Vol 16 No 2, February 2007
1009-1963/2007/16(02)/0365-09

Chinese Physics

© 2007 Chin. Phys. Soc.
and IOP Publishing Ltd

Multi-fractal analysis of highway traffic data*

Shang Peng-Jian(商朋见)[a)†] and Shen Jin-Sheng(申金升)[b)‡]

[a)] *Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing* 100044, *China*

[b)] *School of Traffic and Transportation, Beijing Jiaotong University, Beijing* 100044, *China*

(Received 20 April 2006; revised manuscript received 7 August 2006)

The purpose of the present study is to investigate the presence of multi-fractal behaviours in the traffic time series not only by statistical approaches but also by geometrical approaches. The pointwise Hölder exponent of a function is calculated by developing an algorithm for the numerical evaluation of Hölder exponent of time series. The traffic time series observed on the Beijing Yuquanying highway are analysed. The results from all these methods indicate that the traffic data exhibit the multi-fractal behaviour.

Keywords: fractals, probability theory, stochastic processes and statistics
PACC: 0555, 0250

1. Introduction

It is well known that the geometrical complexity of a 'fractal' set may be described, at least in a global way, by giving its dimension. In order to provide more detailed information, multi-fractal analysis is concerned with describing the local singular behaviour of measures, distributions, or functions in geometrical and statistical fashions. It was first introduced in the context of turbulence,[1,2] and then studied as a mathematical tool[3−6] in many applications such as the investigation of diffusion limited aggregation (DLA) pattern,[7] earth quake distribution analysis,[8] signal processing,[9,10] condensed matter,[11,12] and road traffic analysis.[13]

Transportation systems are complex entities.[14−16] Their current state and future evolution depend greatly on many properties of interactions, often highly variable physical and human elements. The investigation of the existence of a fractal nature in traffic flow has gained considerable attention in recent years. Prigogine and Robert [17] modelled traffic integrating statistical mechanics with individual choice behaviour and showed that it exhibited a high degree of complexity. Disbro and Frame [14] demonstrated how the theoretically derived, Gazis, Herman and Rothery (GHR) traffic model[15] was highly chaotic, even when applied to small (eight-car) systems. Zuylen *et al* [18] discussed the implications of human behaviour, chaos and unpredictability for urban and transportation planning and forecasting. Safanov *et al* [19] showed that fractal behaviour in traffic might be due to the delays in human reaction. Shang and Li [20] have indicated that fractal and chaotic characteristics exist in the traffic system and the techniques based on phase space dynamics can be used to predict the traffic speeds.

The results of such studies, on the one hand, nearly unanimously indicated the presence of a fractal nature in traffic flow and, hence, the possibility of transformation of traffic data from one scale to another. On the other hand, most of the studies have also revealed the insufficiency of the mono-fractal approaches to characterizing the highly irregular traffic flow. Therefore, we are motivated to employ the multi-fractal approaches directly, to identify fractals and, if so, to ascertain whether its type is mono-fractal or multi-fractal. In Ref.[13], which is most closely related to the present work, Vojak *et al* studied the structure of road traffic with the help of multi-fractal tools. They predicted that, under some circumstances, the traffic could possess a multi-fractal structure similar to the ones of multiplicative processes, and verified this behaviour on six-minute data of traffic flows.

The purpose of the present study is to investigate

*Project supported by the State Key Development Program for Basic Research of China (Grant No 2004CB318005) and Beijing Jiaotong University (Grant No 2005SM065).

†Corresponding author. E-mail: pjshang@center.njtu.edu.cn
‡jshshen@center.njtu.edu.cn

the presence of multi-fractal behaviours in the traffic time series not only by statistical approaches but also by geometrical approaches. Especially, we calculate the pointwise Hölder exponent of a function by developing an algorithm for numerical evaluation of Hölder exponent of time series, which becomes our primary quantitative tool. The traffic time series observed on the Beijing Yuquanying highway are analysed, and the study deals with the multi-fractal analyses only in time.

The necessity of a good understanding of the traffic is essential. In fact, multi-fractal analysis can be used for classifying signals that exhibit a rough behaviour.[13] This behaviour is quantified by calculating pointwise Hölder exponents using multi-fractal analysis. Indeed, multi-fractal behaviours in the traffic allow us to classify the highly complex signals of traffic flows, and give some perspectives for solving the difficult problem of nonlinear traffic prediction. If one wants to make robust forecasting, the first important thing to do is identify the structure of the process; if it is approximately linear, then a linear method (for instance an autoregressive one) can be helpful. On the contrary, if it is (multi-)fractal, completely different types should be used. One should further be able to distinguish among server nonlinear behaviours, since it is likely that the more irregular the process is, the more "nonlinear" the methods should be. In other words, one has to quantify it by how much the process departs from a linear one, in order to determine whether linear methods with small corrections can be used, or what sort of more complex model is needed.

