**Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013**

MoB2.1

# Exploring urban traffic flow dynamics

Adam Malcontenti-Wilson[1], Hai L. Vu[1], Peter Cadusch[2], Philip Branch[1], Hans van Lint[3], and Serge Hoogendoorn[3]

***Abstract*— Important statistical characteristics of traffic can be examined via the reconstruction of a so-called phase-space where each point in that space represents a state of traffic and vice versa. Statistical tools such as recurrence analysis can then be used to visualize the temporal evolution of traffic and to assess its non-stationarity structure as an indicator for irregular behaviour. In this paper we investigate, examine and analyze the short-term non-stationary features present in traffic flows based on the historical data available from stop-line detectors collected from a small number of intersections in the Melbourne CBD, Australia and demonstrate that recurrence analysis has considerable potential in identifying changes within the temporal behaviour of traffic.**

## I. INTRODUCTION

Analyzing data collected from road traffic networks can reveal valuable insights that can be used to manage and control that network. Recent studies [15], [16] pointed out that the process underlying traffic is indeed very complex and its analysis requires better methods as an alternative to the well known and often used linear e.g. [3] and non-linear time series analyses e.g. [14]. Statistical tools such as recurrence analysis [7] offer a new way to characterize the behaviour of a dynamical complex system. It is based on the observation that in many systems, similar situations often evolve similarly (i.e. determinism property) and certain events are repeated (the re-occurrences of the system former state). Utilizing these, the recurrence analysis starts with a process of constructing multi-dimensional embedded vector space (phase space) from a scalar traffic data measurement such as traffic volume. It then compares the states of the system (i.e. looking for recurrent state or similarity) by calculating the distance between all pairs of embedded vectors and assessing it against a predefined threshold.

In this paper we will use recurrence analysis to analyze one year worth of traffic data collected from a small number of intersections in the Melbourne CBD, Australia. The data is loop detector data that has been collected by the Sydney Coordinated Adaptive Traffic System (SCATS) which controls traffic signals at intersections around Melbourne. The analysis is part of a larger project in the development of predictive methods which can better model traffic flow and improve road system capacity via more efficient controlling of traffic signalization. Herein we will report preliminary results obtained from the recurrence analysis and demonstrate that this method has considerable potential in identifying changes within the temporal behaviour of traffic.

There has been much work in the literature dealing with traffic analysis based on SCATS or similar data representing typical traffic variables such as volume, lane occupancy and speed. In particular, Nguyen and Gaffney discusses the ARTIS system which identifies congestion and estimates travel times using an unspecified algorithm, based on urban signalized arterials in Melbourne, Australia [11]. Suksri et al. introduces analysis of SCATS data from the Adelaide CBD signalized intersections using spectral density functions (SPF) and autocorrelation function (ACF) plots, allowing them to differentiate the data characteristics and lane configurations from the data set [12]. Vlahogianni et al. proposes a multilayer strategy to identify patterns of traffic based on their structure and evolution in time using both volume and occupancy. The authors use measures in the Cross Recurrence Quantification Analysis (CRQA) as inputs to a Self Organizing Map to reduce the multi-dimensional data to a two-dimensional mapping which then can be clustered using a k-means algorithm with traffic flows on a urban signalized arterial in Athens, Greece [15]. Vlahogianni then builds upon this to make short-term predictions based on the patterns identified for each regime of traffic flow independently by using multilayer feed-forward neural networks (MLP) [16]. Masugi uses a similar approach of CRQA with the non-stationary time-signals of IP-Network Traffic, also including Detrended Fluctuation Analysis (DFA) to analyze the Long-Range Dependence (LRD) to overall asses the dynamical transitions over time [9], [10].

The paper is structured as follows. In section II we begin with brief summary of relevant road traffic theory and give an overview of data provided by the Sydney Coordinated Adaptive Traffic System (SCATS). Section III provides a formal description of recurrence plot and recurrence analysis with measures to be used later in our study. In Section IV we present and discuss our results and observations. Finally, the conclusion is given in Section V.

