Multiscale multifractal analysis of traffic signals to uncover richer structures

Jing Wang,[1,2] Pengjian Shang,[1,*] and Xingran Cui[2]

[1]*Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, People's Republic of China*

[2]*Division of Interdisciplinary Medicine and Biotechnology, Department of Medicine, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, Massachusetts 02215, USA*

(Received 13 June 2013; revised manuscript received 14 December 2013; published 24 March 2014)

Multifractal detrended fluctuation analysis (MF-DFA) is the most popular method to detect multifractal characteristics of considerable signals such as traffic signals. When fractal properties vary from point to point along the series, it leads to multifractality. In this study, we concentrate not only on the fact that traffic signals have multifractal properties, but also that such properties depend on the time scale in which the multifractality is computed. Via the multiscale multifractal analysis (MMA), traffic signals appear to be far more complex and contain more information which MF-DFA cannot explore by using a fixed time scale. More importantly, we do not have to avoid data sets with crossovers or narrow the investigated time scales, which may lead to biased results. Instead, the Hurst surface provides a spectrum of local scaling exponents at different scale ranges, which helps us to easily position these crossovers. Through comparing Hurst surfaces for signals before and after removing periodical trends, we find periodicities of traffic signals are the main source of the crossovers. Besides, the Hurst surface of the weekday series behaves differently from that of the weekend series. Results also show that multifractality of traffic signals is mainly due to both broad probability density function and correlations. The effects of data loss are also discussed, which suggests that we should carefully handle MMA results when the percentage of data loss is larger than 40%.

DOI: 10.1103/PhysRevE.89.032916 PACS number(s): 05.45.Tp, 05.10.−a, 89.40.−a, 05.45.Df

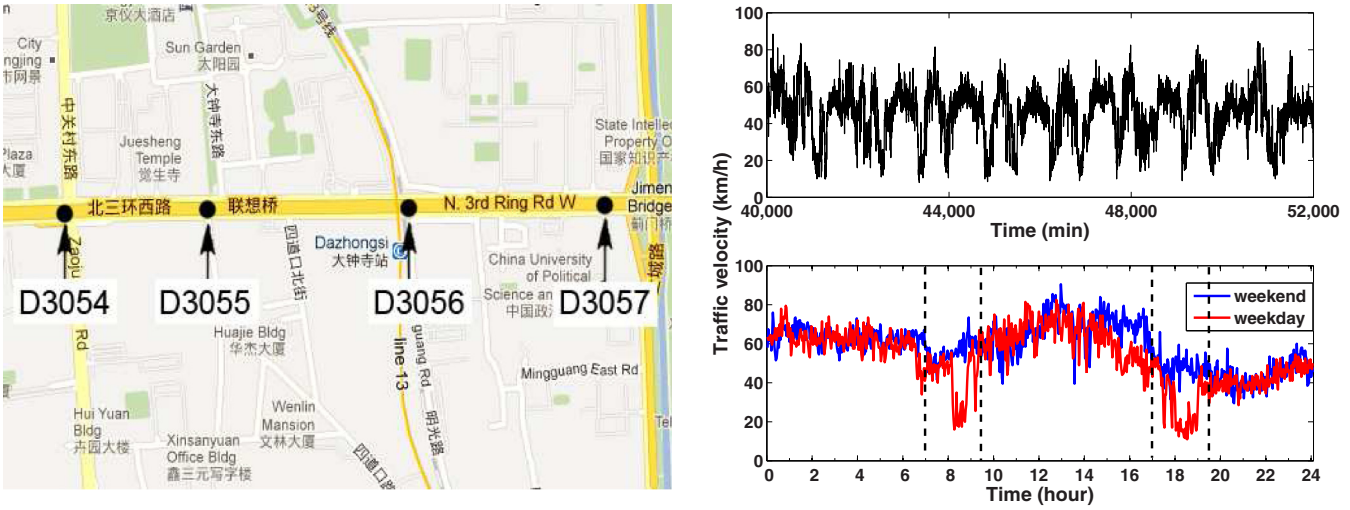
I. INTRODUCTION

Urban traffic congestion is one of the most serious challenges that affects all aspects of the quality of everyday life in metropolitan areas. Traffic systems, especially urban traffic systems, are enforced by well-defined periodicity, together with the existence of unpredictable perturbation and highly variable physical and human elements which may lead to extreme traffic events. Thus traffic flow often exhibits irregular and complex behavior [1–3]. The proper representation of all dynamics in a model is complex: while certain relationships can be developed through analysis, the incorporation of immeasurable quantities, such as laws and social codes, creates further complications [4]. Studies to understand the properties of traffic signals are an important area of research. There have been successful attempts by many researchers when analyzing or forecasting traffic signals in recent decades from different aspects [5–10]. Further, the fractal time series analysis of traffic signals is a developing research area [11–15], which helps to better understand the characteristics of traffic systems.

The simplest type of multifractal analysis is based on standard partition function multifractal formalism, which has been developed for the multifractal characterization of normalized, stationary measurements [16–19]. Unfortunately, this standard formalism may give incorrect results for nonstationary time series that are affected by trends or that cannot be normalized. Thus, in the early 1990s, an improved multifractal formalism was developed, i.e., the wavelet transform modulus maxima (WTMM) method [20], which is based on wavelet analysis and involves tracing the maxima lines in the continuous wavelet transform over all scales. The other method, the multifractal detrended fluctuation analysis (MF-DFA) [21] in the statistical physics field, has become a widely used technique to characterize multifractality of time series in the presence of nonstationarity. It is a generalization of detrended fluctuation analysis (DFA) [22], which was originally established to avoid spurious detection of long-range correlations that are artifacts of nonstationarity in the time series. MF-DFA, as a robust and powerful technique, identifies and quantifies the multiple scaling exponents within a time series. It has been successfully applied in different and heterogeneous scientific fields to study multifractality [14,23–30].

However, when analyzing the multifractality of time series, the well-known MF-DFA usually assumes a much wider time scale, i.e., the scale $s \in [10, N/4]$ [21,31], where N is the length of the series. One problem which may exist is the presence of crossovers in the log-log plot of fluctuation function $F_q(s)$ versus scale s of conventional MF-DFA, i.e., the change of average slope of the fluctuation functions at some scale s, which is commonly found in traffic series. A crossover is usually interpreted as a result of different correlation properties for small and large scales in the signals [32]. When analyzing time series with crossovers, researchers usually narrow the range of investigated scales only to large or only to small scales. A recent study highlighted that a single exponent is inadequate to describe the heart rate (HR) dynamics, showing that α calculated by DFA is poorly reproducible and does not reliably reflect the autonomic influences on HR [33]. A "two coefficient" model [34] has been proposed to estimate short-term and long-term scale coefficients, α_1 and α_2, but it has been viewed as an oversimplification of a more complex phenomenon. The common practice of fixing *a priori* scaling ranges in such methods as DFA and MF-DFA may lead to artifacts in some cases, i.e., if a crossover falls within the scaling range by

*Present address: Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, People's Republic of China.



(a) (b)

FIG. 1. (Color online) (a) The locations of detectors 3054–3057. (b) Part of the profiles of traffic velocity (Km/h) collected by detector 3057 (upper panel). One-day series at weekday and weekend separately (lower panel, black line for weekend, gray line for weekday in printed black-white version); dashed lines show the periods at rush hours: morning peak during 7:00–9:30 and evening peak during 17:00–19:30.

mistake, the results will be biased. In order to avoid mistakes due to improperly predefined scaling ranges, and to obtain all information about the fractal properties among the entire time scales, we applied the multiscale multifractal analysis (MMA) [32] to detect the multiscale structure of traffic signals.

In this study, we focus on the detection of the intriguing long-range correlation nature and multifractal properties of traffic signals, applying a modified MMA model. The structure of the paper is as follows. In Sec. II, we briefly introduce the traffic data used in this study, i.e., MF-DFA, MMA, and, finally, some modifications of MMA. In Sec. III, we employ MMA to analyze traffic speed signals, including a description of multiscale multifractal analysis related to conventional methods, using artificial monofractal and multifractal series. In addition, some discussions on the effects of periodic trends, weekday and weekend patterns, the generating mechanism of multifractality for traffic signals, and data loss on MMA results are included in this section. Finally, a summary is presented in Sec. IV.

II. DATA AND METHOD

A. Data description

Traffic systems have numerous parameters that can be measured, among which, traffic speed, volume, and occupancy are mostly studied. In this study, we use the data observed from detectors 3054 to 3058 which lie on the North 3rd Ring Road (Beijing, China) over a period of about 11 weeks, from August 11 to October 26, 2012. We choose the data from the North 3rd Ring Road as it is located in important areas of the economy, culture, and entertainment of Beijing; the traffic states in this area are closely related to daily life and receive more attention by the people in Beijing. The data were downloaded from the Highway Performance Measurement Project (HPMP) run by the Beijing STONG Intelligent Transportation System Co. Ltd. Here, we focus on the traffic speed time series. The raw data for speed are collected every 20 seconds at each detector location, which are located approximately every half a kilometer. Raw data are screened for errors and then aggregated into two-minute data for average speed. The one-hour recorded traffic speed time series are about 30 data points, and the total number of data points for the whole series is about 55 440.

Figure 1(a) shows the locations of the detectors, and Fig. 1(b) illustrates part of the traffic speed series (about 16 days, upper panel) and one weekday and weekend series (lower panel) collected by detector 3057. As MMA results are similar for the time series recorded by the four detectors, we mainly show the results for the time series collected by detector 3057 in the next section.