The rest of this paper is organized as follows. In Section 2, we discuss the methods employed in the present study to identify the presence of multi-fractals in a time series. The details of the data used, analyses carried out and results achieved are presented in Section 3. Section 4 gives some remarks, and is devoted to introducing other multi-fractal analysis techniques, including the empirical probability distribution function, to verify multi-fractality of road traffic. Finally, important conclusions drawn from this study are provided in Section 5.

2. Proposed methods

Multi-fractal analysis of a set of data can be performed in two different ways: the statistical and the geometrical. The statistical approach is to define an appropriate intensive variable depending on a resolution parameter, and then calculate its statistical moments by averaging over an ensemble realized and at random base points. It is said that the variable is multi-fractal if those moments exhibit power-law dependence on the resolution parameter. On the other hand, the geometrical approach is to assess a local power-law dependence on the resolution parameter for the same intensive variables at each particular point.[19]

While the geometrical approach is informative in the spatial localization of self-similar (fractal) structures, it has been much less used because of the greater technical difficulty in retrieving the recent scaling exponents. However, in the latest years an important effort has been made to improve the sensibility and performance of the geometrical technique has.[19−24] In fact, we will apply the geometrical approach in this paper as a valuable tool to understanding the geometry and dynamics of traffic time series.

2.1. The multi-fractal formalism

In this section, the notion of the multi-fractal spectrum is introduced along with the multi-fractal formalism that is essential in this context. See also Refs.[13, 23–28].

Let μ be a Borel probability measured in a range [0,1]. To show what the multi-fractal spectrum is, we consider a normalized time range $I = [0, 1]$ and divide it into ν_n equal intervals with the ith interval being

$$I_{i,n} = \left[\frac{i}{\nu_n}, \frac{i+1}{\nu_n}\right], \quad i = 0, 1, \ \ldots, \ \nu_n - 1,$$

where ν_n is an ascending sequence of positive integer. The approximate scaling exponent for the ith interval is

$$\alpha^i_{\nu_n} = \frac{\log \mu(I_{i,n})}{-\log \nu_n}. \tag{1}$$

The exact scaling exponent at $x \in I_{i,n}$ can be obtained as

$$\alpha(x) = \lim_{\frac{i}{\nu_n} \to x} \alpha^i_{\nu_n}. \tag{2}$$

Obviously, $\alpha(x)$ may take many different values in the whole range of I. Define

$$P^i_n(\alpha, \varepsilon) = \begin{cases} 1, & \alpha^i_{\nu_n} \in (\alpha - \varepsilon, \alpha + \varepsilon), \\ 0, & \text{otherwise}, \end{cases}$$

Let

$$N_n(\alpha, \varepsilon) = \sum_{i=0}^{\nu_n - 1} P^i_n(\alpha, \varepsilon),$$

where $N_n(\alpha,\varepsilon)$ is the number of $\alpha^i_{\nu_n}$ taking values within $(\alpha-\varepsilon,\alpha+\varepsilon)$. The (grained) multi-fractal spectrum is defined as

$$f_{\mathrm{G}}(\alpha)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\log N_n(\alpha,\varepsilon)}{\log\nu_n}, \tag{3}$$

So $f_{\mathrm{G}}(\alpha)$ is the frequency that $\alpha(t)$ takes the value of α. It is formally called the large deviation multi-fractal spectrum.

In practice, a more numerically accessible multi-fractal spectrum, the Legendre spectrum, is often used as a substitute. Let us introduce this notion step by step.

Firstly, we calculate the following statistical moments:

$$M(n,q)=\sum_{i=0}^{\nu_n-1}[\mu(I_{i,n})]^q, \tag{4}$$

where $\sum$ means that the summation runs over those indices i such that $\mu(I_{i,n})\neq 0$. Secondly, define partition function as

$$\tau(q)=\lim_{n\to\infty}\frac{\log M(n,q)}{-\log\nu_n}. \tag{5}$$

We shall say that μ has a multi-fractal behaviour if the $\tau(q)$ curve departs from a straight line. Since $\tau(q)$ is obtained by averaging, it depends more regularly on the data than $f_{\mathrm{G}}(\alpha)$ and is easier to compute. Consequently, we define the Legendre multi-fractal spectrum as follows:

$$f_{\mathrm{L}}(\alpha)=\inf_{q\in R}(q\alpha-\tau(q)). \tag{6}$$

This spectrum is sometimes referred to the spectrum obtained by the method of moments.

Mathematically $f_{\mathrm{G}}(\alpha)\leq f_{\mathrm{L}}(\alpha)$; when $\tau(q)$ exists and is differentiable with respect to all real values of q, $f_{\mathrm{G}}(\alpha)=f_{\mathrm{L}}(\alpha)$ holds.

