Detecting intrinsic dynamics of traffic flow with recurrence analysis and empirical mode decomposition

Hui Xiong, Pengjian Shang*, Songhan Bian

Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, PR China

HIGHLIGHTS

- The intrinsic dynamics of traffic flow are evaluated from a frequency–time perspective.
- Components of medium- and low-frequencies dominate the signal’s apparent dynamics.
- The denoised RQA diversely characterizes the essential properties of the traffic flow.
- The denoised RQA indicates abrupt changes more accurately.
- The proposed analysis sheds more solid, multiple and inherent lights into the traffic system.

ARTICLE INFO

Article history:
Received 20 October 2016
Received in revised form 7 December 2016
Available online 17 January 2017

Keywords:
Recurrence plot
Recurrence quantification analysis
Empirical mode decomposition
Traffic flow
Frequency- and time-evolution
Noise

ABSTRACT

In this paper, we apply the empirical mode decomposition (EMD) method to the recurrence plot (RP) and recurrence quantification analysis (RQA), to evaluate the frequency- and time-evolving dynamics of the traffic flow. Based on the cumulative intrinsic mode functions extracted by the EMD, the frequency-evolving RP regarding different oscillation of modes suggests that apparent dynamics of the data considered are mainly dominated by its components of medium- and low-frequencies while severely affected by fast oscillated noises contained in the signal. Noises are then eliminated to analyze the intrinsic dynamics and consequently, the denoised time-evolving RQA diversely characterizes the properties of the signal and marks crucial points more accurately where white bands in the RP occur, whereas a strongly qualitative agreement exists between all the non-denoised RQA measures. Generally, the EMD combining with the recurrence analysis sheds more reliable, abundant and inherent lights into the traffic flow, which is meaningful to the empirical analysis of complex systems.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

In last decades, a useful tool named recurrence plot (RP) was comprehensively introduced to describe the fundamental properties of nonlinear time series from complex systems [1]. A recurrence means that the recurrent state is somehow similar to a former state in phase space and the RP is the visualization of a binary symmetric square matrix (i.e., recurrence matrix), whereas it is intuitive to interpret patterns and structures revealed by the RP, leading to the propose of recurrence quantification analysis (RQA) [2]. RQA measures (recurrence rate, determinism and average diagonal line length, etc.) are mainly based on the diagonal and vertical line structures of the recurrence matrix, and these RQA measures calculated from sliding window along the main diagonal of the RP yield the time-dependent RQA, which can be used to detect dynamics of

* Corresponding author.
E-mail address: pjshang@bjtu.edu.cn (P. Shang).

http://dx.doi.org/10.1016/j.physa.2017.01.060
0378-4371/© 2017 Elsevier B.V. All rights reserved.
the underlying system [3–5]. Subsequently, qualitative and quantitative analyses of dynamical system based on the RP and RQA or their modifications and extensions, like cross recurrence plot [6], ordinal patterns recurrence plot [7], multivariate recurrence plot [8] and recurrence network [9], have been widely and successfully applied to real-world data in various disciplines and areas, such as in physiology [10,11], in economy [12,13], the study of ecosystem [14], the assessment on underlying statistical properties of short-term traffic flow [15] and characterization of dynamical attractors [16], etc.

As the RP and RQA were getting modified or generalized, some suggestions for selecting the recurrence parameters or avoiding potential pitfalls were proposed [17–23]. Anyway, the choice of parameters depends on the goal of the analysis, for instance, to minimize the errors due to observational noise, the recurrence threshold \( \varepsilon \) should be increased to at least five times larger than the standard deviation of the added noise if the noise level is not too high [24], while for optimal signal classification, the \( \varepsilon \) should be about 5% of the maximal phase space diameter [25]. Besides, the analysis of combined effects of noise and embedding suggested the sensitivity of RQA to noise [26]. Though the RQA would change considerably even in a low level of noise environment, to our knowledge, how noises contained in but not extra added to the series affect the relevance of RP and RQA is not studied sufficiently.