B. MFDFA method

Multifractal detrended fluctuation analysis (MF-DFA), a generalization of detrended fluctuation analysis (DFA), was developed by Kantelhardt *et al.* [21] for the multifractal characterization of nonstationary time series. Suppose that x_k is a series of length N; determine the profile $X(i) = \sum_{k=1}^{i} x_k$, where $i = 1, \ldots, N$, and then divide the profile $X(i)$ into $N_s = \mathrm{int}(N/s)$ nonoverlapping segments of equal length s. So the ith element within the vth segment is $X_v(i) = X[(v-1)s + i]$ for each segment $v = 1, \ldots, N_s$. Assume that the local trending functions of $\{X_v(i)\}$ are $\{\tilde{X}_v(i)\}$. Then determine the variance,

$$F^2(v,s) = \frac{1}{s}\sum_{i=1}^{s}\{X_v(i) - \tilde{X}_v(i)\}^2. \quad (1)$$

Average over all segments to obtain the qth order fluctuation function,

$$F(q,s) = \left\{\frac{1}{2N_s}\sum_{v=1}^{2N_s}[F^2(v,s)]^{q/2}\right\}^{1/q}. \quad (2)$$

Finally, determine the scaling behavior of fluctuation functions by analyzing log-log plots of $F(q,s)$ versus s for each value of

q. If series x_k is long-range power-law correlated, then $F(q,s)$, for large values of s, increases as a power law,

$$F(q,s) \sim s^{h(q)}. \tag{3}$$

The generalized Hurst exponent $h(q)$ can be interpreted as follows [31]: $h \in (0,0.5)$ indicates antipersistency of the time series, $h = 0.5$ indicates uncorrelated noise, $h \in (0.5,1)$ indicates persistency of the time series, $h = 1.5$ indicates Brownian motion (integrated white noise), and $h \geqslant 2$ indicates black noise.

It is of great importance to choose the proper range of scale s, for which the family of curves $F(q,s)$ should be calculated. An inappropriate choice of the range of scale s may bring about a biased fitting exponent, referring to [32]. A too large scale s results in the division of the time series into a too small number of windows. It is found that a division into fewer than 50 data windows often causes the $F(q,s)$ curves to converge at the scale of saturation. Too small scale s causes the detrending procedure to be executed on a set of only a few points. For instance, for very small scale $s < 20$ (data points), after the detrending procedure, the calculated variance often is of the order of the accuracy of the MATLABdata type double, which may result in an arithmetic underflow. In our tests of the traffic speed time series, the often obtained drastic decrease of the values of $F(q,s)$ for $s < 20$ is a result of an arithmetic underflow. Therefore, we set the usable range of scales in this study to be $s \in [20, N/50]$, where N denotes the length of the time series. In the case of the traffic speed time series analyzed here, N was usually about 55 000. Thus, all of our calculations are made and presented for scales $s \in [20,1100]$, which is also [40 min, 2200 min], as the time interval is two minutes.

C. Multiscale MFDFA method

As a single scaling exponent is inadequate to describe the internal dynamics of signals, Echeverría *et al.* introduced an $\alpha\beta$ filter, based on the Kalman filter approach, to estimate the local exponents and applied it to human cardiac data [35]. Then, Govindan *et al.* quantified the long-range correlation of short fetal cardiac data sets by computing the mean value of the local exponents, associated with the phase randomized surrogate technique [36]. Castiglioni *et al.* introduced an estimation of the temporal spectrum of scale exponents $\alpha_{\mathrm{DFA}}(s)$ to study heart rate variability and blood pressure [37–39]. The authors applied the well-known monofractal DFA method but, instead of defining fixed scale ranges, they continuously varied the scale for which they calculated the DFA exponent $\alpha_{\mathrm{DFA}}(s)$, where $\alpha_{\mathrm{DFA}}(s)$ was defined as the derivative of $\log[F(s)]$ with respect to $\log[n]$. Inspired by their work, J. Gieraltowski *et al.* [32] went further and calculated a multifractal spectrum with variable scale ranges, called multiscale multifractal analysis (MMA).

After calculating all $F(q,s)$ by MF-DFA, we use a moving fitting window, sweeping through the whole range of scale s along the $F(q,s)$ plot. Figure 2 shows the computing procedure. Suppose one fitting window Ri ($i = 1,2,\ldots,n$), where $h_{\mathrm{R}i}$ corresponds to the local scaling exponent calculated in Ri. Then we can calculate the whole temporal spectrum of scale exponents, $h(s) = \{h_{\mathrm{R}1}, h_{\mathrm{R}2}, \ldots, h_{\mathrm{R}n}\}$, for a fixed q, which allows us to study quasicontinuous changes of the $h(q)$

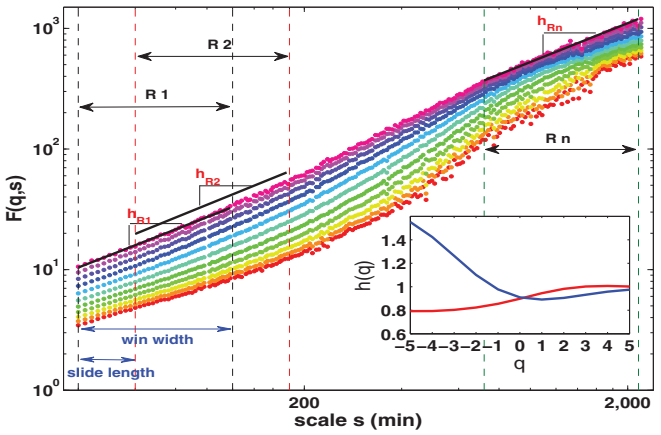


FIG. 2. (Color online) The log-log plots of fluctuation functions $F(q,s)$ vs scale s (color points) calculated for the traffic speed series presented in Fig. 1(b). The curves correspond to q between -5 and 5 in steps of 1 (from red lower points to rose-red upper points). Vertical dashed lines mark three examples of the fitting windows: small scale range $s \in$ [40 min, 120 min] (R1), $s \in$ [60 min, 180 min] (R2), and large scale range $s \in$ [720 min, 2160 min] (Rn). Take the $F(q = 5,s)$ curve (the top curve) for example, where the local scaling exponent $h_{\mathrm{R}1}$ for $q = 5$ corresponds to the slope of the fitting line within window R1. Then we can plot the fitting lines within R2, R3, and Rn and calculate their corresponding local scaling exponent $h_{\mathrm{R}2}, h_{\mathrm{R}3}, \ldots, h_{\mathrm{R}n}$. The subplot shows the $h(q)$ dependence calculated for the small and large scales s: e.g., red lower curve scales show $h(q)$ calculated in R1, while blue upper curve shows $h(q)$ calculated in Rn, indicating a significant difference at the large and small scales. Two parameters, "window width" and "slide length," are also shown. All scales s are presented in units of minutes for the traffic signals in this study.

dependence versus the range of scale s. Then repeat the above procedures for different q. We can define the generalized dependence $h(q,s)$ (Hurst surface, Fig. 3) as follows:

$$h(q,s) = \frac{\log[\Delta F(q,s)_{\mathrm{R}i}]}{\log(\Delta s_{\mathrm{R}i})}, \tag{4}$$

where $F(q,s)_{\mathrm{R}i}$ and $s_{\mathrm{R}i}$ represent points falling into window Ri. As fluctuation functions $F(q,s)$ are presented in log-log coordinates, the moving fitting window expands logarithmically, so that it seems to be of constant width.

The fit is made only for points currently inside the moving window. The starting window includes scale $s \in$ [40 min, 120 min], the second subwindow includes scale $s \in$ [60 min, 180 min], and then it is moved and expanded to reach the final subwindow $s \in$ [720 min, 2160 min] (see Fig. 2). In the Hurst surface (Fig. 3), the scale axis is calibrated so as to show the beginning of the fitting window—it starts from $s = 40$ min (i.e., the beginning of the first subwindow [40 min, 120 min]) and ends at $s = 720$ min (beginning of the final subwindow [720 min, 2160 min]). For clarity, the points of the $h(q,s)$ graph are connected to form a colored surface. Notably, the Hurst surface is presented using a linear axis, different from the plot of $F(q,s)$ versus s, which is plotted using a logarithmic axis for the scale s.

Similarly, for $q < 0$, the $h(q,s)$ plot corresponds to those fragments of the signal that have a low variance (small

fluctuations), while for $q > 0$, it characterizes the signal fragments with a large variance (large fluctuations). The results for $q > 0$ are much more stable and encumbered by smaller errors, while the results for $q < 0$ should always be used with caution.