Moreover, $\tau(q)$ is related to a notion of generalized dimensions. Indeed, if we define

$$D_q=\frac{\tau(q)}{q-1},\quad q\neq 1, \tag{7}$$

$$D_1=\lim_{q\to 1}\frac{\tau(q)}{q-1}, \tag{8}$$

then D_0 is the fractal dimension of the support of μ, D_1 is the information dimension, D_2 the correlation dimension, etc.

Set $I_n(x)=\{I_{i,n}/x\in I_{i,n}\}$. We define

$$E_\alpha=x\in(0,1)/\lim_{n\to\infty}\frac{\log\mu(I_n(x))}{-\log\nu_n}=\alpha. \tag{9}$$

E_α can then be seen as the subset of points having the same scaling behaviour, described by α. The geometry of the singular distribution μ can then be characterized by giving the 'size' of the set E_α, more precisely their Hausdorff dimension[19]

$$f_{\mathrm{H}}(\alpha)=\dim_{\mathrm{H}}(E_\alpha).$$

This definition is most useful in purely mathematical settings. Under very general assumptions, it has been proven that (see Ref.[14]) $f_{\mathrm{H}}(\alpha)\leq f(\mathrm{G})(\alpha)\leq \mathrm{f_L}(\alpha)$. For certain special classes of measures, including multinomial measures, we have equality $f_{\mathrm{H}}(\alpha)=f_{\mathrm{L}}(\alpha)$. For the interested reader we refer to Refs.[5,6,27] for further details.

2.2. The Hölder exponent method

The basic numerical expression used in the multi-fractal analysis is the so-called Hölder exponent. Essentially, it is the parameter α, defined in the expression (2) of Section 2.1. However, this parameter is more sensitive to noise,[22,6] since it is computed independently at each point. Moreover, such an analysis is complicated by the fact that the traffic flow is both irregular and non-stationary. Here we focus on the pointwise Hölder exponent of a function, which becomes our primary quantitative tool.

The Hölder exponent of a function is a local characteristic value calculated at each point where the function is defined. It reflects the decay rate of the amplitude of the function fluctuation in the neighbourhood of that point as the size of the neighbourhood shrinks to zero. A highly irregular point (or singularity) in a function is characterized by a lower value of Hölder exponent and a smoother portion of a function will have a higher value of Hölder exponent.[28] A function may have different Hölder exponents at different points due to a variation in the local degree of irregularity (or singularity).

A function f defined in a neighbourhood of a point x_0 has a Hölder exponent α at x_0 if and only if (see Ref.[29])

(a) for every real $\gamma<\alpha$: $\lim\limits_{h\to 0}\dfrac{|f(x_0+h)-P(h)|}{|h|^\gamma}=0$ and

(b) if $\alpha<+\infty$, for each real $\gamma>\alpha$: $\limsup\limits_{h\to 0}\dfrac{|f(x_0+h)-P(h)|}{|h|^\gamma}=+\infty$,

where P is a polynomial whose degree is less than or equal to α. We denote the Hölder exponent of a function at x as $H_f(x)$. It is easy to see that $H_f(x_0)>1$

implies that the function $f(x)$ is differentiable at x_0, while the function $f(x)$ is differentiable at x_0, which implies $H_f(x_0) \geq 1$. For instance, in the extreme case when the function is constant near x, the definition of the Hölder exponent clearly implies that function $H_f(x) = +\infty$. A non-constant linear function $f(x) = ax + b$ with $a \neq 0$ has Hölder exponent 1.

The value of Hölder exponent $H_f(x) = \alpha$ strictly below 1 indicates that the function is not differentiable at the given point and the amplitude of its oscillations within ε-neighbourhood of the point scales as ε^α for $\varepsilon \to 0$. Classical fractal functions (for example the Weierstrass function) are strictly self-similar, i.e. their singularities have the same scaling at each point. It is generally believed[30] that, however, in the collected data set, the Hölder exponent stays within the range (0,1) that characterizes a non-differentiable function.

It is well known that the Hölder exponent at a point x of function $f(x)$ can be expressed as[29]

$$H_f(x) = \liminf_{h \to 0} \frac{|f(x+h) - f(x)|}{\log(|h|)}. \tag{10}$$

It should be noted that in the definition and formula for the Hölder exponent given above, f is assumed to be a function of continuous variable x. In the experimental context however we deal with discrete time series, and therefore the numerical computation of Hölder exponent presents a problem. In the following, we develop an algorithm for numerical evaluation of Hölder exponent of a discrete function based on the expression (10).

