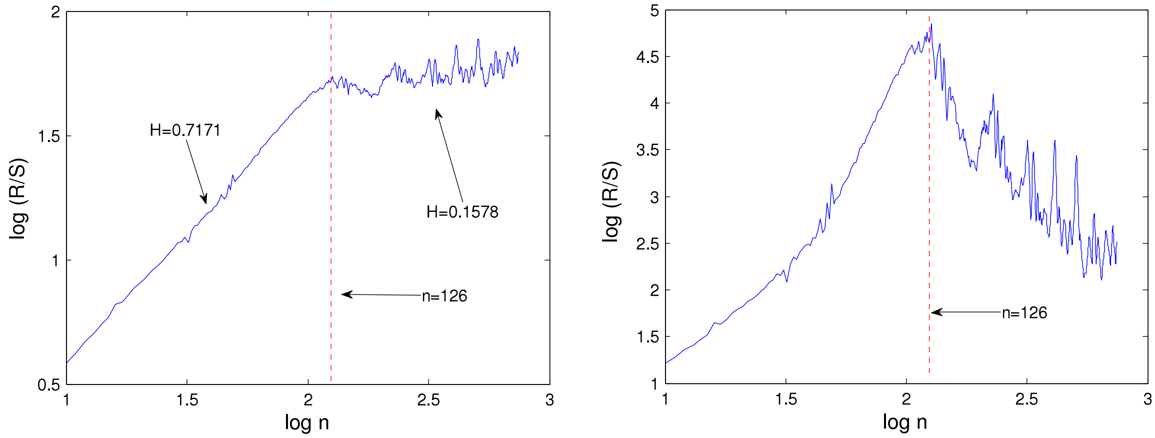
As before, we use another method to estimate the cycling period of the freeway traffic flow time series in order to verify the conclusions above. By the method of 2.3 we can get the value of ground way traffic flow time series $L_v = 0.00148$, $L_s = 0.00147$ and $L_o = 0.00145$. then we can get: $T_v = 1/L_v = 675$, $T_s = 1/L_s = 680$, and $T_o = 1/L_o = 689$. These results are comparatively approximate with the modified R/S analysis results.

5. Conclusion

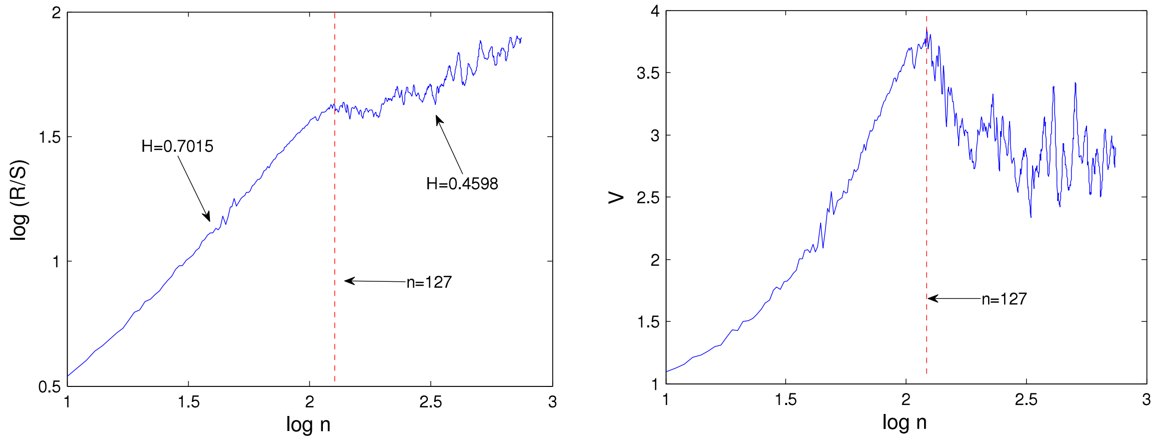
In this paper, we investigate how long the historical traffic flow time series keep efficacious to forecast future, which may provide a theoretical basis for the traffic flow data forecasting. In this frame, first, we collect the traffic flow time series data (include three types of parameters: traffic flow velocity, average vehicles speed and the average vehicles time occupancy) of freeway and ground way with different granularity, 5 min/25 min for freeway and 3 min/15 min for ground way. Then, using the modified R/S analysis method, we test for long memory property of the traffic flow time series by computing the Hurst exponent, our results suggest that in some time scale, all of the Hurst exponent values of the parameters of the freeway and ground way are greater than 0.5, it means the freeway and ground way traffic flow time series have long-term memory property really. Second, we calculate the long-term memory cycle and test the significance of the long-term memory property, the results show that in the time scale, the traffic flow time series long-term memory property is significant, and the analysis of the modified R/S results is reliable. Third, we compare the long-term memory cycle by modified R/S analysis and the maximum Lyapunov exponent value (L) of the traffic flow time series, we find the $1/L$ is so close to the memory cycle. This again reflects the rationality of the modified R/S analysis results.