D. Modifications of MMA method

In the MMA model, there are two important parameters: "window width" (WW) and "slide length" (SL), shown in Fig. 2, which may change the shape of $h(q,s)$ (Hurst surface). As $F(q,s)$ are presented in log-log coordinates, the moving fitting window should expand logarithmically, so that it seems to be of constant width. Window width is defined as the ratio between the right endpoint and left endpoint of window Ri, e.g., for window [40 min, 120 min], [720 min, 2160 min], WW is 3; for window [20 min, 40 min], [40 min, 80 min], WW is 2. WW determines the length of the moving window, inside which the local scaling exponents are calculated. Too large values of WW may result in ignoring some variations of the local exponents. Conversely, too small values of WW may lead to suspicious results due to noises, especially for those $F(q,s)$ with big fluctuations at negative q's (e.g., big fluctuations exist at $s \in$ [600 min, 1400 min] for $q = -5$ in Fig. 2). Figures 3, 4(a), and 4(b) illustrate the results of $h(q,s)$ using the same SL but WW = 3, 2, and 4, separately. For positive q's, the shapes of the Hurst surfaces with WW = 2,3,4 are similar. Hurst surfaces with higher resolutions for smaller WW may provide more accurate positions of crossovers. On the other side, for negative q's, the change of the shape of the Hurst surfaces is more pronounced. An obvious drop exists in $h(q,s)$ for windows beginning from 440 min to 680 min when WW = 2 [Fig. 4(a)], whereas Hurst surfaces, when WW = 3 (Fig. 3) and 4 [Fig. 4(b)], do not show such a drop. This is due to large fluctuations in the log-log plot of $F(q,s)$ versus s at such

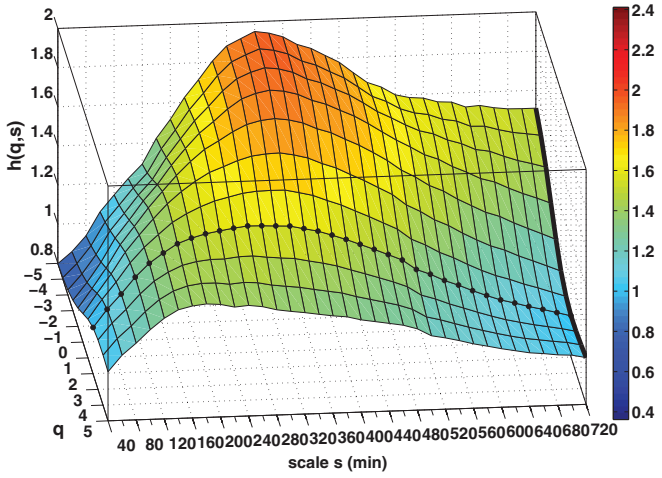


FIG. 3. (Color online) Hurst surface $h(q,s)$ calculated for the traffic speed time series collected by detector 3057 of length 55 440. Parameters here are WW = 3 and SL = 20 min. The black dots show Hurst curves $h(s)$ for $q = 2$. The black thick line at the right side of the plot corresponds to $h(q)$ calculated with the standard MF-DFA. Note that we use the same color bar standard for all $h(q,s)$ in this study.

windows, which may give unreliable results. On the contrary, fluctuation functions are more stable for positive q's, hence the $h(q,s)$ are more stable. We need to take the MMA results at negative q's when WW = 2 with caution.

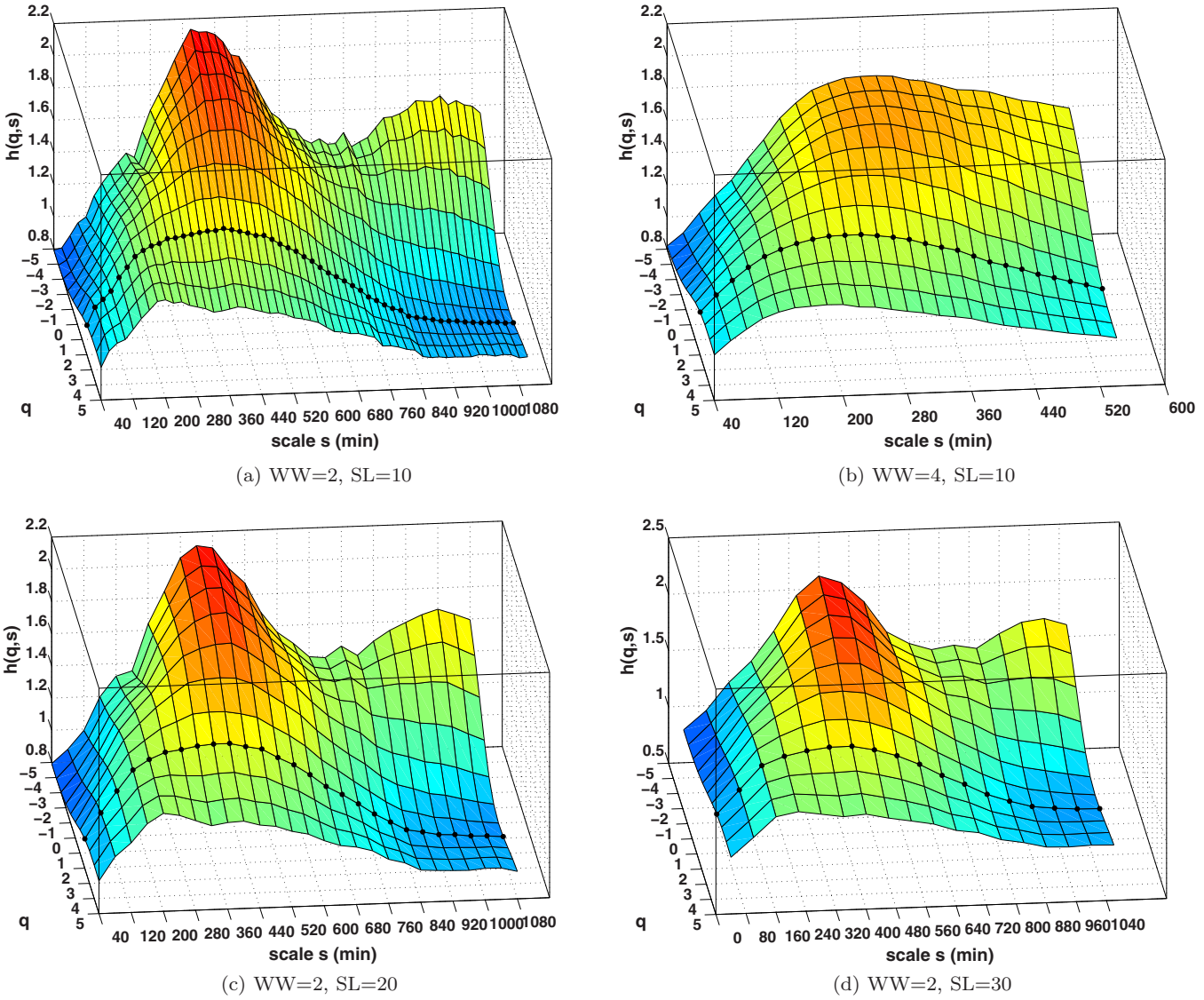
Another parameter is slide length (SL), which gives the shift length of windows. When too large values are given to SL, e.g., SL is larger than the length of the moving window, the local scaling exponents are calculated within entirely different windows, which may result in big variations in the values of $h(q,s)$, and then big fluctuations in the Hurst surface. When too small values are given to SL, e.g., SL = 1 data point, the windows are nearly overlapped. As a result, appropriate values (SL $\in$ [10, 0.5 window length]) for SL are necessary. Figures 4(a), 4(c), and 4(d) plot the Hurst surface using SL = 10, 20, and 30, from which the shapes are similar but with different resolutions. From the analysis above, we choose SL = 10 (20 min traffic series) and WW = 3 in the following parts; in particular, when we need to find the exact positions of crossovers, we may choose WW = 2.

III. RESULTS

A. Correspondence with earlier multifractal analysis methods

The MMA method is tested for two artificial series with known generalized Hurst exponent $h(q)$, which has been analyzed by earlier multifractal analysis methods [21,40–42]. First, it is tested on a monofractal series generated by the autoregressive fractionally integrated moving average (ARFIMA) process [43]: $z(t) = Z(d,t) + \epsilon(t)$, where $d \in (0,0.5)$ is a memory parameter and ϵ is an independent and identically distributed Gaussian variable, and $Z(d,t) = \sum_{n=1}^{\infty} a_n(d) z(t-n)$ in which $a_n(d) = \Gamma(n-d)/[\Gamma(1-d)\Gamma(n+1)]$. Then the Hurst exponent h_z is $h_z = 0.5 + d$ [40,44]. A monofractal series with $h(q) = 0.7$ (setting $d = 0.2$) of length 55 440 is generated, and we obtain satisfying results [Figs. 5(a) and 5(b)]. As expected, for every scale s, we obtain a flat Hurst surface [Fig. 5(b)], situated at 0.7.

A multifractal time series generated by multifractal binomial measures with known analytic multifractal properties [45] is obtained in an iterative way: start with the 0th iteration $k = 0$, where the data set $z(i)$ consists of one value, $z^{(0)}(1) = 1$. In the kth iteration, the data set $\{z^{(k)}(i) : i = 1,2,\ldots,2^k\}$ is obtained from $z^{(k)}(2i-1) = pz^{(k-1)}(i)$ and $z^{(k)}(2i) = (1-p)z^{(k-1)}(i)$ for $i = 1,2,\ldots,2^{k-1}$. When $k \to \infty$, $z^{(k)}(i)$ approaches to a binomial measure, whose scaling exponent $h(q)$ has an analytic form [45]: $h(q) = 1/q - \log_2[p^q + (1-p)^q]/q$. Here, we have performed $k = 16$ iterations with $p = 0.25$, and also gained satisfying results [Figs. 5(c) and 5(d)]. The shapes of $h(q,s)$ for all available scales are very similar to the standard shape of the multifractal $h(q)$ dependence [embedded plot in Fig. 5(c)], but with slight fluctuations at negative q's. Some fluctuations are obvious for negative q's in the log-log plot of $F(q,s)$ versus s [Fig. 5(c)], leading to the variations of the Hurst surface at negative q's, as the Hurst surface is very sensitive to fluctuations in the fluctuation functions. But we can still get a multifractal characterization and an accurate estimate of the scaling exponent for the binomial multifractal series



(a) WW=2, SL=10

(b) WW=4, SL=10

(c) WW=2, SL=20

(d) WW=2, SL=30

FIG. 4. (Color online) Hurst surface $h(q,s)$ dependence calculated for the traffic speed time series for different values of parameters: window width (WW) and slide length (SL). The black points corresponds to $h(q = 2,s)$.

from the Hurst surface. The above results indicate that the MMA method can successfully reproduce the results obtained from earlier multifractal analysis methods.