Suppose that $\{y_0, y_1, y_2, \ \ldots, \ y_n\}$ is a time series of length $n+1$ representing some quantity measured in uniform intervals. The idea of this algorithm is that the degree to which the function has the strength of its Hölder exponent at a certain data point y_i is a function of a number of data points preceding and following it. This number, s, is called the window width and is not fixed but rather is a parameter set by the user when running the algorithm. The weight that each of the $2s$ values (there exist s values on each side of the data point) has on the ultimate calculation of the Hölder exponent is controlled by another parameter λ. The λ is called the weighted regression coefficient and must satisfy $0 < \lambda < 1$. Adjusting the strength of λ will allow us to adjust the degree to which the data points further depart from the point we are interested in and influence the calculation of the Hölder exponent at that point. Obviously points nearby should have a greater weight than distant points.

To calculate the Hölder exponent at the ith point in the time series, we begin by choosing a window width, s, and a value of λ (we use $\lambda = 0.5$ and $s = 10$ for all the Hölder exponent calculations appearing in Section 3.2).

Next, we calculate a value, $R_{i,k}$, for each integer $k \neq 0$ from $-s$ to s in the following manner:

$$R_{i,k} = \frac{\log\left(\frac{|y_{i+k} - y_i|}{200}\right)}{\log\left(\frac{|k|}{1000}\right)}. \tag{11}$$

In the above case, k must satisfy $0 \leq i + k \leq n$. For data points at the very beginning or end of the time series for which s values to the left or right are not available, s is shrunk to an appropriate size. The numbers "200" and "1000" in the expression (11) are only the normalizing parameters which are chosen for the convenience of computation.

Then, in the second step, for each integer j, $1 \leq j \leq s$, we calculate

$$h_{i,j} = \min\{R_{i,k} : |k| \leq j\}, \tag{12}$$

which corresponds to taking the lim inf in the equation for the Hölder exponent.

In the third step, we calculate H_i, i.e. estimate the Hölder exponent at the ith data point by computing the weighted average of the approximations $h_{i,j}$:

$$\begin{aligned} H_i = \ & \frac{1-\lambda}{1-\lambda^s}(h_{i,1} + \lambda h_{i,2} + \lambda^2 h_{i,3} \\ & + \cdots + \lambda^{s-1} h_{i,s}), \end{aligned} \tag{13}$$

where more weights are given to terms with smaller values of j (i.e. h is close to 0 in the expression (10)).

We can test our algorithm of calculating the Hölder exponent by using the so-called generalized Weierstrass functions,[28] whose Hölder exponent can be predetermined and then compared with the result obtained using this numerical algorithm. The generalized Weierstrass function is given by

$$f(x) = \sum_{k=0}^{\infty} 3^{-ks(x)} \sin(3^k x), \tag{14}$$

where $s(x)$ is the seed function ranging between 0 and 1. It is known[29] that $s(x) = H_f(x)$ for all x. Figure 1 shows a comparison of our estimated Hölder exponent with the predetermined Hölder exponent $s(x)$ for the case $s(x) = |\sin(5\pi x)|$. It demonstrates that the Hölder exponent estimated by our algorithm is almost consistent with the predetermined Hölder exponent.

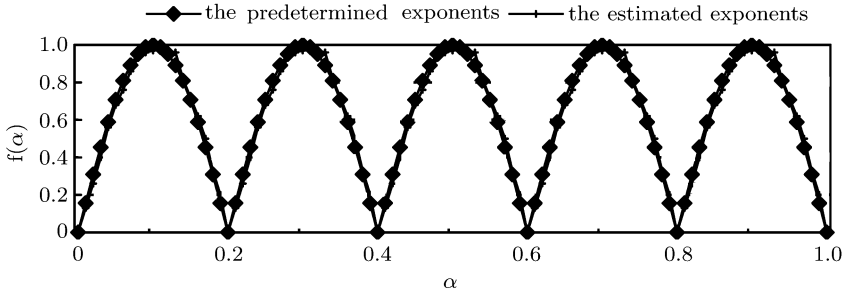


Fig.1. A comparison of our estimated Hölder exponent with the predetermined Hölder exponent.

3. Results and analyses

3.1. Traffic data

We use the data observed on the Beijing Yuquanying highway over a period of about 40 months, from 16 January 2001 to 17 June 2004. During this period the traffic data set was chosen not to be biased due to road constructions or bad weather conditions. The data were downloaded from the Highway Performance Measurement Project (FPMP) run by Beijing STONG Intelligent Transportation System CO. LTD, Beijing. The raw data for speed, volume and occupancy were collected every twenty seconds for each lane of the instrumented highway locations. Then these data were aggregated into one-minute data for average speed (i.e. average speed in one minute). We analyse mainly speed data under free flow conditions (that amounts to about 85% of all data). Similar results have been obtained also from volume and occupancy measurements.