Moreover, we find either the 5 min/25 min time scale of the freeway traffic flow time series or the 3 min/15 min time scale of the ground way traffic flow time series displays similar long-term memory property, this means all of them have the property of scale invariance, and the traffic flow time series has significant fractal property. The long-term memory cycle of three parameters of freeway traffic flow and ground way traffic flow time series is all about 30 h, it suggests both of the freeway traffic flow and ground way traffic flow historical time series data will keep efficacious for about 30 h. In other words, if we will use the data t_0 to forecast the traffic flow at time $t_0 + t$, the maximum t-value is about 30 h.

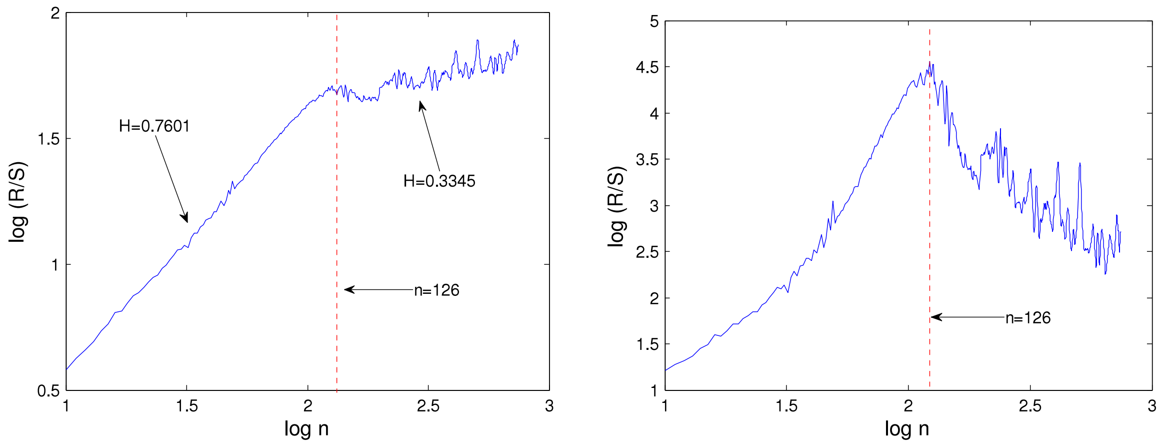
While promising, the results of this paper should be considered as preliminary and their general applicability, at this point in time, is not guaranteed. However, to be of any practical value, this intuitive observation has to be converted into a precise quantitative indicator. This problem required further investigation, both experimental and theoretical.



(a) Traffic flow velocity.



(b) Average vehicle speed.



(c) Average vehicle time occupancy.

Fig. 7. The Hurst exponents and cycle of ground way traffic flow time series (15 min data).

Acknowledgments

This research has been supported by Shanghai Natural Science Funds (No. 15ZR1429200), Doctoral Fund of USST (No. BSQD201407). We would like to thank the editors and anonymous reviewers for their valuable support and comments.