B. Effect of periodical trends

We first focus on the MMA results for original traffic signals X. The Hurst surface (Fig. 3) calculated for signal X shows abundant information, which may be hidden by traditional MF-DFA. The Hurst surface shapes like a hill, experiencing (a) a rising stage, i.e., a transition from the color blue to red for negative q's and a transition from the color blue to green for positive q's; (b) a stable stage, i.e., red areas for negative q's and yellow or yellow-green area for positive q's, where the values of $h(q,s)$ are stable only with small variations; and (c) a decreasing stage, i.e., a transition from red to green for negative q's, and a transition from yellow-green to blue for positive q's. The Hurst surface also helps us to locate the crossovers. We mainly focus on $h(q > 0,s)$ with parameters WW $= 2$ and SL $= 10$ [Fig. 4(a)] as it provides results with higher resolutions. Through observing the log-log plot of $F(q,s)$ versus s (Fig. 2) and its corresponding Hurst surface (Fig. 3), we find that when the color of the Hurst surface consecutively changes, it corresponds to windows containing crossovers; when the color stabilizes at a certain color, it corresponds to windows which just step over the crossover. Then we can locate the crossovers when the color of the Hurst surface starts to stay at a certain color: i.e., we can find that the color of $h(q > 0,s)$ is stable in two areas, i.e., yellow green (approximately at $s = 180$ min) and light blue (roughly at scale $s = 360$ min), corresponding to 3-hour and 12-hour periods for traffic speed signals [Fig. 6(a)].

Then, we turn to focus on curves of $h(q,s)$ from the perspectives of q and s, respectively. For a fixing scale s (i.e., in the same window Ri), when q changes from -5 to 5, there are downtrends for all $h(q,s_{\mathrm{R}i})$ curves, showing the

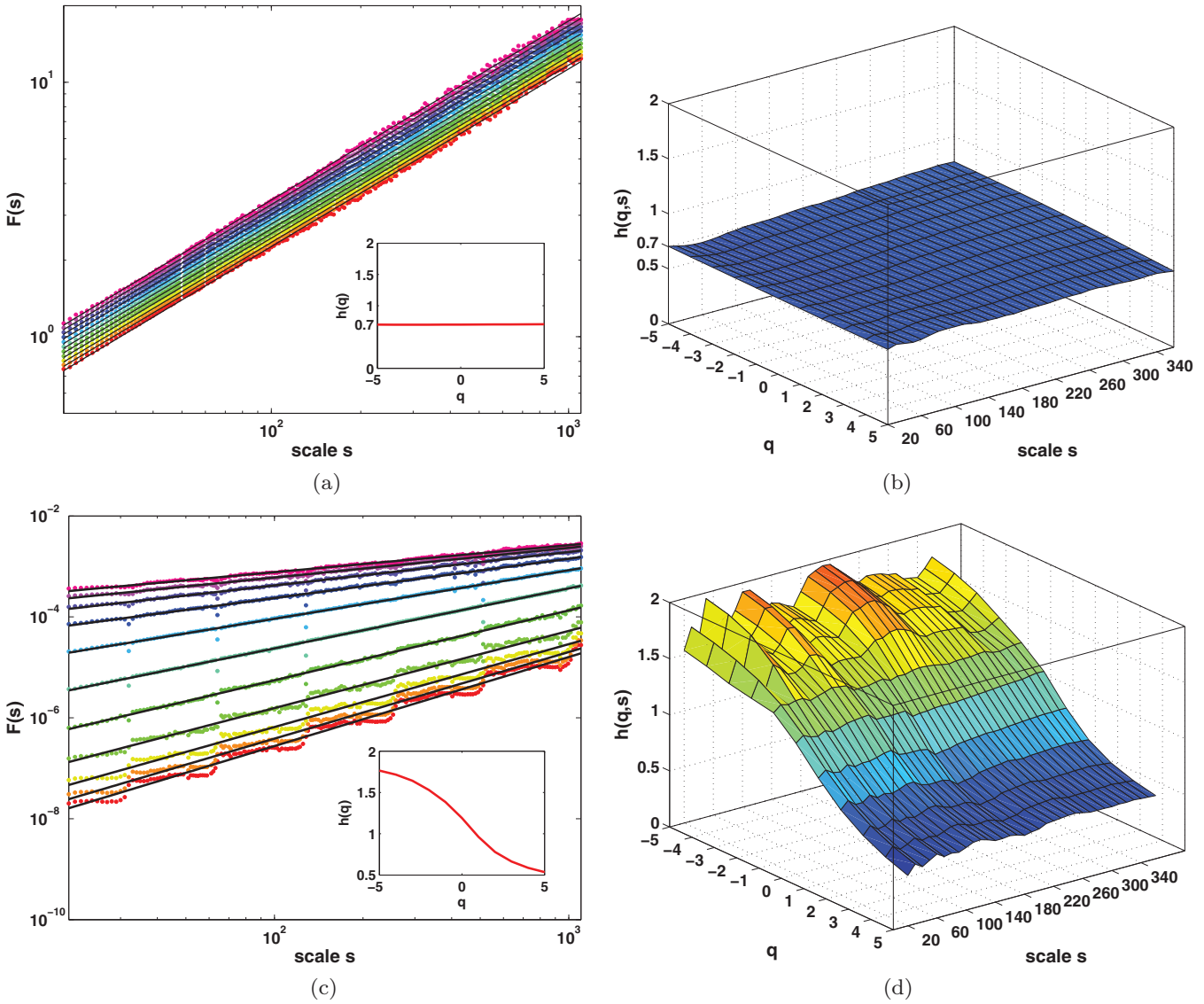


FIG. 5. (Color online) (a) The fluctuation function $F(s)$ vs scale s calculated using MF-DFA for a monofractal series generated by the ARFIMA model [$d = 0.2$, $h(q) = 0.7$, length of series 55 440]. Inset: corresponding $h(q)$ dependence. (b) The $h(q,s)$ dependence calculated for the monofractal series in (a). (c) $F(s)$ vs s for a multifractal series generated by multifractal binomial measures (parameter $p = 0.25$, length of series 65 536). Inset: corresponding $h(q)$ dependence. (d) The $h(q,s)$ dependence calculated for the multifractal series in (c).

multifractal properties for the traffic signal, which is consistent with former results [14,15]. For a fixed q (e.g., $q = 2$, black points in Fig. 3), when the window moves from small scales to large scales, all curves of $h(q = 2,s)$ first show a rapid rise to their peak values, then become stable for a while, and finally decrease to certain values. The above information could be covered by traditional MF-DFA as it only calculates one scaling exponent for each q at the whole scale range, but, actually, abundant structure exists, with different local exponents shown in the corresponding Hurst surface.

The traffic system is enforced by well-defined periodicity as well as the existence of unpredictable perturbation. Such periodic trends may have some undesired effects, such as the crossovers in the log-log plot of $F(q,s)$ versus s, rendering the variations on the Hurst surface. From the time domain to frequency domain using Fourier transform [Fig. 6(a)], the periods of original traffic speed signal X are consistent with the real case, such as 7 days, 24, 12, 8, 6, 3 hours, which is similar to [15]. There is a simple way for us to remove the periodic trends of signal X using Fourier transform [15,46]: first, transform the signal X to Fourier space, and locate the corresponding frequency $f_k(X)$ of the dominant periodic trends embedded in the signals; set $|F(f^*(X))|$ to 0 at frequency $f_k(X)$ for convenience [47] and inverse transform $F(f^*(X))$; then, signal Y [Fig. 6(b)] is finally derived without periodic trends.

The Hurst surface $h(q,s)$ [Fig. 6(d)] calculated for signal Y ranges from 0.7177 to 1.4794, which is much smaller compared to that of signal X ([0.7981, 1.9499]), but still indicates a long-range correlation and multifractal property. The Hurst surface still has a hill-like shape, but the peak (highest values) shifts to smaller scales, and blue areas dominate the surface. Comparing the log-log plot of $F(q,s)$ versus s for signal X (Fig. 2) and signal Y [Fig. 6(c)], the