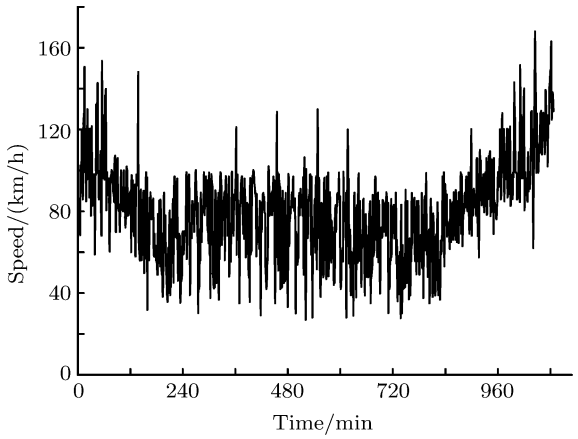


Fig.2. Time series plot of the speed data from Beijing Yuquanying highway (from 5:30 AM to 11:30 PM on12 January 2004)

Figure 2 shows the time series plot of the speed data observed at the Beijing Yuquanying highway (from 5:30 AM to 11:30 PM on 12 January 2004). A visual inspection of the speed series indicates only significant variations in speed, but this is not enough to say anything regarding the existence of fractals. Hence, additional tools are necessary to identify the possible presence of fractals.

3.2. Statistical moment scaling method

To investigate the scaling behaviour of the time series, the variation of empirical moments with scale is examined. This analysis is a common technique to study scaling properties of data[29] and has been applied to hydrography by, for example, Over and Gupta.[31] Among the existing multi-fractal identification methods, the statistical moment scaling method[2,31] is the most widely used one in time series. In this method, the time series is divided into non-overlapping intervals with a certain time resolution. The ratio of the maximum scale of the field to this interval is termed the 'scale ratio' k. Thus k is inversely proportional to the size of the scale examined. For different values of scale ratio k, we compute the average intensity in each interval i, and denote it by $\mu(k,i)$ as follows:

$$\mu(k,i) = \frac{N(k,i)}{S}, \tag{15}$$

where $N(k,i)$ is the measurement of the quantity in time interval i, S is the sum of all the values in the time series. Raising $\mu(k,i)$ to a power q, we obtain the statistical moments

$$M(k,q) = \sum_i [\mu(k,i)]^q. \tag{16}$$

For a scaling field, the moment $M(k,q)$ is related to the scale ratio k by

$$M(k,q) \propto k^{\tau(q)}, \tag{17}$$

where $\tau(q)$ may be regarded as a characteristic function of the fractal behaviour. If $\tau(q)$ versus q is a

straight line, the data set is mono-fractal. On the other hand, if $\tau(q)$ versus q is a convex (concave) function, the data set is multi-fractal.[2,25,26]

The validity of the expression (17) may be directly tested on the speed time series. By plotting $M(k,q)$ as a function of k in a log–log diagram, the scaling behaviour of the time series as expressed by the expression (17) may be investigated. If the expression (17) is valid the curve will exhibit an approximately linear behaviour with a slope that is an estimate of $\tau(q)$. By performing the same procedure for different values of q, the whole $\tau(q)$ function may be estimated. Owing to uncertainties in the estimation of high-order moments,[32] the investigation is limited to orders in the range $0.2 \leq q \leq 2.6$. Figure 3 shows the $M(k,q)$ as a function of k in a log–log diagram for the speed time series (5:30 AM on 16 January 2001 through 11:30 PM on 17 June 2004) observed on the Beijing Yuquanying highway for $0.2 \leq q \leq 2.6$. The curves in Fig.3 exhibit a well-defined straight-line behaviour over the whole scaling region for $0.2 \leq q \leq 2.2$. However, the higher-order moments do not follow a straight line for the time series. In addition, it can be noted that the higher the order of the moment, the less linear the character of $M(k,q)$ is.

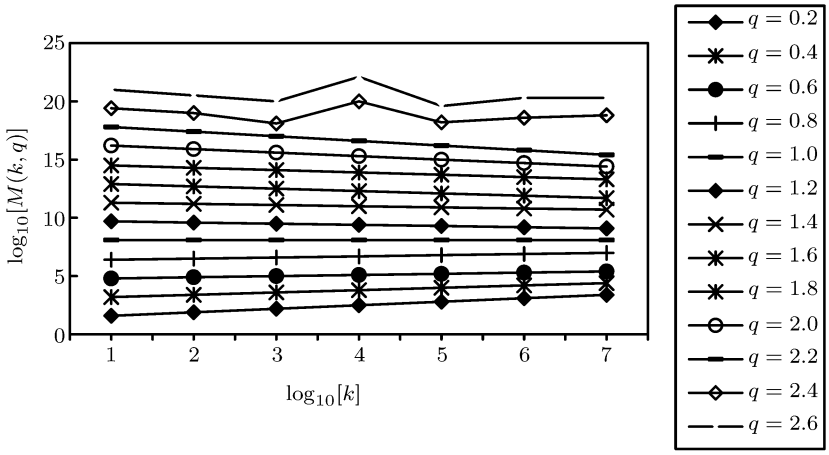


Fig.3. The qth statistical moment, $M(k,q)$, as a function of the scale ratio, k, in the scaling regime for the speed time series from the Beijing Yuquanying highway (5:30 AM on 16 January 2001 through 11:30 PM on 17 June 2004).