References

[1] P. Shang, Y. Lu, S. Kama, The application of Hölder exponent to traffic congestion warning, Physica A 370 (2) (2006) 769–776.
[2] P. Shang, J. Yu, Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis, Chaos Solitons Fractals 36 (1) (2008) 82–90.
[3] B.B. Su, H. Chang, Y.Z. Chen, et al., A game theory model of urban public traffic networks, Physica A 379 (1) (2007) 291–297.
[4] B. Shen, Z.Y. Gao, Dynamical properties of transportation on complex networks, Physica A 387 (5) (2008) 1352–1360.
[5] J.J. Wu, H.J. Sun, Z.Y. Gao, Dynamic urban traffic flow behavior on scale-free networks, Physica A 387 (2–3) (2008) 653–660.
[6] J.J. Wu, H.J. Sun, Z.Y. Gao, Long-range correlations of density fluctuations in the Kerner–Klenov–Wolf cellular automata three-phase traffic flow model, Phys. Rev. E 78 (3 Pt 2) (2008) 417–423.
[7] R. Song, C. Bian, Q.D.Y. Ma, Multifractal analysis of heartbeat dynamics during meditation training, Physica A 392 (8) (2013) 1858–1862.
[8] J. Thollot, C.E. Zair, E. Tosan, et al., Modeling fractal shapes using generalizations of IFS techniques, in: Fractals in Engineering, 1997.
[9] J. Wu, H. Sun, D.Z.W. Wang, et al., Bounded-rationality based day-to-day evolution model for travel behavior analysis of urban railway network, Transp. Res. C 31 (2) (2013) 73–82.
[10] Y. Wang, C. Wu, Y. Wei, Can GARCH-class models capture long memory in WTI crude oil markets? Econ. Model. 28 (3) (2011) 921–927.
[11] Y.Y. Zhang, W.U. Zheng, M.M. Guo, Time series analysis for the measured data of traffic flow, J. Fudan Univ. 50 (6) (2011) 767–2206.
[12] E.I. Vlahogianni, M.G. Karlaftis, J.C. Golias, Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume, Transp. Res. C 14 (5) (2006) 351–367.
[13] J. Tang, Y. Wang, H. Wang, et al., Dynamic analysis of traffic time series at different temporal scales: A complex networks approach, Physica A 405 (2014) 303–315.
[14] W. Shi, P. Shang, J. Wang, Large deviations estimates for the multiscale analysis of traffic speed time series, Physica A 421 (2015) 562–570.
[15] D.S. Dendrinos, Traffic-flow dynamics: A search for chaos, Chaos Solitons Fractals 4 (4) (1994) 605–617.
[16] D.J. Low, P.S. Addison, A nonlinear temporal headway model of traffic dynamics, Nonlinear Dynam. 16 (2) (1998) 127–151.
[17] D.S. Dendrinos, Traffic-flow dynamics: A search for chaos, Chaos Solitons Fractals 4 (4) (1994) 605–617.
[18] T. Nagatani, Chaos and dynamical transition of a single vehicle induced by traffic light and speedup, Physica A 348 (2005) 561–571.
[19] T. Nagatani, Self-similar behavior of a single vehicle through periodic traffic lights, Physica A 347 (2005) 673–682.
[20] T. Nagatani, Vehicular traffic through a self-similar sequence of traffic lights, Physica A 386 (1) (2007) 381–387.
[21] Y. Wang, L. Liu, Is WTI crude oil market becoming weakly efficient over time?: New evidence from multiscale analysis based on detrended fluctuation analysis, Energy Econ. 32 (5) (2010) 987–992.
[22] X. Lin, F. Fei, Y. Wang, Analysis of the efficiency of the Shanghai stock market: A volatility perspective, Physica A 390 (20) (2011) 3486–3495.
[23] Y. Huang, A study on fractal structure in chinese stock market—theoretical & empirical discussion, XiaMen University, 2004.
[24] A.W. Lo, Long-term memory in stock market prices, Social Science Electronic Publishing 59 (5) (2007) 1279–1313.
[25] E.E. Peters, A chaotic attractor for the S&p 500, Financ. Anal. J. 47 (2) (1991) 55–81.