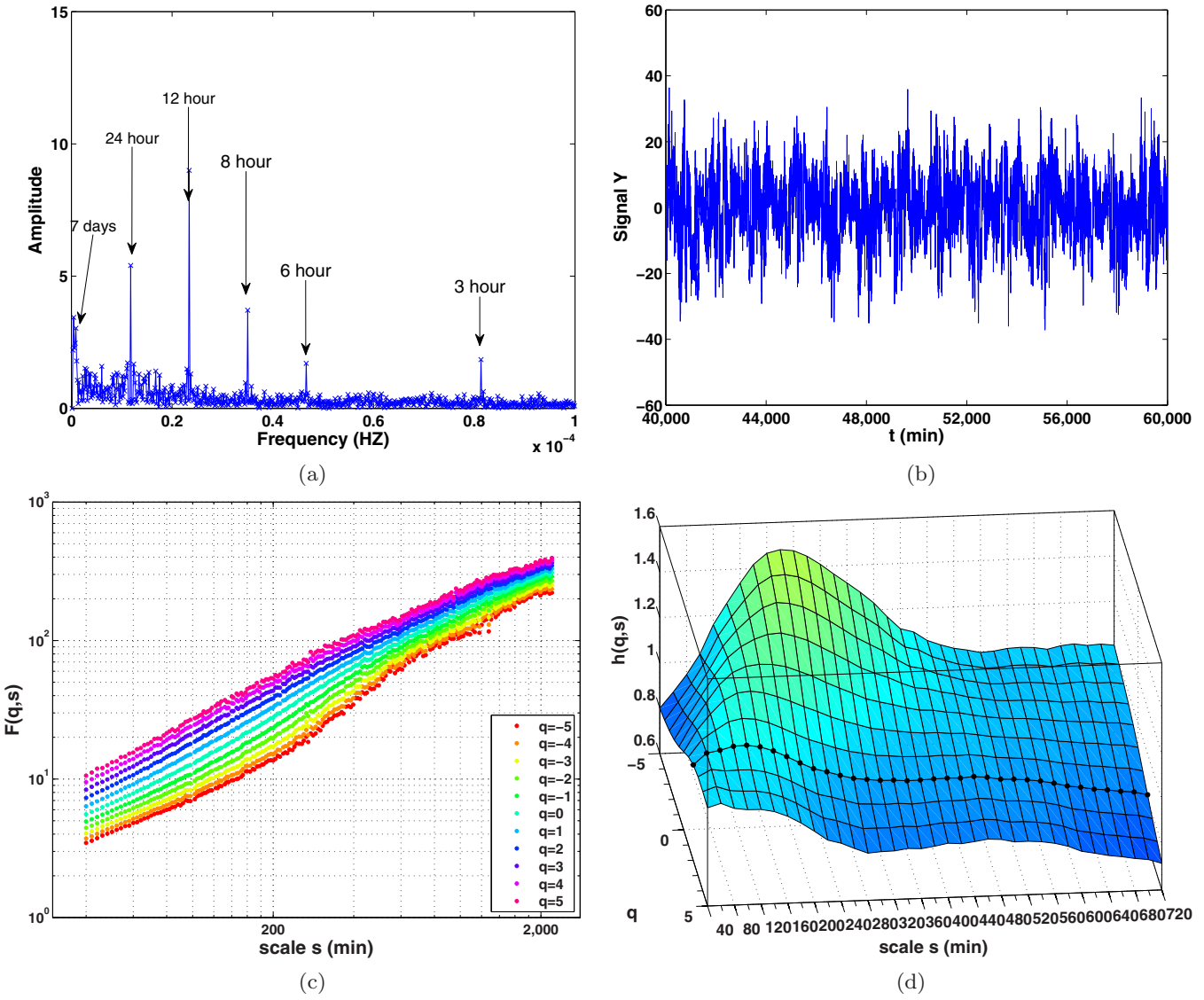


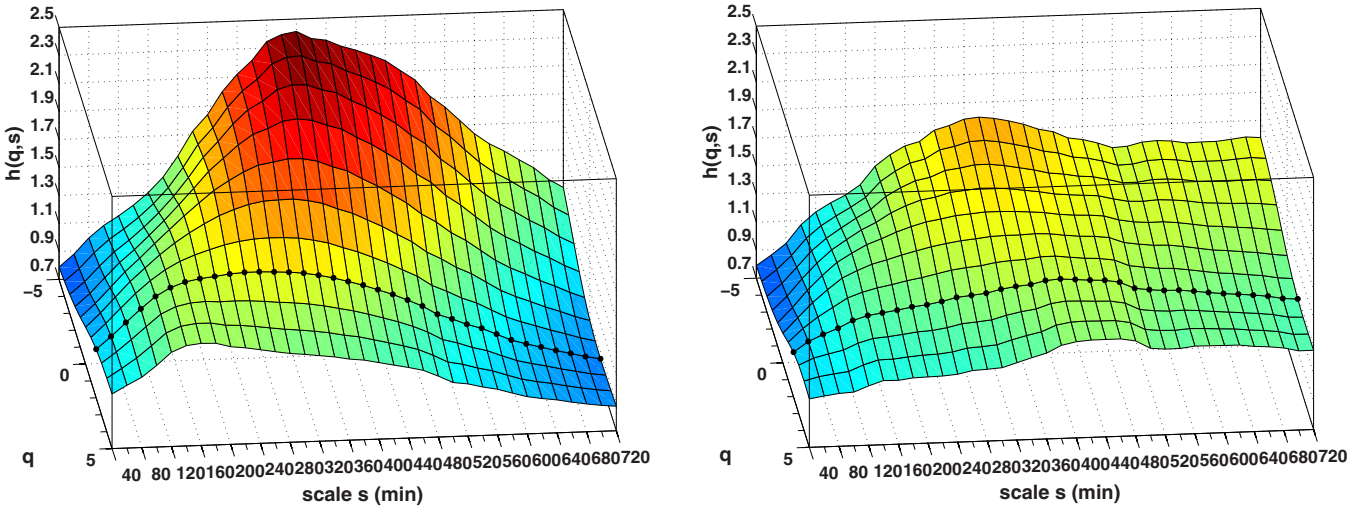
FIG. 6. (Color online) (a) Dominant frequencies and amplitudes for the traffic speed series after Fourier transform. Dominant periods of 7 days, 24, 12, 8, 6, and 3 hours can be observed. After removing the dominant periods via Fourier transform, we obtained signal Y shown in (b). (c) Fluctuation function $F(q,s)$ vs scale s calculated for signal Y. (d) The Hurst surface $h(q,s)$ calculated for signal Y.

bending degree drops for all curves and some crossovers disappear after removing periodic trends, indicating that some crossovers are caused by the periodicities in the signal. But the bump at negative q's of the Hurst surface or the remaining bending curves in Fig. 6(c) indicate that there are other interesting factors or underlying traffic mechanisms which dominate traffic signals.

C. Effect of weekday and weekend patterns

We investigate two important patterns in traffic signals: weekday (or working day) and weekend (or holiday) patterns. The traffic speed series at weekday exhibit a two-sudden-drop pattern at rush hours, indicating a periodic pattern with some slight variations from Monday to Friday [lower subplot in Fig. 1(b)]. Different from that of the weekday speed series, the data at the weekend do not display such periodic pattern, which may reflect on the Hurst surface. We then divide the whole original traffic series X into two parts: weekday (X_{weekday}) and weekend (X_{weekend}), and analyze their periodicities separately. The dominant periodicities of signal X_{weekday} are similar to signal X, showing a dominant period of 7 days, 24, 12, 8, 6 and 3 hours; while that of signal X_{weekend} only shows period of 12 and 24 hours. We also remove the periodic trends of the two subseries using Fourier transform (shown in Sec. III B) and obtain signal Y_{weekday} and Y_{weekend} to investigate the effect of weekday and weekend patterns on the Hurst surface $h(q,s)$.

It is observed that the Hurst surface for signal X_{weekday} is similar to that for signal X, but with the shape of a steeper hill: higher values at higher area and lower values at lower area, ranging in [0.7885, 2.4114] [Fig. 7(a)]. The Hurst surface for signal X_{weekend}, on the contrary, appears to be a smoother shape compared to that for signal X, ranging in [0.7897, 1.7945] [Fig. 7(b)]. We know that the weekday signal before removing periodic trends (X_{weekday}) has a more typical periodicity than the whole signal X. And signal X can be considered as



(a) Signal $X_{weekday}$ (b) Signal $X_{weekend}$

(c) Signal $Y_{weekday}$ (d) Signal $Y_{weekend}$

FIG. 7. (Color online) The Hurst surface $h(q,s)$ calculated for weekday (left side) and weekend (right side) traffic speed signals. Upper plots shows results before removing periodic trends (represented in $X_{\text{weekday/weekend}}$), while lower plots shows $h(q,s)$ for signals after removing periodic trends ($Y_{\text{weekday/weekend}}$). Note that the weekend series also include holidays such as the Chinese National Day holiday from 01/10 to 05/10. The length of the weekday and weekend data is 35 318/19 176. We still use the same range of scale s in order to compare with signal X.

a combination of both weekday and weekend parts. Thus the Hurst surface for whole signal X (Fig. 3) represents the features of both weekday and weekend signals. After removing dominant periodicities, the Hurst surfaces [shown in Figs. 7(c) and 7(d)] of both weekday and weekend signals behave differently from results for corresponding signals before removing trends as follows. (a) The Hurst surfaces for both signals Y_{weekday} and Y_{weekend} become similar, unlike the significant difference between the Hurst surfaces for signals X_{weekday} and X_{weekend}, which means the main distinction between weekday and weekend patterns is their periodic patterns. (b) The Hurst surfaces for both parts after taking out their trends become flat, with much smaller values compared to those before removing trends: $h(q,s)_{Y_{\text{weekday}}} \in [0.4724, 1.2461]$ and $h(q,s)_{Y_{\text{weekend}}} \in [0.3621, 1.2929]$. (c) For signal Y_{weekday}, the Hurst surface slowly decreases when $q > 0$; it first rises and then gradually decreases when $q < 0$. For signal Y_{weekend}, the Hurst surface first shows an increase and then a gradual decline for all q's. (d) Multifractality still exists whether or not the trends are removed for both parts. From the above analysis, we know that the Hurst surface is very sensitive to the weekday and weekend patterns of traffic signals, namely, to different periodic trends. Such rich information contained in the Hurst surface may be hidden by the traditional MF-DFA method.