[1]A. Malcontenti-Wilson, H. L. Vu and P. Branch are with the Faculty of ICT, Swinburne University of Technology.
`email: 9509453 at student.swin.edu.au, {hvu, pbranch} at swin.edu.au`

[2]P. Cadusch is with the Faculty of Engineering and Industrial Science, Swinburne University of Technology.
`email: pcadusch at swin.edu.au`

[3]H. van Lint and S. Hoogendoorn are with the Faculty of Civil Engineering and Geosciences of the Delft University of Technology.
`email: {h.v.lint, s.p.hoogendoorn} at tudelft.nl`

Hai L. Vu was supported by the Australian Research Council (ARC) Future Fellowships grant FT120100723.

## II. Urban traffic operations and SCATS data

### A. Traffic Theory

Road traffic is typically understood in terms of different phases where each phase has its own behavioural characteristics. The classical theory of traffic flow assumes two phases: free flow and congested flow. Kerner et al. [6] extend this classical theory by replacing congested flow with two phases of congestion: synchronized flow and wide moving jam. In this paper we adopt the classical theory but note the existence of a transition zone between the phases.

The fundamental mathematical relationship that describes traffic flow is that volume (sometimes "flow") is the product of density and mean speed. Volume is the number of vehicles per unit time passing a given point on the road, usually expressed as vehicles per hour (veh/h). Density is the number of vehicles per kilometer (veh/km).

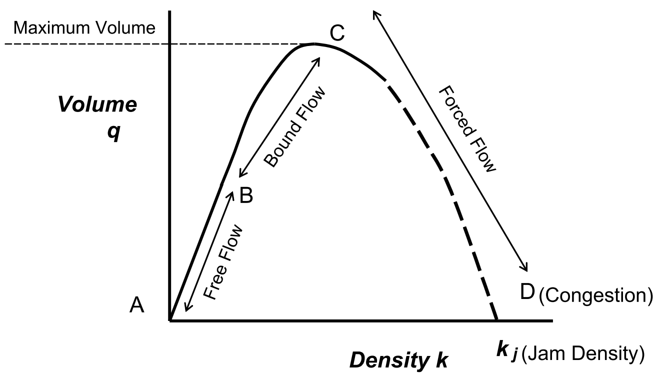


Fig. 1. Typical Volume-Density relationship of uninterrupted flow [1, Fig. 2.3], modified to include flow annotations.

The relationship between volume and density can be described by a Volume Density diagram (or fundamental diagram) as shown in Fig. 1. 'Free flow' traffic condition is characterized by a strong positive correlation between Volume and Density up to the point of maximum volume and critical density. In this phase individual vehicles suffer little restriction from the presence of other traffic. As the density increases traffic enters 'bound' or 'transitional' flow as vehicle speed is bound by that of vehicles around it. This phase is characterized by unstable flow conditions. Beyond the critical density traffic enters 'forced flow' conditions in which the speed of the traffic decreases until the traffic is stationary ('jam').

In experimental observations of a single lane or detector, typically the volume-density plot is more triangular in shape, due to smaller amount of data available as well as speed limitations restricting road users from travelling faster than the speed limit posted for that road [4], [1, pp. 65 - 66].

### B. SCATS

The Sydney Coordinated Adaptive Traffic System (SCATS) is an adaptive traffic signal control system used in Melbourne by VicRoads to manage signalized intersections in the Melbourne road network. Vehicle loop detectors are installed at intersections which provide SCATS with traffic count and occupancy data. The former is essentially a vehicle count during a green time for each traffic movement, while the latter is the portion of time when vehicles are within the sensor zone of the loop detectors.

SCATS data consists of a sequence of records, each is produced every cycle and includes:

- Site identifier,
- Lane number and detector identifiers,
- Maximum hourly flow,
- Location,
- Traffic movements possible from lane,
- Sampling date and time (to nearest minute),
- Cycle time in seconds,
- Green (Phase) time in seconds,
- Number of cars that passed the detector during the cycle ($VO$),
- Degree of saturation ($DS$) defined below,
- Estimated traffic count or reconstituted volume ($VK$) defined below.

The $DS$ is a percentage measure of the utilization of the carrying capacity of the lane. It is defined as:

$$DS = \frac{GT - S_{act} + nS_{MF}}{GT} \quad \% \tag{1}$$

where $GT$ is the green time in seconds, $S_{act}$ is the total space time in that period, $S_{MF}$ is the maximum flow and $n$ is the number of vehicle spaces counted.