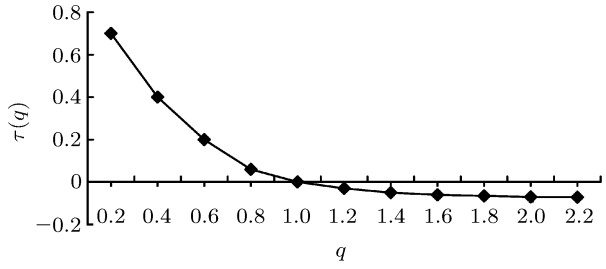


Fig.4. Estimated function $\tau(q)$ for the speed time series from the Beijing Yuquanying highway.

The relationship between $\tau(q)$ and q is shown in Fig.4. The function $\tau(q)$ is estimated from the slope of the fitted regression line for the traffic time series. The function $\tau(q)$ is shown for moment orders up to 2.2 for the speed time series. As seen in the figure, the $\tau(q)$ versus q is a concave curve, rather than a straight line, indicating that the traffic data are multi-fractal. The function $\tau(q)$ is slightly curved for $0.2 \leq q \leq 2.2$ but becomes more linear as q increases.

3.3. Hölder exponent method

The Hölder exponent series obtained for the speed time series observed over a period of 18 h (5:30 AM to 11:30 PM on 12 January 2004) on the Beijing Yuquanying highway is shown in Fig.5. Even a cursory glance at the Hölder exponent plot depicting the temporal dynamics of the traffic system can find that the speed time series displays little singularity for the first 120 min or so (5:30 AM to 7:30 AM), and the bulk of the Hölder exponent values for that time period fall to between 0.8 and 1. This is followed by an outburst of singularity, which is captured by the Hölder exponent as it falls to between 0.3 and 0.6 for the period from 120 min to 230 min (7:30 AM to 9:20 AM). When the period of singular behaviour is over, the Hölder exponent rises again, which is between 0.5 and 0.8 (9:20 AM to 5:00 PM). However, in the following period from 690 min to 810 min (5:00 PM to 7:00 PM) the Hölder exponent falls to between 0.2 and 0.5.

Lastly, the chart shown in Fig.5 displays an increasing trend in the value of Hölder exponent (i.e. decrease in singularity) of the traffic dynamics, which varies from 0.6 up to 1 for the period from 810 min to 1080 min (7:00 PM to 11:30 PM).

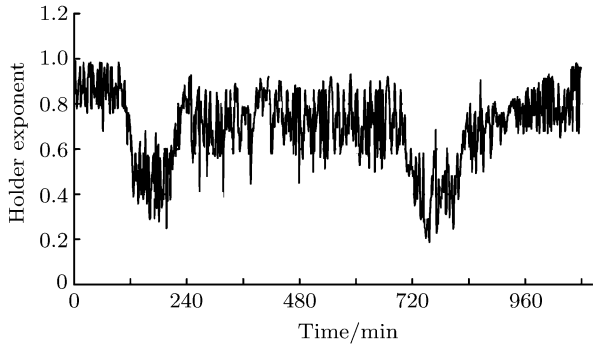


Fig.5. Hölder exponent series for the speed dada observed over a period of about 18 h, from 5:30 AM to 11:30 PM on 12 January 2004, at the Beijing Yuquanying highway.

Though the Hölder exponent series obtained is rather noisy (this is typical for experimentally estimated Hölder exponent values), certain characteristics can be distinguished. By a mere visual inspection of the plots, one can observe that the estimated Hölder exponent intensities are consistent with those in the different traffic periods in time (i.e. low hours, busy hours and normal hours). Moreover, the Hölder exponents for the speed time series display a marked change in magnitude, thereby enabling the detecting of the degree of singularity of the traffic flow. However, to be of any practical value, this intuitive observation must be converted into a precise quantitative indicator. This problem requires further investigation, both experimental and theoretical.