D. The generating mechanism of multifractality of traffic signal

There are two different types of multifractality in the time series: (i) multifractality due to a broad probability density function (PDF) for the values of the time series, which cannot be removed by shuffling the series; and (ii) multifractality due to different long-range correlations of the small and large fluctuations, which can be destroyed by the shuffling procedure. If both types of multifractality exist, the shuffled

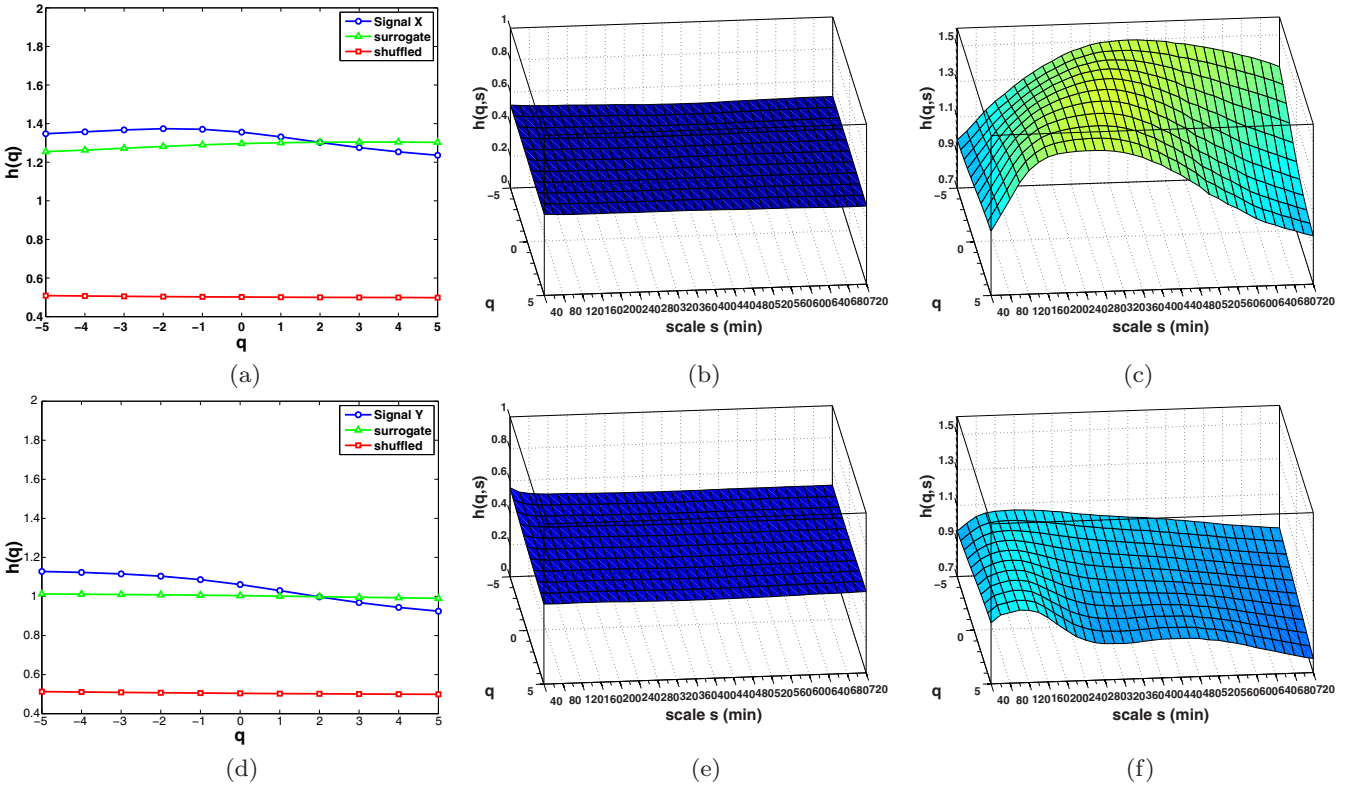


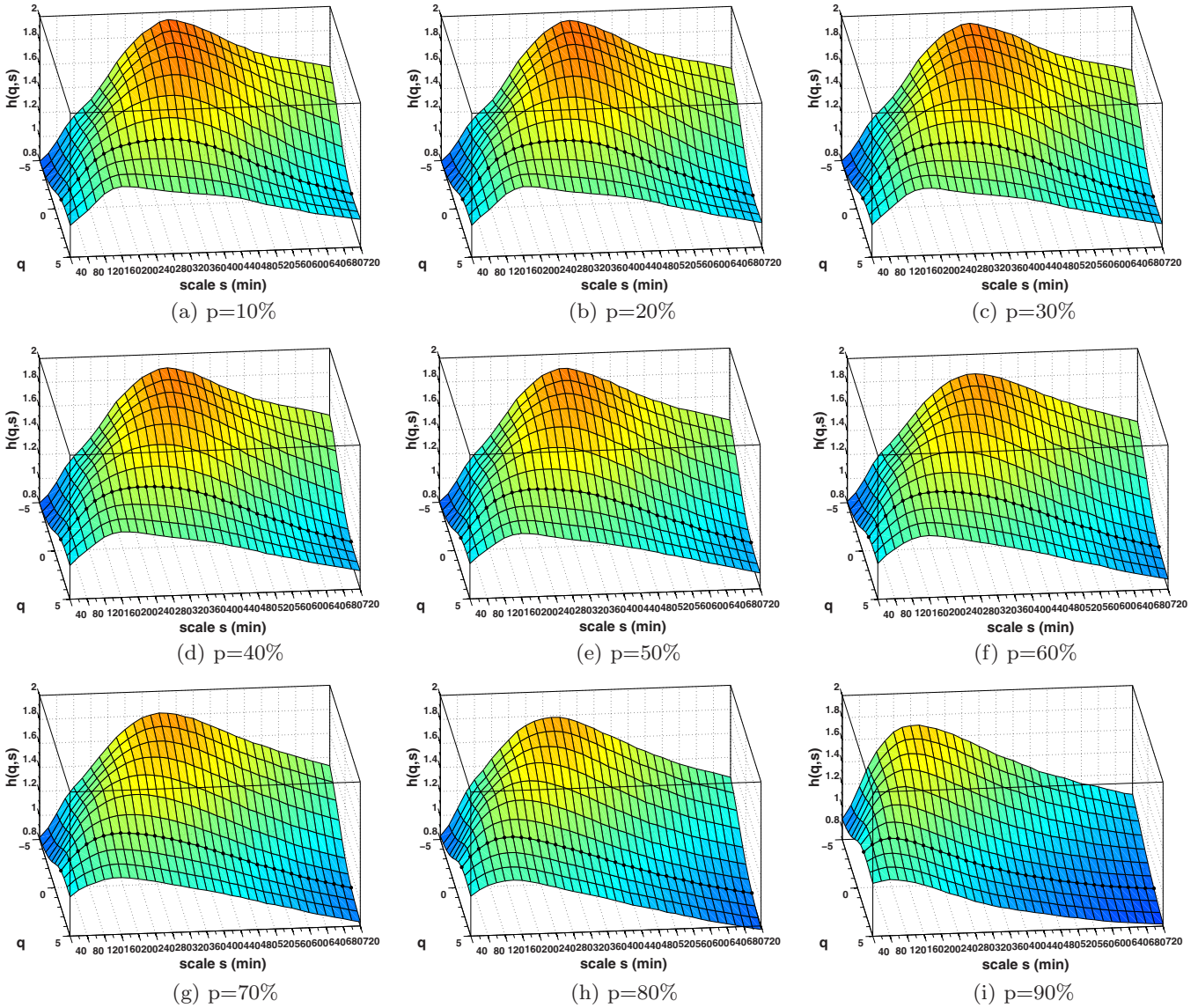
FIG. 8. (Color online) Generalized Hurst exponent $h(q)$ as a function of q for original, surrogate, and shuffled time series calculated by MF-DFA, (a) for data before removing trends and (d) for data after removing trends. Hurst surface $h(q,s)$ calculated for (b),(e) shuffled series and (c),(f) surrogate series of the traffic speed signals. Presented are results averaged over 500 realizations of the shuffled and surrogate procedures. Upper figures are results for data before removing trends, while lower figures are results for data after removing trends.

series will show weaker multifractality than the original series. The easiest way to distinguish between these two types of multifractality is by analyzing the corresponding randomly shuffled [21] and surrogate time series [48,49]. The shuffling of time series destroys the long-range correlation. Hence, if the multifractality only belongs to the long-range correlation, we will find $h_{\text{shuf}} = 0.5$. The multifractality nature due to the fatness of the PDF will not be affected by the shuffling procedure. On the other hand, to determine the multifractality due to the broadness of the PDF, the phase of discrete fourier transform (DFT) coefficients of the traffic series are replaced with a set of pseudoindependent distributed uniform $(-\pi,\pi)$ quantities in the surrogate method [50,51]. The correlations in the surrogate series do not change, but the probability function changes to a Gaussian distribution. If multifractality in the times series is only due to a broad PDF, then $h(q)$ obtained by the surrogate method will be independent of q. If both kinds of multifractality are present in the traffic time series, then the shuffled and surrogate series will show weaker multifractality than the original one. As the MMA method is a generalized version of the MF-DFA method, the Hurst surface will exhibit the same results after the shuffling procedure with traditional MF-DFA.

To show the source of multifractality of the traffic speed signals analyzed in this study, we first compare the generalized Hurst exponent $h(q)$ for the original series with the result of the corresponding shuffled and surrogate series before and after cancellation of trends, shown in Figs. 8(a) and 8(d). The dependence of $h_{\text{shuf}}(q)$ and $h_{\text{sur}}(q)$ shows that the multifractality nature of the traffic speed time series is due to both long-range correlation and broad PDF. But the multifractality due to fatness is much weaker than that due to correlation. We then calculate $h(q,s)$ for the corresponding shuffled and surrogate series of the traffic speed signals before and after removing periodic trends. Figure 8 also presents the average results of 500 realizations of the shuffled and surrogate procedures of the original traffic speed signals. The shuffled procedure destroys the correlations—the results closely resemble those for white noise ($\langle h(q,s)\rangle \approx 0.5$) for both series before and after removing trends. The generating procedure of the surrogate series also changes the Hurst surface. Through comparing the mean value of $h(q,s)$ (Table I), we also find that the multifractality of the traffic speed signals in this study is due to both broad PDF and correlation, but the source of multifractality is mainly correlation, which is consistent with the results of MF-DFA. We also have tested the source of multifractality of the traffic speed signals collected from the other three detectors, which also shows similar results.

