ARTICLE IN PRESS

Physica A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physica A

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

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

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

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

HIGHLIGHTS

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

ARTICLE INFO

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

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

ABSTRACT

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

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

https://doi.org/10.1016/j.physa.2017.10.006
0378-4371/© 2017 Published by Elsevier B.V.

1. Introduction

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

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

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

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

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

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

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

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

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

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

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

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

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

Please cite this article in press as: S. Feng, et al., A better understanding of long-range temporal dependence of traffic flow time series, Physica A (2017), https://doi.org/10.1016/j.physa.2017.10.006.