Time series generated from complex systems in real world usually exhibit strong fluctuation and unpredictable perturbation with high level of nonlinearity and non-stationarity, thus it is natural that they are very likely to be affected by noise, which may lead to spurious results. Traffic system is a typical complex system, where signals, like traffic flow (i.e., volume) and travel speed, are mostly studied to uncover underlying structures of the system [15,27–30]. In the field of time series analysis, the cellular automaton models of vehicular traffic were utilized in the characterization of transitions or short-term prediction using complex networks [31–35]. However, considering that traffic signals are interacted by intrinsic and extrinsic factors (e.g., the existence of noise effect), detecting the inherent nature of traffic system through noise reduction is necessary and helpful to unveil rich and accurate information about the dynamical system considered. In various fields, such as traffic signals [36–38], financial market [39], physiologic time series [40,41], wind power [42] and climate [43,44], empirical mode decomposition (EMD) [45] has been used as a powerful method to reduce noises, filter out trends and predict in empirical and complicated data analysis. The EMD, based on the local characteristic time scale of the data, is an adaptive, direct and highly efficient process that will generates a limited number of intrinsic mode functions (IMFs) over frequency and a residual embedded in the signal. Thus, the decomposition can be regarded as an expansion of the data over its IMFs and as an overall frequency–time analysis. Meanwhile, relevant information contained in different-frequency components is preserved, allowing an efficient search of discriminating features based on the cumulative IMFs [46]. Therefore, in this paper, we combine the EMD with RP and RQA and then apply them to traffic time series to analyze its frequency-evolving RP structures and time-evolving RQA dynamics. Noise effect is considered as well. Existing publications using a combination of both approaches like [47,48] are acknowledged. However, to our best of knowledge, this is the first application of the RP and RQA in such a brand-new perspective.

The remainder of this work is organized as follows. We briefly introduce the RP and RQA in Section 2 and the EMD method in Section 3. Then, in Section 4, these methods are applied to traffic time series to reveal its inherent complexity. Finally, Section 5 concludes and some supplementary materials are presented in Appendix.

2. Recurrence plot and recurrence quantification analysis

Recurrence plot (RP) [1,2] measures recurrences of a trajectory \( x_i \in \mathbb{R}^m \) in its m-dimensional phase space by recurrence matrix \( R \), which is defined as

\[
R_{ij}(\varepsilon) = \Theta(\varepsilon - \|x_i - x_j\|), \quad i, j = 1, 2, \ldots, N, \tag{1}
\]

where \( N \) is the trajectory length, \( \varepsilon \) is a recurrence threshold, \( \Theta(\cdot) \) is the Heaviside function and \( \| \cdot \| \) is a norm. Then, \( R_{ij}(\varepsilon) \) is one if the state of the system at time \( i \) and that at time \( j \) have a distance less than \( \varepsilon \), and zero otherwise. The phase space trajectory \( x_i \) is reconstructed from a time series \( \{u_i\}_{i=1}^N \) by time delay embedding

\[
x_i = (u_i, u_{i+\tau}, u_{i+2\tau}, \ldots, u_{i+(m-1)\tau}), \tag{2}
\]

where \( m \) is the embedding dimension and \( \tau \) is the time delay.

The RP is obtained by plotting the recurrence matrix \( R \) in its binary entries: a black dot if \( R_{ij} = 1 \) (recurrence) and a white dot if \( R_{ij} = 0 \) (no recurrence). Thus, by the definition of \( R \), the RP has always a black main diagonal line and is symmetric with respect to the main diagonal. Instead of plotting the recurrence matrix, an unthresholded RP based on the distance matrix \( D \),

\[
D_{ij} = \|x_i - x_j\|, \quad i, j = 1, 2, \ldots, N, \tag{3}
\]

can be plotted and called distance plot (DP).

As the initial purpose of RPs was to visualize trajectories in phase space, the interpretation of RPs is qualitative and subjective, therefore, lacking quantitative standards. Hence, a quantification of the obtained structures of RPs is necessary for a more objective investigation of the considered system. To this end, recurrence quantification analysis (RQA) was proposed as a quantitative description of RPs, whose diagonal and vertical lines are the base of the RQA [3]. Furthermore, the RQA can
provide information about different characteristics of the RP and can be used to detect transitions in dynamical systems. In this work, the used RQA measures are as follows [2]:

**Recurrence rate**, $RR$, is a measure of the density of recurrence points in the RP,

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}(\varepsilon),$$

(4)

where $R_{i,j}(\varepsilon)$ is the recurrence matrix defined in Eq. (1) and the main diagonal ($i = j$) is usually excluded. The RR can be used to reveal changes in dynamical systems.