TABLE I. The values of $h(q,s)$ of the original series, and average values for the corresponding shuffled and surrogate series before and after removing periodic trends.

	$\langle h(q,s)\rangle$	$\langle h_{\text{shuf}}(q,s)\rangle$	$\langle h_{\text{sur}}(q,s)\rangle$
Before	1.4094	0.4993	1.3335
After	1.0096	0.5010	0.9603



(a) p=10% (b) p=20% (c) p=30%

(d) p=40% (e) p=50% (f) p=60%

(g) p=70% (h) p=80% (i) p=90%

FIG. 9. (Color online) The Hurst surface $h(q,s)$ for the traffic signal X with different level of data loss: (a)–(i) 10%–90% with increments of 10%. Presented are results averaged over 500 realizations of the test series.

E. Effect of data loss

For real traffic signals, data can be missing or unavailable to a very large extent which, once recorded in the past, often cannot be generated again. Knowledge of what the effects of data loss may have on the correlations and other dynamical properties of the output signals for a given system is instrumental in accurately quantifying and modeling the underlying mechanisms driving the dynamics of the system [52]. Data loss of traffic signals can be caused by failure of the data collection equipment, as well as by the removal of artifacts or noise-contaminated data segments. Thus, it is necessary to find out the effect of data loss for traffic signals on the MMA results. Initially, we calculate $h(q,s)$ for the whole time series (Fig. 3). We then analyze the variations of Hurst

TABLE II. Mean values $\langle h_2(q,s)\rangle$ and intersurface distance d between the results for X and $\tilde{X}$ at different p's. In each case, the mean value μ and standard deviation δ of the results for the 500 realizations of each test series are given.

	p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d	μ	0.0196	0.0276	0.0347	0.0442	0.0572	0.0689	0.0870	0.1237	0.1934
	δ	0.0071	0.0097	0.0125	0.0119	0.0159	0.0151	0.0175	0.0209	0.0320
$\langle h_2(q,s)\rangle$	μ	1.4073	1.4000	1.3903	1.3720	1.3577	1.3335	1.2975	1.2430	1.1262
	δ	0.2632	0.2618	0.2599	0.2585	0.2574	0.2576	0.2570	0.2645	0.2742

surfaces for different percentages p (from 10% to 90% with increment of 10%) of data loss. We prepare groups of test series by introducing a segmentation approach to generate surrogate series $\tilde{X}_k$ by randomly removing data segments from signal X and stitching together the remaining parts of X [52]. For the removed segment, we use uniform distribution, range [1, 1000], which is a common distribution of missing segments of traffic recordings. As the length of two-minute traffic recordings for a day is about 720, the data loss is usually less than 720; here, 1000 represents the extreme case for data loss. The results plotted in Fig. 9 and all of the distances d presented below are averaged over 500 realizations of the segmentation procedure and presented together with the standard deviation of the results.

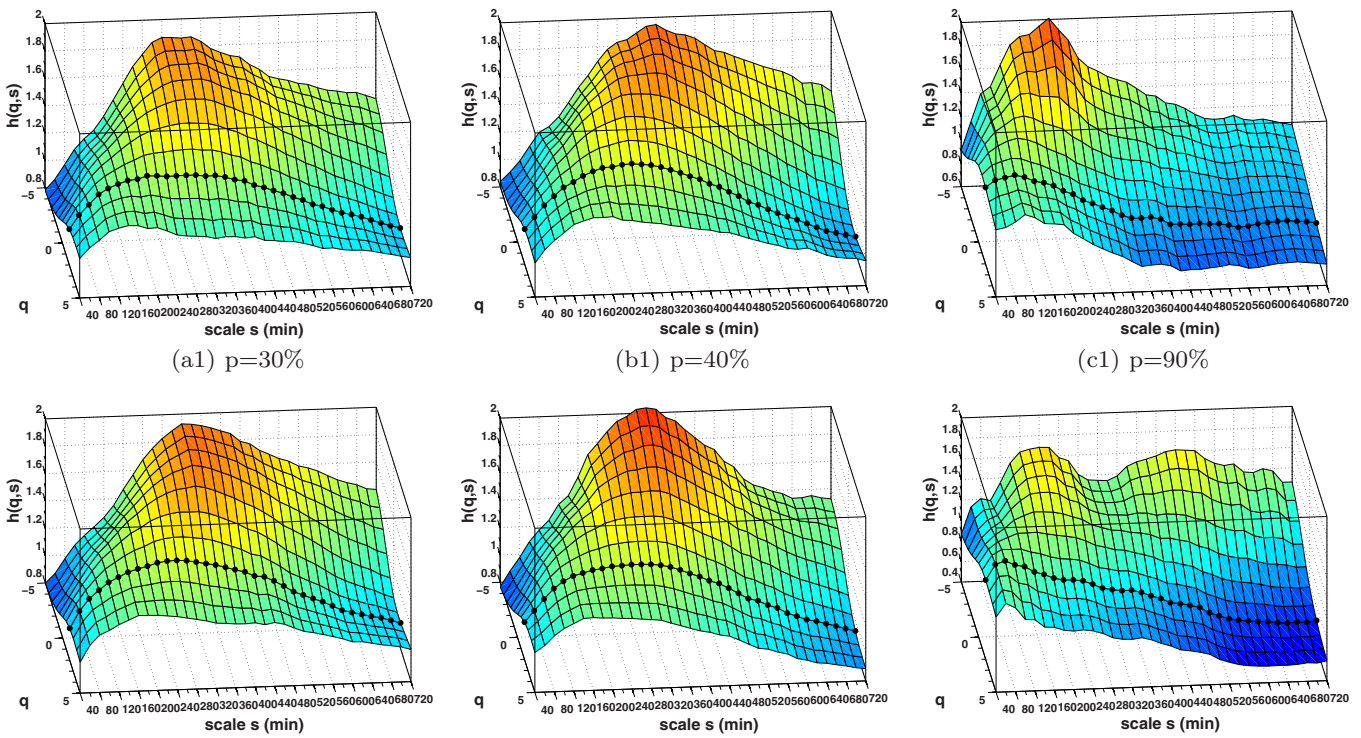
The Hurst surface for data loss of $p = 0.1$ [Fig. 9(a)] is very similar to that for the whole available series X (Fig. 3). As all results plotted in Fig. 9 are averaged over 500 realizations of the test series, the Hurst surfaces at different p's are smoother. There are some variations between the Hurst surfaces of each realization. The Hurst surfaces at $p = 0.2$ and 0.3 are still very similar, while the results for data loss of $p = 0.4$ begin to behave differently, but preserve the shape. As p rises, it is visible that the Hurst surfaces become flat, with values of $h(q,s)$ becoming smaller at all scale s and q (from the color of the Hurst surface). Moreover, the red areas [higher $h(q,s)$'s] start to shift to smaller s ranges, with a visible "left shift," while the blue areas at large scales expand gradually. Especially for $p = 90\%$, the changes are even more pronounced, losing the similarity in shape.

To reliably describe the differences between two Hurst surfaces, intersurface distance d, a generalized mean distance between two $h(q,s)$ surfaces is calculated [32]: (1) Calculate the mean $\langle h(q,s)\rangle$ for each surface. (2) Adjust the values of one of the Hurst surfaces so that the means of both surfaces are the same by shifting the surface up or down, shown in equation $h_2'(q,s) = h_2(q,s) + [\langle h_1(q,s)\rangle - \langle h_2(q,s)\rangle]$, where, for example, $h_1(q,s)$ is the reference surface for all data X; $h_2(q,s)$ is the result for test series $\tilde{X}_k$; $h_2'(q,s)$ is the shifted surface of equal mean value with $h_1(q,s)$; and $\langle\cdot\rangle$ denotes the mean (in this way, we are able to focus only on the differences in the shape of the surfaces). (3) Then calculate the intersurface distance d using the equation as follows:

$$d = \{\langle [h_1(q,s) - h_2'(q,s)]^2\rangle\}^{\frac{1}{2}} / [\langle h_1(q,s)\rangle]^{-1}. \qquad (5)$$

Here, we set $d = 0.05$ as the threshold so that if $d \leqslant 0.05$, we consider two surfaces to have similar shape, and if $d \geqslant 0.05$, we consider them to have different shape.

Table II shows the generalized distance d between Hurst surfaces for signal X (full length series) and signal $\tilde{X}$ (series with data loss at p's). The mean distances between the results (taking into account the standard deviation) for X and $\tilde{X}$ at level $p = 0.1$ to 0.3 are less than 0.05, which is thus below the preassumed threshold. The mean distance for $\tilde{X}$ at $p = 0.4$ is also less than 0.05, but the standard deviation δ for this case shows that some of the test series yielded $h(q,s)$, exceeding the threshold criterion of 0.05. Both μ and δ of intersurface distances d increase with p, which means deviations become larger with p; and increase of δ illustrates that the deviations of



(a1) p=30% (b1) p=40% (c1) p=90%

(a2) p=30% (b2) p=40% (c2) p=90%

FIG. 10. (Color online) Hurst surface for the traffic speed series when randomly removed at percentage (a1),(a2) $p = 30\%$, (b1),(b2) $p = 40\%$, and (c1),(c2) $p = 90\%$. Presented are results of the MMA of a randomly selected test series at different p.

each realization at the same data loss level become distant as p rises. We also plot the Hurst surface of some randomly selected test series at the same data loss level p in Fig. 10. Through comparing with the results averaging 500 realizations, the Hurst surfaces of the selected test series at $p = 30\%$ are almost consistent with its average results. It is observable that the results of the selected series at $p = 40\%$ sometimes differ from its average results. More significant deviations from its average results are observed at the Hurst surface of the selected series at $p = 90\%$, and the shapes of the Hurst surface of each case vary greatly. The above analysis reminds us to take the MMA results with data loss $p \geqslant 40\%$ with caution.