The estimated traffic count (or reconstituted volume) is a measure of the estimated traffic flow having the same DS under free flow conditions. It is defined as:

$$VK = GT \times DS \times MF. \tag{2}$$

The Maximum Hourly Flow ($MF$) is calculated per lane, daily as the moving average of the daily Maximum Volume taken over one week. Within SCATS the $DS$, $VK$ and $VO$ values can be used to indicate congestion whenever they fulfill the following conditions: $VK/VO$ exceeds 2.4 and $DS$ exceeds 0.95.

## III. Methods

Recurrence plots and Cross Recurrence plots are visual, non-linear analysis techniques based on two fundamental aspects of time-dependent systems:

1) similar situations often evolve in a similar way,
2) some situations recur [7].

The method of Recurrence Plots, introduced by Eckmann et al. [2], visualizes the recurrences of dynamical systems in a phase space trajectory by considering a series of vectors in an abstract mathematical space, and plotting at which points in time they recur (approximately) to a previous state.

Extending the recurrence plot to bi-variate case is the Cross Recurrence Plot, allowing analysis of dependencies between two different systems [17], [8], [7]. The corresponding cross recurrence matrix of two dynamical systems, each one represented by the trajectories $\vec{x}_i$ and $\vec{y}_i$, is defined by

$$\mathbf{CR}_{i,j}^{\vec{x},\vec{y}} = \Theta(\varepsilon - \|\vec{x}_i - \vec{y}_j\|),\ i = 1,\ldots,N,\ j = 1,\ldots,M \tag{3}$$

978-1-4799-2914-613/$31.00 ©2013 IEEE

where N is the number of measured points $\vec{x}_i$, $\varepsilon$ is a threshold distance, $\Theta(\cdot)$ the Heaviside function and $\|\cdot\|$ is a norm (typically the Euclidian norm) [7].

Selection of the threshold parameter of the recurrence plot ($\varepsilon$) is crucial, and must be chosen appropriately. When the threshold is too small, there will be an inadequate number of recurrence points, leading to insufficient information for learning anything about the recurring structure of the underlying system. However if the threshold is chosen to be too large, almost every point on the phase-space trajectory is similar to every other, leading to meaningless artefact on the plot. Tangential motion, which causes thicker and longer diagonal structures in the recurrence plot than they should be, are caused by consecutive points on the trajectory being considered as neighbours due to a large threshold. The influence of noise however may suggest choosing a larger threshold to preserve the structure of the Recurrence Plot, as otherwise noise may distort any existing structure. Therefore, the threshold must be a carefully selected compromise taking these items into account [7].

The multi-dimensional trajectories of a dynamical system are often not known and cannot be directly observed. Instead only a time-discrete measurement of a single observable parameter is available. In this case, the phase-space must be reconstructed in order to analyze the time series by the recurrence plot method.

A frequently used method of phase-space reconstruction is the time delay method of embedding,

$$\hat{\vec{x}_i} = \sum_{j=1}^{m} u_{i+(j-1)\tau}\vec{e_j} \tag{4}$$

where $m$ is the embedding dimension, $\tau$ is the time delay and $u_{i+(j-1)\tau}$ are the sample points. The vectors $\vec{e_i}$ are unit vectors and span an orthogonal coordinate system. [7]. Both of these parameters must be chosen appropriately, the false nearest-neighbours algorithm [5] provides a reasonable estimation of the smallest sufficient embedding dimension.

The phase-space trajectory, as made up of vectors of arbitrary dimension, allows for the use of multiple variables in the analysis either as separate components or by concatenating the vectors of multiple time-delay embedded trajectories.

### A. Recurrence Quantification Analysis

Recurrence Quantification Analysis (RQA) and Cross Recurrence Quantification Analysis (CRQA) quantify the small-scale structures in Recurrence Plots and Cross Recurrence Plots respectively. Note that Recurrence Plots will have a diagonal Line of Identity (LOI) where the same states are compared and found to be recurrent, hence is excluded from the measures below.

RR Recurrence Rate ($rr$) is the simplest measure of RQA and is a measure of the density of recurrence points in the plot [7]

$$rr = \frac{1}{N^2}\sum_{i,j=1}^{N}\mathbf{R}_{i,j}. \tag{5}$$

DET Determinism (DET) is the ratio between recurrence points forming diagonal lines greater than the minimum length ( $l_{min}$ ) with the number of recurrent points in total.