**Determinism**, $DET$, is the ratio of recurrence points that form diagonal structures of at least length $l_{\text{min}}$ to all recurrence points,

$$DET = \frac{\sum_{l=l_{\text{min}}}^{N} P(l)}{\sum_{l=1}^{N} P(l)},$$

(5)

where $P(l)$ is the histogram of diagonal lines of length $l$ given the threshold $\varepsilon$. Processes with uncorrelated or weakly correlated, stochastic or chaotic behavior cause none or very short diagonals, whereas deterministic processes cause longer diagonals and less single, isolated recurrence points. Therefore, the DET provides a measure for determinism or predictability of the system.

**Average diagonal line length**, $L$, is the average time that two segments of the trajectory are close to each other,

$$L = \frac{\sum_{l=l_{\text{min}}}^{N} IP(l)}{\sum_{l=1}^{N} P(l)},$$

(6)

which can be interpreted as the mean prediction time.

**Longest diagonal line**, $L_{\text{max}}$, is related to the exponential divergence of the phase space trajectory,

$$L_{\text{max}} = \max(l_i)_{l_i=l_{\text{min}}}^{N},$$

(7)

where $N_l = \sum_{l=l_{\text{min}}}^{N} P(l)$ is the total number of diagonal lines. The faster the trajectory segments diverge, the lower is the $L_{\text{max}}$. Analogously to the $L_{\text{max}}$, the maximal length of the vertical lines, $V_{\text{max}}$, in the RP can be regarded as

$$V_{\text{max}} = \max(v_i)_{v_i=v_{\text{min}}}^{N},$$

(8)

where $N_v$ is the absolute number of vertical lines.

**Entropy**, $ENTR$, refers to the Shannon entropy of the probability $p(l) = P(l)/N_l$ to find a diagonal line of exactly length $l$ in the RP,

$$ENTR = - \sum_{l=l_{\text{min}}}^{N} p(l) \ln p(l).$$

(9)

$ENTR$ reflects the complexity of the RP in respect of the diagonal lines. Therefore, for uncorrelated noise, the value of $ENTR$ is rather small, indicating its low complexity.

**Laminarity**, $LAM$, is the ratio between the recurrence points forming the vertical structures and the entire set of recurrence points,

$$LAM = \frac{\sum_{v=v_{\text{min}}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}.$$

(10)

It is analogous to the $DET$ in Eq. (5), representing the occurrence of laminar states in the system without describing the length of these laminar phases.

### 3. Empirical mode decomposition

Empirical mode decomposition (EMD) is a data-driven and efficient decomposition of a time series into a finite number of independent and concretely implicational intrinsic mode functions (IMFs) and a residual [45]. The IMFs, which characterize the oscillation modes embedded in the data, are iteratively extracted by a sifting process if it satisfies two conditions: (1) in the whole data set, the number of local extrema and the number of zero crossings must either equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. Hence, an IMF involves only one mode of oscillation, which can be non-stationary and both amplitude and frequency modulated.

---

1. The implementation of the RP and RQA in this paper is based on the CRP toolbox for Matlab provided by TOCSY at [http://tocsy.agnld.uni-potsdam.de](http://tocsy.agnld.uni-potsdam.de).
The extraction of the IMFs of a series $x$ with length $n$ is summarized as follows:

1. Extract all the extrema of the series $x$.
2. Generate the upper and lower envelopes, $e_{\text{max}}$ and $e_{\text{min}}$, by interpolating between minima and maxima separately.
3. Compute the average of the two envelopes, $a = (e_{\text{min}} + e_{\text{max}})/2$.
4. Extract an IMF candidate $c$ by subtracting $a$ from the data, $c = x - a$.
5. Iterate the above steps on the residual $r$, $r = x - c$, if $c$ is an IMF. Otherwise, replace $x$ with $c$.
6. Repeat the above steps and end the shifting process when it satisfies a predefined stopping criterion.

Therefore, the series $x$ can be finally represented as the sum of a collection of IMFs with frequencies from high to low and a residual representing the trend of the series, namely,

$$x = \sum_{t=1}^{k} c_t + r,$$

where $k$ is the total number of extracted IMFs and $c_t$ denotes the $t$th IMF in descending order of frequency. Given an IMF $c$, its energy [37] is defined as

$$|c| = \sqrt{c^2(1) + c^2(2) + \cdots + c^2(n)}.$$
Fig. 3. The RPs for part of the traffic flow with varied lengths of 5000 (A), 2000 (B) and 720 (C). Parameters $m = 1$ and $\varepsilon$ is 10% of the standard deviation of the series considered, using the Euclidean norm.