4. Remarks

Research over the past two decades has led to the development of a wide variety of methods to identify fractals in a time series.[27,33−37] The scaling of the statistical moments $M(k,q)$ is indeed a valid way to investigate multi-fractality. However, one should use other multi-fractal analysis technique, for example, the Legendre spectrum $f_{\mathrm{L}}(\alpha)$ and the large deviation spectrum $f_{\mathrm{G}}(\alpha)$, to verify multi-fractality of traffic flow. The former is more attractive multi-fractal spectrum from numerical point of view, since it is obtained through the Legendre transform: $f_{\mathrm{L}}(\alpha) = \inf_{q\in R}(q\alpha - \tau(q))$. Figure 6 gives the results for the Legendre spectrum $f_{\mathrm{L}}(\alpha)$ and the large deviation spectrum $f_{\mathrm{G}}(\alpha)$. Here we employ the Fraclab—a Matlab toolbox based on fractal and multi-fractal approaches, which features these methods. We see that $f_{\mathrm{L}}(\alpha)$ includes a rich set of α and the curve of $f_{\mathrm{L}}(\alpha)$ is nearly the same as that of $f_{\mathrm{G}}(\alpha)$.

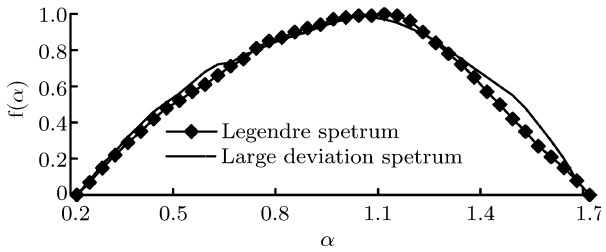


Fig.6. The Legendre spectrum $f_{\mathrm{L}}(\alpha)$ and the large deviation spectrum $f_{\mathrm{G}}(\alpha)$ for the traffic time series.

In addition, the general information about the multi-fractal behaviour may be obtained by using standard statistical descriptions—the empirical probability distribution function (PDF).[34,35,37] The PDF of a time series describes the fractal of the intensity thresholds of the time series fluctuations at a given scale, generally the scale corresponds to the measurement resolution. If, for high-intensity threshold values x, the tail of the probability distribution of the time series X follows a power law of form

$$\Pr(X > x) \propto x^{-D},$$

where D is the probability exponent, then the series is characterized by hyperbolic intermittency.[35] This is a general, but not necessary, feature of multi-fractal processes.[37] In fact, a value of $D < 2$ indicates that mono-fractal models may be sufficient to characterize the traffic behaviour, whereas multiplicative (and generally multi-fractal) models are necessary if $D > 2$.[34,35]

The empirical PDF obtained for the speed time series (5:30 AM on 16 January 2001 through 11:30 PM on 17 June 2004) observed on the Beijing Yuquanying highway is shown in Fig.7. The figure shows that the traffic data exhibit hyperbolic tail behaviour with a probability exponent D of about 10 estimated from the fitted regression line. Since the D value obtained from the traffic time series is greater than 2.0, a mono-fractal model is not sufficient to model the traffic behaviour, as it cannot accommodate a value of $D > 2$.

This suggests that a multi-fractal model, which can accommodate any value of D, is necessary to model the traffic behaviour.

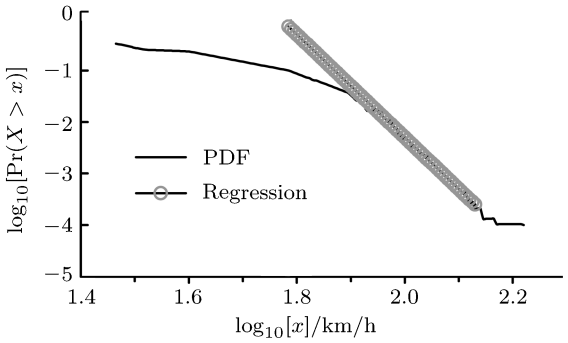


Fig.7. Empirical probability distribution function, $\Pr(X > x)$, for the speed time series (5:30 AM on 16 January 2001 through 11:30 PM on 17 June 2004) from the Beijing Yuquanying highway.

In conclusion, all these methods show that the traffic data are multi-fractal.

5. Conclusions

Application of the ideas gained from multi-fractal theory to characterizing time series has been one of the most active areas of research in recent years. This paper has presented a detailed analysis of multi-fractal properties of the highway speed time series. The data originated from millions of measurements collected on China highways. Our analysis aims to examine whether these data possess any fractal structure and whether its type is mono-fractal or multi-fractal. The existence of a multi-fractal nature is investigated using both the statistical approach and the geometrical approach. Especially, we have calculated the point-wise Hölder exponent of a function, by developing an algorithm of numerically evaluating the Hölder exponent of time series. The results from all these methods provide positive evidence regarding the existence of multi-fractal behaviour in the traffic process observed on the Beijing Yuquanying highway. Further, the present investigation indicates the insufficiency of a mono-fractal model and the necessity for a multi-fractal model to characterize the traffic behaviour.