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \tag{6}$$

where $P(l)$ is the histogram of diagonal lines of length $l$ [7]:

$$P(l) = \sum_{i,j=1}^{N}(1-\mathbf{R}_{i-1,j-1})(1-\mathbf{R}_{i+l,j+l})\prod_{k=0}^{l-1}\mathbf{R}_{i+k,j+k}. \tag{7}$$

In the above equation $\mathbf{R}_{i,j}$ is the distance between samples $i$ and $j$. As processes with uncorrelated or weakly correlated, stochastic or chaotic (i.e. random) behaviour cause none or very short diagonals and deterministic processes cause longer diagonals and less single, isolated recurrence points, DET is a measure of the determinism of the system. The threshold $l_{min}$ excludes the diagonal lines which are formed by the tangential motion of the phase space trajectory [7].

The DET measure based on diagonal lines allows for the chaos-order transitions to be identified [7], [13] and will be utilized in the next section.

## IV. Results and discussion

### A. Data

As mentioned earlier SCATS utilizes inductive loop detectors at the stop line of intersections to collect data, such as traffic count, maximum traffic flow and the corresponding vehicle occupancy, in order to detect changes of the traffic and to adjust the signal timings appropriately. For this reason the availability and accuracy of SCATS data depend on the reliability of the loop detectors and their robustness in terms of detecting vehicles at intersections. For example the detector might under-count traffic by performing only a single count for a possible platoon of vehicles where the gap between them is shorter than the detector loop length of 4.5m.

Our initial study into the available SCATS data for Melbourne indicates that the data appears to be limited in terms of the number of intersections and lane detectors available, and contains spurious (green) phase time information. This can cause large spikes in vehicle $volume$ [veh/h] defined as the traffic count (VO) divided by the corresponding phase time that was abnormally short. To this end, only data at intersections with few to none abnormalities is studied, and data is filtered to remove any unrealistically high volumes in the traffic count. It should also be noted that stop line detectors by their very nature cannot differentiate turning traffic from through traffic and that recorded counts may be artificially decreased when vehicles must wait for turning traffic ahead of them. Likewise, outermost lanes can exhibit abnormal traffic flows due to stopped traffic, and other miscellaneous reasons.

In particular, a route along Flinders St, a major thoroughfare within the Melbourne CBD, is considered, focusing on the Westbound stop-line detectors of the Exhibition St (Site 4560) and Swanston St (Site 4562) intersections, as indicated in Fig. 2. The layout of the Flinders and Exhibition intersection is shown in Fig. 3 where data from the through lane on the Flinders St (Detector 7) is of interest.


Fig. 2. Extract of the Melbourne CBD from Google map with studied intersections highlighted.


Fig. 3. Layout of an intersection (Site 4560) at Flinders St and Exhibition St.

The volume (calculated as VO/(phase time) [veh/h]), the degree of saturation (DS) [%] and congestion indicator as defined in Section II-B are plotted for the above two intersections in (Fig. 4) where the data points were collected on the thru lane and smoothed using a moving average. Although small time scales cannot be considered when smoothed, it provides valuable insight in regard to the daily trend. Near the location of congestion (indicated by green bars), the volume decreased while the DS increased which is expected from the fundamental diagram in the area of congested or forced flow.

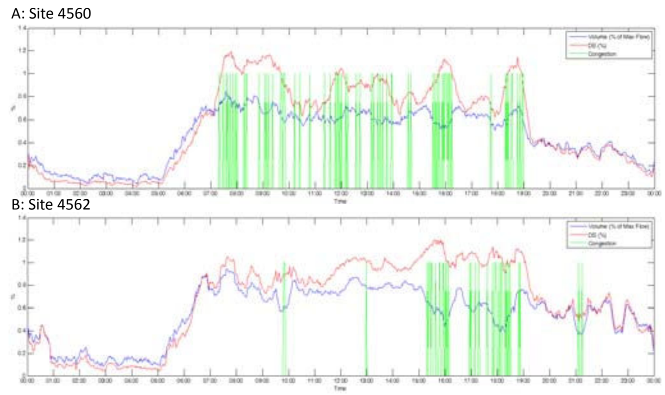


Fig. 4. Smoothed Time Series Plots of normalized parameters from through lane vehicle detectors on Flinders St, 12th July 2012.

Figure 5 shows the resulting graph from the samples of volume (i.e. VO/(phase time) [veh/h]) and DS over a single day for a single detector in one of the considered intersections. The graph reassembles the shape of a typical volume-density fundamental diagram where the volume is plotted against the DS instead of density.