Fig. 4. The extracted IMFs and residual of the selected traffic flow by the EMD method.

In the denoising process, the first group of IMFs whose energies are monotonically decreasing can be viewed as noises and then, we can directly remove these noisy components to reconstruct a denoised signal. As a main criterion to evaluate the quality of noise reduction, SNR (signal to noise ratio) is computed by

$$\text{SNR} = 10 \times \lg \frac{\sum_{s=1}^{n} [x(s) - \bar{x}]^2}{\text{MSE}},$$

where MSE is the mean square error, $\sum_{s=1}^{n} [x(s) - x^\dagger(s)]^2 / n$, between the original and denoised signals ($x$ and $x^\dagger$, respectively), and $\bar{x}$ is the average of $x$. A large SNR means that the reconstructed signal is with low level of noises, indicating that noises contained in the original signal are effectively reduced.
Fig. 5. Frequency-evolving RPs of the traffic flow according to Fig. 4 and Eq. (14).

Furthermore, the cumulative sums of the IMFs, beginning with the fastest oscillated IMF, yield a multilevel filtering of the original signal [46]. The cumulative IMF (CIMF) summing up to order $k$ is defined by

$$CIMF_k = \sum_{j=1}^{k} c_j,$$

(14)
As components of lower frequency are gradually added and relevant information carried by different-frequency components is preserved, evaluation based on the CIMF enables the analysis of the evolution of the signal regularity over different oscillation levels.

4. Empirical analysis of traffic flow

4.1. Data set

The data involved here are the volume data (unit: vehicles/km, i.e., the number of vehicles in one kilometer around the detector) gathered from the detector D3054 on the North 3rd Ring Road in Beijing, China, as in Fig. 1. The raw data, downloaded from the Highway Performance Measurement Project run by Beijing STONG Intelligent Transportation System Co. Ltd., were recorded about every two minutes from September 8th to October 5th, 2012. The total sample size is 20,067. Fig. 2 depicts the normalized volume after averaging over three lanes, where we can see that it increases and then decreases about every three hours during the day, and that peaks occur over the periods 7–10AM (including the early rush-hour) and 16–19PM (including the late rush-hour). Furthermore, as suggested in [27], traffic signals have the period of 24 h that can be considered as a quasi-period because of the similar interday road conditions. Consequently, the structure of the corresponding RP is intuitively periodic and same pattern reappears as illustrated in Fig. 3. Therefore, it is redundant to evaluate the entire data set since the RP in Fig. 3(C) with size that is about 24 h (720 data points) will suffice to reveal essential properties of the signal. In the following, we put our emphasis on the data subset presented in the middle panel of Fig. 2, to reveal the inherent intraday and interday dynamics of the traffic flow.

4.2. Selection of the parameters

Generally, the mutual information (MI) function and false nearest neighbors (FNN) method are common and robust tools to respectively decide appropriate time delay \( \tau \) and embedding dimension \( m \) [18,19]. Nevertheless, time-series embedding is not absolutely required for RPs and RQA as trajectories might be over-reconstructed due to an inappropriate \( m \) or the effect of noise. As suggested in [20], striking structural similarity between embedded and unembedded RPs exists and the only apparent difference is a lightening of the RP with increasing embedding dimension (same as RPs and DPs in Fig. A.2 for the series of interest). Hence, it is not strictly necessary to reconstruct RPs for a variety of embedding parameters, because an embedded RP can be generated from an unembedded RP (i.e., \( m = 1 \)) that contains the information about statistics of all possible embedded RPs [20,21]. Besides, the embedding procedure for time series with observational noise deviates from the RP of the underlying process [24]. As our main interest here involves the noise effect on dynamics of the traffic flow, in this work, we apply the recurrence analysis with no embedding to exclude the influence of noise on the structures of reconstructed trajectories. The threshold \( \varepsilon \) is selected to be 10\% of the standard deviation of the considered series, using the Euclidean norm, and other parameters are defaulted using the CRP toolbox if not specified.