Now that the multi-fractal structure of traffic flow has been demonstrated, several tasks lay ahead. First, exploring other essential statistical information such as short-term forecasting for traffic flow which would be useful. The knowledge of the whole multi-fractal spectrum can give more precise statistics than the knowledge of mono-fractal dimension only. Second, developing the models 'physically' relevant with multi-fractal properties which are needed for a better understanding of traffic flow. These issues will be addressed in forthcoming papers.

Acknowledgments

The authors wish to thank Beijing STONG Intelligent Transportation System Co. Ltd, Beijing (SITS) for providing the traffic data of the Beijing Yuquanying highway.

References

[1] Mandelbrot B B1974 *J. Fluid. Mech* **62** 331
[2] Frisch U and Parisi G 1985 *Proc. Int. Summer School on 'Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics'* (Amsterdam: North-Holland) p84
[3] Falconer K J 1994 *J. Theory. Prob.* **7** 681
[4] Brown G, Michon G and Peyriere J 1992 *J. Stat. Phys.* **66** 775
[5] Riedi R H 1995 *J. Math. Anal. Appl.* **189** 462
[6] Véhel L J and Vojak R 1998 *Adv. Appl. Mat.* **20** 1
[7] Evertsz C J G and Mandelbrot B B 1992 *Physica* A **185** 77
[8] Wu Z L 2001 *Chin. Phys.* **10** 23
[9] Zhang S, Liu H X, Gao D T and Du S D2003 *Chin. Phys.* **12** 14
[10] Véhel J L 1996 *Fractal Geometry and Analysis, the Mandelbrot Festschrift,* Curacao 1995, C.J.G. Evertsz, H.-O. Peitgen, R.F. Voss (eds.) (Singapore: World Scientific)
[11] Wu F M 2005 *Chin. Phys.* **14** 9
[12] Zhou G H, Pan L X and Yan J R 2001 *Chin. Phys.* **10** 689
[13] Vojak R and Véhel J L 1994 *IFAC'94,* Tianjin (China), August 24
[14] Disbro J E and Frame M 1989 *Transportation Res. Rec.* **1225** 109
[15] Gazis D C, Herman R and Rothery R W 1961 *Oper. Res.* **9** 545
[16] Nair A S, Liu J C, Rilett L and Gupta S 2001 *Non-Linear Analysis of Traffic Flow,* http://translink. tamu.edu /docs/Research/Linear Analysis Traffic Flow/ chaos1. PDF. Accessed May 27, 2003
[17] Prigogine I and Robert H 1971 *Kinetic Theory of Vehicular Traffic* (New York: Elsevier)
[18] Zuylen V, Henk J, Marina S, Geenhuizen V and Peter N 1999 *Transportation Res. Rec.* **1685** 21
[19] Safanov L A, Tomer E, Strygin V V, Ashkenazy Y, Havlin S and Safanov L A 2002 *Europhys. Lett.* **57** 151

[20] Shang P J and Li X W 2005 *Chaos, Solitons and Fractals* **25** 121

[21] Crovella M and Bestavros A1997 *IEEE/ACM Trans. on Networking* **5** 835

[22] Jaffard S 1999 *Probab. Theory Rel. Fields* **114** 207

[23] Riedi R H and Véhel J L 1997 *Inria Tech. Rep.* RR 3129

[24] Véhel J L and Sikdar B 2001 *a Multiplicative Multifractal Model for TCP Traffic. Proceedings of IEEE ISCC* 714

[25] Barral J and Véhel J L 2004 *Multifractality of TCP explained.* Inria Tech. Rep

[26] Barral J and Véhel J L 2004 *Ele. J. Prob.* **9** 508

[27] Falconer K 1990 *Fractal Geometry* (New York: Wiley)

[28] Jaffard S 1993 *Multifractal Formalism for Functions,* Ph.D. thesis, see also *Comptes Rendus de l'Acadmie des Sciences* **317** 745

[29] Daoudi K, Véhel J L and Meyer Y 1998 *J. Cons. Approxim.* **14** 349

[30] Véhel J L and Daoudi K 1996 *Generalized IFS for signal processing,* IEEE DSP Workshop

[31] Over T M and Gupta V K 1994 *J. Appl. Meteorol.* **33** 1526

[32] Kumar P P, Guttarp P and Foufoula G E 1994 *Stochastic Hydro. Hydraul.* **81** 73

[33] Kantz H and Schreiber T 1996 *Nonlinear Time Series Analysis* (Cambridge: Cambridge University Press)

[34] Tessier Y, Lovejoy S and Schertzer D 1993 *J. Appl. Meteorol.* **32** 223

[35] Lovejoy S Mandelbrot B 1985 *Tellus* **37**A 209

[36] Struzik Z R 2000 *Technical Report INS-R*0015, **C**WI, (Amsterdam, the Netherlands)

[37] Fraedrich K and Larnder C 1993 *Tellus* **45**A 289