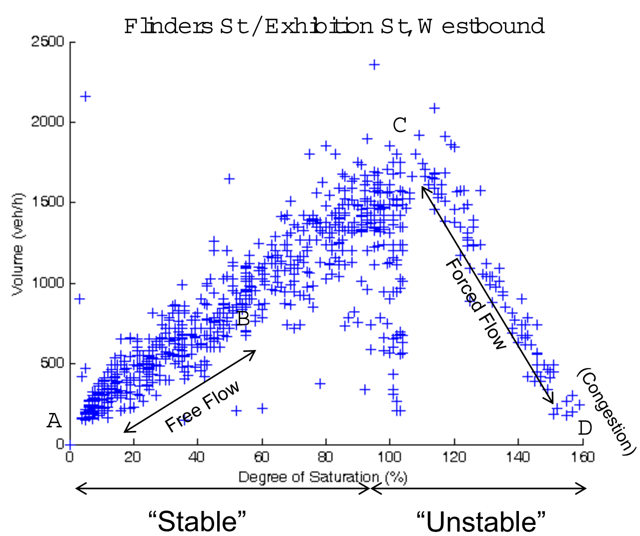


Fig. 5. Volume vs. DS Plot of Detector 7, Site 4560.

Note that the triangular shape of volume-density fundamental diagram is typical for a single lane/detector over a 24 hour period as described in II-A. The nature of the volume reaching maximum flow at the point where DS reaches 100% and decreasing thereafter suggests that the DS of greater than 100% being a reasonable indicator of congestion and over-saturation in traffic.

## B. Recurrence Plot Analysis

In the following the results obtained by recurrence plot (RP) analysis of traffic at both particular times of day and over the whole day is described. The SCATS data of 15 intersections along the Victoria Street and Victoria Parade in Melbourne is analyzed for both west (inbound) and east (outbound) traffic flows as shown in Fig. 2 (route A-B on the map). The detectors chosen at each intersection were, as far as possible, on straight-through (typically non-turning) lanes. Traffic volume is interpolated and re-sampled at 90 second intervals. All subsequent processing uses this re-sampled data. The DET measure is calculated from each day's time series in a set of adjacent, non-overlapping 1-hour (40 samples) windows.

The RP measures calculated from the data for each selected intersection and selected day of the week were averaged over one year (i.e. over 52 weeks). A phase-space trajectory was reconstructed from each 1-dimensional time series signal by using time-delay embedding with an embedding dimension of $m = 5$ based on the false nearest neighbour algorithm, and a fixed time delay of $\tau = 1$ to ensure no data points are skipped. Two methods of displaying the averaged data are used:

- Averaged traffic volume and DET measures are plotted on the same axes, the volume sampled at 90 second intervals and the DET sampled hourly.
- The DS vs. VO/VK plane is segmented into a 32 x 40 grid and the average DET in each cell over one year's data for each selected week day and intersection.

The calculations and plots in this report utilized the Matlab toolbox, `crptoolbox`, using the methods presented by Marwan [7] and available online at `http://tocsy.pik-potsdam.de/CRPtoolbox/`.

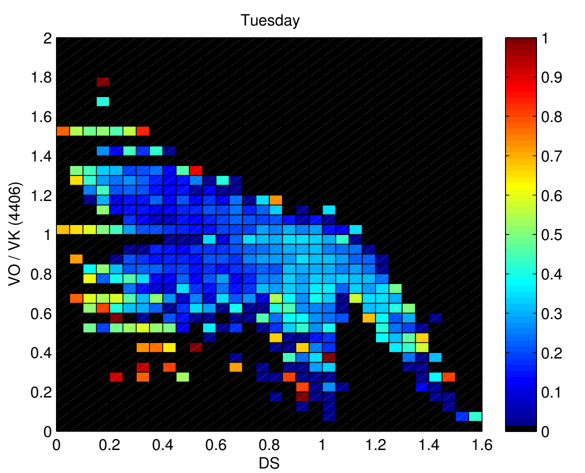


Fig. 7. Average DET values are plotted against the congestion indicator measure.