4.3. Frequency-evolution of the RP structures

Typology and texture [2] are typical patterns in the RP in respective large- and small-scale, and on the basis of these patterns that are linked to specific behavior of the system, we can gain insights into the RP in a qualitative way. For the traffic flow, its RP structure is quasi-periodically recurrent and from Fig. 3(C), we can find that the analyzed data within a period (24 h, about 720 data points) are non-stationary so that the RP pales away from the main diagonal, vanishing to the upper-left and lower-right corners. On the other hand, there is approximately homogeneous structure in the bigger
Fig. 7. The RP of the denoised volume.

Fig. 8. Time-evolving RQA of the traffic flow before (black) and after (red) noise reduction. Parameters window-size $WL = 60$ and shift-size $SL = 10$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

black squared block, indicating the stationary process embedded in the non-stationary system. White bands in the RP imply abrupt changes that the traffic flow increases or decreases sharply. Besides, the processes trapping to the two black blocks form vertical and horizontal clusters, and minor intermittencies exist within the bigger laminar zone that results from more frequent and active human activities during this period of time. Moreover, there is chaotic process between the two laminar zones. Therefore, for the traffic flow, its RP structure repeats every 24 h with a general laminar–chaotic–laminar pattern.

In addition to analyzing the considered data set completely, we evaluated the RP structure of the series over different levels of oscillation on the basis of the CIMF defined in Eq. (14) and then examined the corresponding RPs to show its frequency-evolutionary process. We first used the EMD method to decompose the traffic flow and Fig. 4 shows the extracted IMFs and a residual. Correspondingly, the RPs and RQA of the CIMFs are separately presented in Fig. 5 and Table 1. An evolution regarding changes in the RP structures over descending frequency can be clearly seen in Fig. 5. Specifically, the RPs of the first four IMFs are nearly homogeneous or in random states. Visible white bands within the block 00:00–06:40
Table 1
The RQA measures of the traffic flow in terms of frequency according to the RPs in Fig. 5.

| CIMF1 | 0.0563 | 0.2176 | 2.2282 | 10 | 0.5877 | 0.0969 | 4 |
| CIMF2 | 0.0601 | 0.1316 | 2.0899 | 6 | 0.3090 | 0.1498 | 4 |
| CIMF3 | 0.0616 | 0.1310 | 2.0920 | 6 | 0.3123 | 0.1680 | 4 |
| CIMF4 | 0.0615 | 0.1410 | 2.0971 | 6 | 0.3271 | 0.1852 | 4 |
| CIMF5 | 0.0565 | 0.1386 | 2.0930 | 5 | 0.3177 | 0.2043 | 5 |
| CIMF6 | 0.0571 | 0.1988 | 2.1428 | 6 | 0.4296 | 0.2846 | 8 |
| CIMF7 | 0.0559 | 0.2900 | 2.2620 | 11 | 0.6386 | 0.3980 | 12 |
| CIMF8 | 0.0728 | 0.3837 | 2.4371 | 15 | 0.8597 | 0.5012 | 23 |
| Volume | 0.0964 | 0.4116 | 2.4386 | 21 | 0.8641 | 0.5227 | 23 |

Table 2
The SNRs of the CIMFs and the denoised volume.

<table>
<thead>
<tr>
<th>SNR</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIMF1</td>
<td>31.63</td>
</tr>
<tr>
<td>CIMF2</td>
<td>31.68</td>
</tr>
<tr>
<td>CIMF3</td>
<td>31.72</td>
</tr>
<tr>
<td>CIMF4</td>
<td>31.82</td>
</tr>
<tr>
<td>CIMF5</td>
<td>31.43</td>
</tr>
<tr>
<td>CIMF6</td>
<td>33.09</td>
</tr>
<tr>
<td>CIMF7</td>
<td>33.50</td>
</tr>
<tr>
<td>CIMF8</td>
<td>39.34</td>
</tr>
<tr>
<td>Denoised volume</td>
<td>45.54</td>
</tr>
</tbody>
</table>