IV. CONCLUSION

In this study, we applied multiscale multifractal analysis (MMA) to investigate the fractal properties of traffic signals using a spectrum of local Hurst exponents: $h(q,s)$ (Hurst surface). On this surface, the results of standard DFA are represented as one or two single points, and the results of standard MF-DFA are represented as a single line. The MMA method allows us to perform multifractal analysis without initial assumptions about the time scale of the problem investigated. And it is able to characterize the multifractality of the time series in a wide range of frequencies (scales) simultaneously. The Hurst surface is also a far better way of analyzing time series with crossovers, and can easily position the crossovers; therefore, we do not have to avoid data sets with crossovers or narrow the range of scales to only large or small scales.

We focus on the results of the MMA for traffic speed signals and obtain some interesting findings. First, traffic speed signals show multifractal properties and strong long-range correlations, which is consistent with previous studies. Second, the Hurst surface presents a rising phase at small scales, a stable phase, and then a decreasing phase at large scales, which shows crossovers exist in Hurst curves at different q's. To position the crossovers, we adjust values of window width and slide length, and find crossovers are approximately at $s = 180$ min and 720 min, corresponding to 3-hour and 12-hour periods. And this cannot be easily obtained from Hurst curves of MF-DFA. We also study the effect of periodical trends on the Hurst surface. By comparing Hurst surfaces before and after removing dominant periodicities, we find that crossovers are mainly coming from periodicities. But traffic signals after removing such trends still have multifractality and long-range correlations. Third, regarding two special traffic patterns, i.e., the weekday and weekend pattern, we divide the whole data into a weekday series and a weekend series, and obtain that the Hurst surfaces behave differently. In addition, it is shown that multifractality of the traffic speed signals is stemming from both broad PDF and correlations (the main source), via comparing the original series with their shuffled and surrogate series. We also discuss the effect of data loss on MMA results for traffic series. Through a segmentation approach, we generate test series of different percent p of data loss by randomly removing data segments from its original series. It shows that MMA results with data loss of percentages from 0.1 to 0.3 are consistent with the full-length data, while some deviations occur when $p \geqslant 0.4$, which suggests taking MMA results for time series with data loss $p \geqslant 0.4$ with caution.

However, there are still some areas that need further study. The MMA method can be applied to different time series of various fields, such as economics, finance, climate systems, etc. Similar to MF-DFA, we can also generalize MF-DCCA (multifractal detrended cross-correlation analysis [41]) to multiscale MF-DCCA to investigate the cross correlations between two time series. Besides, in this study, we simply analyzed the effect on the MMA results of data loss at different percentages. However, it is of great importance to discuss the average length of the removed data segments as well as the functional form of the distribution of length of the removed data segments, e.g., Gaussian or exponential distribution. In addition, we only discuss the effect of data loss for the traffic speed time series, which is strongly positively correlated; it is also necessary to apply it to an anticorrelated series, and to the time series of some certain distributions.

ACKNOWLEDGMENTS

Financial support by China National Science (Grants No. 61071142 and No. 61371130), Beijing National Science (Grant No. 4122059), the National High Technology Research Development Program of China (863 Program) (Grant No. 2011AA110306), and the Fundamental Research Funds for the Central Universities (Grant No. S14JB00050) is gratefully acknowledged.

[1] D. Chowdhury, L. Santen, and A. Schadschneider, Phys. Rep. **329**, 199 (2000).
[2] D. Helbing, Rev. Mod. Phys. **73**, 1067 (2001).
[3] B. S. Kerner, *The Physics of Traffic* (Springer, Berlin-Heidelberg, 2004).
[4] P. Shang, Y. Lu, and S. Kamae, Chaos, Solitons Fractals **36**, 82 (2008).
[5] L. A. Safonov, E. Tomer, V. V. Strygin, Y. Ashkenazy, and S. Havlin, Europhys. Lett. **57**, 151 (2002).
[6] K. Daoudi and J. Lévy Véhel, Signal Process. **82**, 2015 (2002).
[7] I. Gasser, G. Sirito, and B. Werner, Physica D **197**, 222 (2004).
[8] R. E. Wilson, Philos. Trans. R. Soc. A **366**, 2017 (2008).
[9] M.-Y. Bai and H.-B. Zhu, Physica A **389**, 1883 (2010).
[10] J. Wang, P. Shang, X. Zhao, and J. Xia, Int. J. Mod. Phys. C **24**, 1350006 (2013).
[11] M. Y. Choi and H. Y. Lee, Phys. Rev. E **52**, 5979 (1995).
[12] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1996).
[13] P. Shang and X. Li, Chaos, Solitons Fractals **25**, 121 (2005).
[14] P. Shang, X. Li, and S. Kamae, Phys. Lett. A **357**, 314 (2006).
[15] X. Zhao, P. Shang, A. Lin, and G. Chen, Physica A **390**, 3670 (2011).
[16] J. Feder, *Fractals* (Plenum, New York, 1988).
[17] A.-L. Barabási and T. Vicsek, Phys. Rev. A **44**, 2730 (1991).
[18] H.-O. Peitgen, H. Jurgens, and D. Saupe, *Chaos and Fractals* (Springer, New York, 1992), Appendix B.

[19] E. Bacry, J. Delour, and J. F. Muzy, Phys. Rev. E **64**, 026103 (2001).
[20] J. F. Muzy, E. Bacry, and A. Arneodo, Phys. Rev. E **47**, 875 (1993).
[21] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, Physica A **316**, 87 (2002).
[22] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).
[23] L. Telesca, G. Colangelo, V. Lapenna, and M. Macchiato, Phys. Lett. A **332**, 398 (2004).
[24] N. K. Vitanov and E. D. Yankulova, Chaos, Solitons Fractals **28**, 768 (2006).
[25] G. R. Jafari, P. Pedram, and L. Hedayatifar, J. Stat. Mech. (2007) P04012.
[26] G. Lin and Z. Fu, Physica A **387**, 573 (2008).
[27] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, Nature (London) **399**, 461 (1999).
[28] P. Oswiecimka, J. Kwapien, and S. Drozdz, Physica A **347**, 626 (2005).
[29] Z.-Q. Jiang, W. Chen, and W.-X. Zhou, Physica A **388**, 433 (2009).
[30] J. Kwapien and S. Drozdz, Phys. Rep. **515**, 115 (2012).
[31] J. W. Kantelhardt, *Mathematics of Complexity and Dynamical Systems* (Springer, New York, 2011), pp. 463–487.
[32] J. Gieraltowski, J. J. Zebrowski, and R. Baranowski, Phys. Rev. E **85**, 021915 (2012).
[33] C. O. Tan, M. A. Cohen, D. L. Eckberg, and J. A. Taylor, J. Physiol. **587**, 3929 (2009).
[34] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, Chaos **5**, 82 (1995).
[35] J. C. Echeverría, M. S. Woolfson, J. A. Crowe, B. R. Hayes-Gill, G. D. H. Croaker, and H. Vyas, Chaos **13**, 467 (2004).
[36] R. B. Govindan, J. D. Wilson, H. Preißl, H. Eswaran, J. Q. Campbell, and C. L. Lowery, Physica D **226**, 23 (2007).
[37] P. Castiglioni, G. Parati, A. Civijian, L. Quintin, and M. Di Rienzo, IEEE Trans. Biomed. Eng. **56**, 675 (2009).
[38] P. Castiglioni, G. Parati, M. Di Rienzo, R. Carabalona, A. Cividjan, and L. Quintin, J. Physiol. **589**, 355 (2011).
[39] P. Castiglioni, G. Parati, C. Lombardi, L. Quintin, and M. Di Rienzo, Biomed. Eng. **56**, 175 (2011).
[40] B. Podobnik and H. E. Stanley, Phys. Rev. Lett. **100**, 084102 (2008).
[41] W.-X. Zhou, Phys. Rev. E **77**, 066211 (2008).
[42] Z.-Q. Jiang and W.-X. Zhou, Phys. Rev. E **84**, 016106 (2011).
[43] J. R. M. Hosking, Frac. Differenc. Biometrika **68**, 165 (1981).
[44] B. Podobnik, P. Ch. Ivanov, K. Biljakovic, D. Horvatic, H. E. Stanley, and I. Grosse, Phys. Rev. E **72**, 026121 (2005).
[45] C. Meneveau and K. R. Sreenivasan, Phys. Rev. Lett. **59**, 1424 (1987).
[46] J. Wang, P. Shang, and K. Dong, Applied Math. Comput. **224**, 337 (2013).
[47] C. V. Chianca, A. Tinoca, and T. J. P. Penna, Physica A **357**, 447 (2005).
[48] J. Kwapień, P. Oświęcimka, and S. Drożdż, Physica A **350**, 466 (2005).
[49] S. Drożdż, J. Kwapień, P. Oświęcimka, and R. Rak, Europhys. Lett. **88**, 60003 (2009).
[50] M. S. Movehed and E. Hermanis, Physica A **387**, 915 (2008).
[51] M. S. Movahed, G. R. Jafari, F. Ghasemi, S. Rahvar, and M. R. R. Tabar, J. Stat. Mech. (2006) P02003.
[52] Q. D. Y. Ma, R. P. Bartsch, P. Bernaola-Galván, M. Yoneyama, and P. C. Ivanov, Phys. Rev. E **81**, 031101 (2010).