It can be seen in Fig. 6 that the DET value is high during more congested conditions suggesting that the traffic is less random in those periods. The same can be said in very

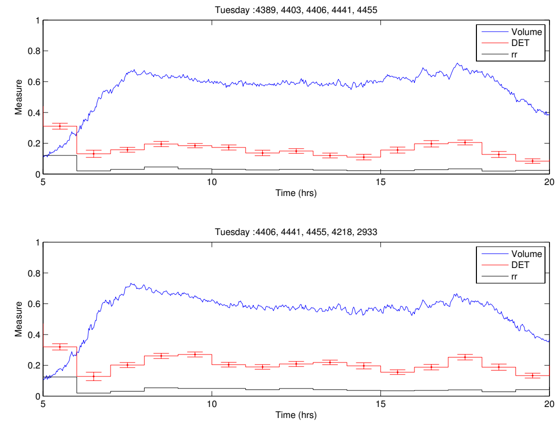


Fig. 8. CRQA measures, average DET values and recurrence rate ($rr$), plotted through the day for particular days of the week.

uncongested conditions where cars are moving in a free-flow pattern. Note that although the traffic volume are very similar, the DET value is higher for inbound traffic compared to that of outbound at the same intersection. This is likely due to the fact that the inbound traffic is terminated at the CBD and traffic is more predictable given the same condition compared to the outbound flow. It is also interesting to note that for inbound traffic, the Monday traffic exhibits a very different characteristic compared to that of the rest of the week (represented by the Tuesday traffic. In particular, Monday traffic seemed to be more random between the morning and afternoon peaks, whilst it is less random on the other days. The traffic at the weekend is again different with only one peak at about late morning as expected where there is a high level of determinism of traffic during the peak time.

In Fig. 7 there are similar observations where the average DET values are high in the region of low traffic (to the left of Fig. 7) or congestion (far right bottom of Fig. 7). It is interesting to note a triangular shape beyond the point where the DS is 100% which may represent the transient state (i.e. non-stationarity) of traffic at the beginning of the congestion forming process.

Finally, the CRQA measures (DET and $rr$) are plotted based on five intersections averaged over one year for inbound traffic in Fig. 8. The patterns observed in the time delay embedded data also appear in this spatial data (i.e. over a set of intersections), but require much larger neighbourhoods to get a reasonable level of recurrence. Here the DET and $rr$ measures are calculated using the individual volumes for the five intersections as the components of the phase space vector at each sampled instant, while the volume is an Euclidean norm of the 5-intersection volume vector divided by the square root of 5. Given that a pure noise signal with comparable setting for neighbourhood radius (as a fraction of mean separation) would be expected to have an $rr$ of about 0.002; most of the studied data have the $rr$ values at least 4 times as large, and frequently 10 times the