emerge when the CIMF sums up to order 5, indicating the presence of abrupt changes or rare events that the traffic flow drastically declines or increases. At the same time, the two laminar states start appearing apparently, supported by the gradually increasing LAM. Moreover, the RQA measures of the CIMFs are generally increasing across decreasing frequency. The rising RQA from the 5th CIMF signifies a more deterministic, complex and slower divergent structure of the traffic flow in terms of frequency. Furthermore, we computed the SNRs of the CIMFs with the original signal, viewing the CIMFs as denoised ones (Table 2). Relatively small SNRs suggest that there are more noises ($\approx 3.16\% = 1/\text{SNR} \cdot 100\%$) contained in highly oscillated components, i.e., the first three IMFs, which is supported by the corresponding smaller ENTR that implies low complexity. Overall, dynamics of the analyzed data are mainly dominated by components of medium- and low-frequencies while severely affected by high-frequency components that cause homogenous states embedded in the nonlinear and non-stationary process (same result can be obtained with embedding as shown in Fig. A.3 and with varied sample size). Sudden changes of the RQA as well as the SNRs from CIMF4 to CIMF5 and from CIMF7 to CIMF8 mark critical stage points in the frequency-evolutionary process of the RP.

4.4. Intrinsic time-evolving dynamics of the traffic flow

As illustrated before, noise affects the reliability of the RP and RQA, and dynamics of traffic signals are influenced by the existence of noises. Therefore, to investigate the intrinsic dynamics of the traffic flow, the first three highly oscillated IMFs were viewed as noises and then removed according to the energy distribution of the IMFs in Fig. 6.

The large SNR of the denoised volume exhibits the ability of the filtering process, after which only 2.20\% (SNR = 45.54) of noises remain (Table 2). As a consequence, difference between the RPs before (Fig. 5(I)) and after (Fig. 7) noise reduction is evident. Specifically, the denoised RP is darker-shaded as random process embedded in the non-stationary system has basically been removed, implying the increase of recurrence density. Meanwhile, white bands within clusters become more obvious and indicate abrupt changes more accurately, e.g., at 16:20 and 19:39. Quantitatively, comparison of the
Fig. A2. The RPs and corresponding DPs of the traffic flow for $\tau = 1, 2$ and $m = 1, 2, 3, 4$, respectively, according to Fig. A1.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>RR</th>
<th>DET</th>
<th>L</th>
<th>$l_{\text{max}}$</th>
<th>ENTR</th>
<th>LAM</th>
<th>$V_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>0.0964</td>
<td>0.4116</td>
<td>2.4386</td>
<td>21</td>
<td>0.8641</td>
<td>0.5227</td>
<td>23</td>
</tr>
<tr>
<td>Denoised volume</td>
<td>0.1235</td>
<td>0.9985</td>
<td>9.5298</td>
<td>987</td>
<td>3.0018</td>
<td>0.9993</td>
<td>86</td>
</tr>
</tbody>
</table>

The RQA measures of the original and denoised volume.

corresponding RQA measures for the volume with and without noises is displayed in Table 3, where we can find that all the RQA measures of the denoised volume are much larger than those of the original one, identical with the result that the magnitude of all RQA variables would decrease with increasing noise amplitude in [26] (through viewing the noise amplitude of the denoised signal as zero). In general, larger RQA measures manifest that the denoised system is more predictable, slower divergent and much more complex than the one with noises. In addition, the increased LAM implies that there are more vertical structures in the denoised RP, suggesting rich laminar states in the underlying system.
In order to detect intrinsic dynamical changes of the traffic flow across time, we then utilized the RQA with sliding window of window-size WL = 60 ($\approx$2 h) and shift-size SL = 10 ($\approx$20 min) to get the time-dependent RQA to reveal the time evolution of the series considered. The results are shown in Fig. 8, where it is clear to see that all the RQA measures of the denoised volume are quite larger than those of the noisy one. Moreover, there are more intriguing findings, qualitatively and quantitatively, (1) for the original time series, a strongly qualitative agreement exists between all the RQA measures and details of the dynamics are lost, whereas the denoised RQA exhibits pronounced diversity and their peaks or lower points better label corresponding crucial points that the volume changes drastically, e.g., at 12:59; (2) two types of RRs are highly synchronous over the period 23:00-06:40, which corresponds to the interday transition; (3) the DETs of the denoised volume fluctuate in the interval of 0.96 to 1 and drop or rise dramatically over white areas. Though relatively small DETs occur within the bigger laminar phase between 10:00 and 19:39, 100% of the denoised data has the DET above 0.96, implying the long diagonals produced in the corresponding RP and the presence of plentiful deterministic structures in the underlying dynamical system; (4) correspondingly, the denoised Ls largely increase, ranging from 5.36 to 30.43 and larger than the maximal L ($=4.63$) of the non-denoised data; (5) same as the DETs of the denoised volume, the denoised ENTRs fluctuate relatively stably over the range of 1.84 to 4.04, and 94.48% of them are larger than the maximal ENTR ($=2.08$) of the non-denoised one, manifesting greater intrinsic complexity of the system; (6) the denoised L and ENTR are basically...
synchronous with the denoised DET and have larger values at more deterministic parts; (7) the LAMs of the denoised volume vanish over non-laminar phases, providing an indicator for the detection of laminarity–chaos transition. Moreover, 100% of the denoised LAMs are above 0.93, signifying rich trapping states and rare single recurrent points in the system; (8) above analyses are quantitatively in accordance with the overall results from Table 3; (9) similar results can be obtained with embedding (Fig. A.4) and for different sliding parameters WL = 30, 40, 70 and SL = 5, 10, 15 with no embedding (Fig. A.5); and (10) at last, due to the periodic trend of the traffic flow, similar results can be obtained for the analyzed data with varied length (not shown), showing the feasibility of the proposed analysis. Generally, on the basis of the above analyses, we can draw the conclusion that the combination of EMD with RP and RQA can provide more solid and essential information about the traffic flow, enabling the revelation of inherent dynamics and providing multiple insights into the underlying system.