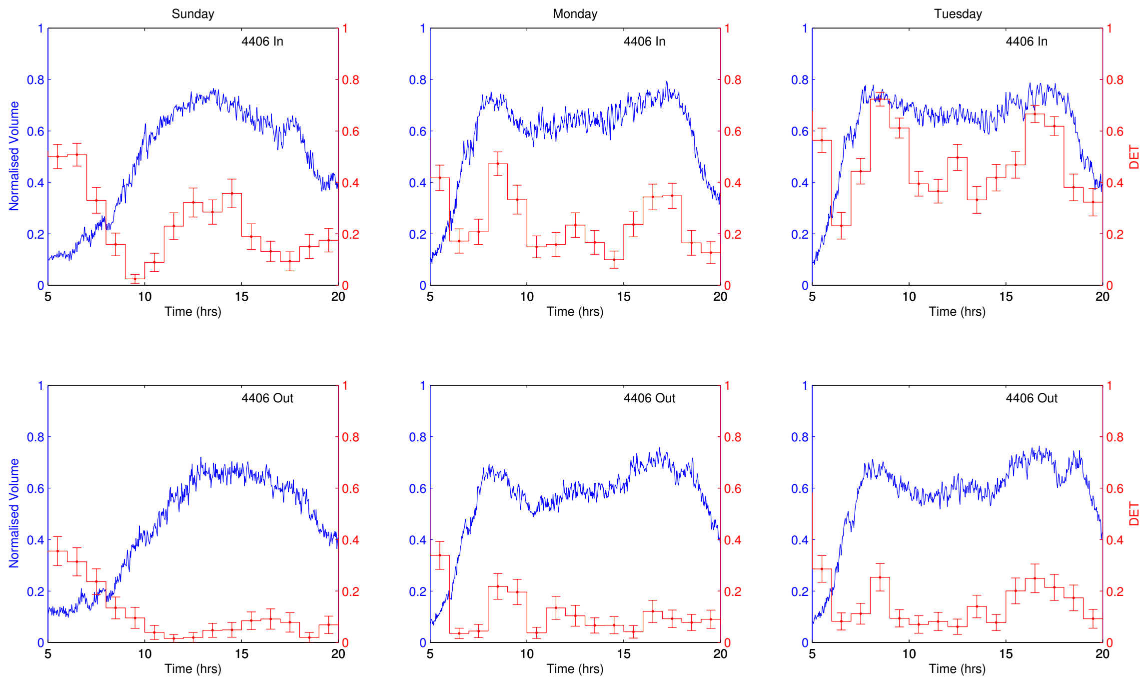


Fig. 6. Volume and Determinism measure (DET) of in-bound and out-bound traffic at an intersection (Site 4406), averaged over a year, are plotted through the day for particular days of the week.

noise level, which is a modest but encouraging margin as shown in Fig. 8.

## V. Conclusion

In this paper the short-term non-stationary features present in traffic flows based on the historical data available from stop-line detectors were investigated. It was shown that recurrence plot and recurrence quantification analysis can be used to identify non-stationary features of the traffic that otherwise could not be detected based on time series visualization. It was observed that periods of low traffic flow in the early morning are highly deterministic, as are the peak time periods, but the rest of the day are much more random. We also discovered that inbound traffic has a higher level of determinism as a result of the fact that most of the traffic is terminated in the city CBD.

## References

[1] D. Bennett, "Guide to Traffic Management Part 2: Traffic Theory," in *Guide to Traffic Management*, 1st ed. Austroads, 2008.
[2] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *EPL (Europhysics Letters)*, vol. 4, no. 9, p. 973, 1987.
[3] M.M. Hamed, H.R. Al-Masaeid, Z.M. Bani Said, "Short-term prediction of traffic volume in urban arterials.," ASCE Journal of Transportation Engineering 121 (3), 249, 1995.
[4] A. Hegyi, B. D. Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transportation Research Part C*, vol. 13, no. 3, pp. 185–209, 2005.
[5] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis.*, ser. Cambridge nonlinear science series. Cambridge University Press, 2003.
[6] B. S. Kerner and S. L. Klenov,"Deterministic microscopic three-phase traffic flow models," Journal of Physics A: Mathematical and General, vol. 39, no. 8, 2006.
[7] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237 – 329, 2007.
[8] N. Marwan and J. Kurths, "Nonlinear analysis of bivariate data with cross recurrence plots," *Physics Letters A*, vol. 302, no. 5, pp. 299–307, 2002.
[9] M. Masugi, "Recurrence plot-based approach to the analysis of ip-network traffic in terms of assessing nonstationary transitions over time," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 53, no. 10, pp. 2318 –2326, oct. 2006.
[10] M. Masugi, "Applying a recurrence plot scheme to analyze non-stationary transition patterns of ip-network traffic," *Communications in Nonlinear Science and Numerical Simulation*, vol. 14, no. 4, pp. 1418 – 1430, 2009.
[11] N. Nguyen and J. Gaffney, "Real-time traffic performance measures of adaptive traffic signal control systems," in *Research into Practice: 22nd ARRB Conference*, 2006.
[12] C. Suksri, M. Taylor, and W. Yue, "Exploring stop-line traffic flows at signalised intersections in the adelaide cbd," in *Australasian Transport Research Forum (ATRF), 34th, 2011, Adelaide, South Australia, Australia*, vol. 34, no. 0143, 2011.
[13] L. Trulla, A. Giuliani, J. Zbilut, and C. W. Jr., "Recurrence quantification analysis of the logistic equation with transients," *Physics Letters A*, vol. 223, no. 4, pp. 255 – 260, 1996.
[14] J.W.C. van Lint, S.P. Hoogendoorn, H.J.V. Zuylen, "Freeway travel time prediction with state-space neural networks," Transportation Research Record 1811, pp. 30–39, 2002.
[15] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Temporal evolution of short-term urban traffic flow: A nonlinear dynamics approach," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 7, pp. 536–548, 2008.
[16] E. I. Vlahogianni, "Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics," *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 73–84, 2009.
[17] J. P. Zbilut, A. Giuliani, and C. L. W. Jr., "Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification," *Physics Letters A*, vol. 246, no. 1-2, pp. 122 – 128, 1998.