5. Conclusive remarks

In this paper, we analyzed the different-frequency components and the time-evolving dynamics of the traffic flow (i.e., volume), through applying the EMD method to the RP and RQA. For the traffic flow, its RP structure is quasi-periodically recurrent with a laminar–chaotic–laminar pattern and there are homogenous states embedded in the nonlinear and non-stationary traffic system. Based on the CLMFs, the frequency-evolving RP of the volume data indicates that apparent dynamics of the considered data are mainly controlled by its medium- and low-frequency components, and in the meantime, strongly affected by components of high-frequency. Furthermore, the approximately rising RQA measures are able to mark important stages in the process of frequency evolution. Then, to detect the intrinsic dynamics of the traffic flow, we filtered out noises contained in the volume data and compared the results with those from the non-denoised one. Consequently, we found that stationary processes embedded in the non-stationary system have basically been removed and that the denoised volume is less divergent, much more complex and with richer deterministic and laminar structures than the one containing noises. Moreover, the time-dependent RQA suggested that all RQA measures of the original signal have strongly qualitatively identical evolving patterns, while the denoised RQA diversely characterizes the properties of the system so that they better label critical points where white bands in the RP occur. Overall, through the frequency evolution and inherent complexity detection of the traffic flow over time, we gained deeper and multiple insights into the underlying dynamical system.

As far as we are aware, though the influence of noises extra added to generated time series on the RP/RQA has been studied before, how noises contained in empirical data from real-world complex system affects their performances is studied here for the first time, and this is the first application of RP and RQA in traffic system from such a perspective that concerns different oscillation of modes and noise effect. Moreover, the combination of EMD with RP and RQA is feasible under different cases, showing the effectiveness of the proposed method. Conclusively, in this work, the combination of both approaches offered us more reliable and essential information about the traffic system from an overall frequency–time perspective, which enables the revelation of inherent dynamics of the series and provides a suggestion on the empirical analysis of complicated data with noises for other complex systems using the recurrence approach.
(A) WL = 30, SL = 5.  
(B) WL = 30, SL = 10.  
(C) WL = 30, SL = 15.  
(D) WL = 40, SL = 5.  
(E) WL = 40, SL = 10.  
(F) WL = 40, SL = 15.  
(G) WL = 70, SL = 5.  
(H) WL = 70, SL = 10.  
(I) WL = 70, SL = 15.

Fig. A.5. Time-evolving RQA of the traffic flow with sliding parameters WL = 30, 40, 70 and SL = 5, 10, 15, respectively.

Acknowledgment

We acknowledge valuable comments from anonymous referees and financial support from the China Postdoctoral Science Foundation Funded Project (B16M00090), the China National Science (61371130) and the Beijing National Science (4162047).

Appendix

Here, we present some supplementary figures. Fig. A.1 shows the optimal recurrence parameters for the data considered by the MI and FNN methods. According to Fig. A.1, the corresponding RPs and DPs in some cases are illustrated in Figs. A.2,
and A.3 and A.4 give the frequency-evolving DPs (for clearer graphic presentation) and time-evolving RQA with embedding. Finally, Fig. A.5 gives the unembedded time-dependent RQA for different sliding parameters.

References