

Systematic Identification of Peak Traffic Period

Sara Wibawaning Respati¹, Ashish Bhaskar², Zuduo Zheng³, Edward Chung²

¹PhD Student, School of Civil Engineering & Build Environment, Queensland University of Technology (QUT), Brisbane, QLD 4000, Australia

² Smart Transport Research Centre, School of Civil Engineering & Build Environment, Queensland University of Technology (QUT), Brisbane, QLD 4000, Australia

³ Principal Research Fellow (DECRA), School of Civil Engineering & Build Environment, Queensland University of Technology (QUT), Brisbane, QLD 4000, Australia

Email for correspondence: sara.respati@hdr.qut.edu.au

Abstract

To identify peak periods in systematic manner, this paper proposes and tests a framework to pinpoint the periods by mining travel time time-series data. In differentiating between peak and off-peak periods in the data, Bottom-Up algorithm, a method of segmenting time series, is presented. Each day's travel time time-series is segmented and the time points of the start and the end of the peak period is identified. These points (start and end of the peak) are obtained for different days of interest and the respective distribution is defined to estimate a statistically significant time for start and end of the peak for the site. The applicability of the proposed framework is tested using real Bluetooth data from Brisbane. The case study analysis indicates that the peak periods can be systematically estimated using the proposed framework and the results are better than the existing threshold based approach. The paper also presents practical applications that can be improved by implementing the proposed method.

Keywords: peak period, time series segmentation, Bottom-Up, peak period application

1. Introduction

Road traffic is a dynamic system where its state change over space and time. Understanding the elements is critical for traffic management system, especially a traffic state when traffic congestion is high (or when speed is low), which is referred to peak period or peak hours. Failure to manage the traffic during peak demand may lead to severe traffic conditions (as gridlocks and significant delays) and incidents (crashes). Peak period is widely used by both practitioners and researchers because of its fundamental role in both network planning and operations. For instance, the knowledge of the peak period is needed for: estimating congestion minutes and signal/corridor prioritization; time of the day-based arterial traffic signal plans updates; congestion cost modelling; identification of congestion charging duration; public transport frequency planning and parking management.

There is no systematic way to define peak period and generally, it is defined considering practitioners' judgement. Practitioners has also incorporated travel survey (Corpuz 2006) to define peak period. Surveys are expensive and time consuming. The frequency of the survey is generally not enough to capture the changes of traffic condition over different days and over different sites. Implementing threshold value has also been applied by practitioners to define the period, i.e. demand or speed threshold, nonetheless, different sites may have different threshold because the nature of traffic for different sites varies.

The advancement of technology provides the availability of big traffic data such as, Bluetooth, Wifi (Abbott-Jard, Shah and Bhaskar 2013), GPS and third-party data (Here, Intelmatics etc).

The accessibility of these traffic data gives opportunity to mine real data, as such, peak period can be identified accordingly. This study proposed a framework to determine the peak period in systematic manner, and the framework is applied on the real Bluetooth data from Brisbane. In expectation of the study contributes to the practitioners' planning and operation, review of some applications that can be improved by using the proposed framework such as congestion cost calculation, signal plan optimisation and congestion measurement is discussed.

2. Background

2.1. Bluetooth data collection

The advancement of technology has led to the utilization of Bluetooth Media Access Control Scanner (BMS) as one of the traffic data source. These days, most of electronic devices, such as mobile phones and car navigation system, are enhanced by Bluetooth technology. The Bluetooth devices have their own unique MAC (Media Access Control) addresses. BMS scanners detect the MAC ID within its detection zone, which is around 100 m for most traffic applications (Bhaskar and Chung 2013). The time difference of a MAC ID detected by upstream and downstream BMS scanner yields the travel time for the MAC ID. Data collected by BMS scanners is stored in table format with several fields:

- *ID*: MAC ID of devices detected by BMS Scanner
- *Timestamp*: time when scanners first detect the Mac ID
- *Duration*: duration of device presence in the detection zone.

Interested reader should refer to Bhaskar and Chung (2013) for the fundamental understanding on the use of BMS as a complementary transport data source. BMS can provide a seamless time-series of travel time between 2 BMS stations.

2.2. Time series segmentation

Piecewise Linear Representation (PLR), approximates a time series with a number of straight lines splitting the series. The algorithms based on PLR includes: *Sliding Window*, *Top-Down* and *Bottom-Up*. Here, these algorithms are summarized. Interested readers should refer to Keogh et al. (2004) for detailed explanation.

Top-Down algorithm starts with a single segment (unsegmented sequence), and it creates a new segment at the next step in such way that minimize the overall error. The boundary is introduced recursively, that is, in i -th time step, the algorithm creates i -th segment boundary by dividing the certain current segment in to 2.

Sliding Window initiates the segmentation process by anchoring the left point to be a boundary, then scanning the data to the right. In that manner, a longer potential segment is considered, and at some point i , the error is greater than the predefined threshold, the subsequence points from the anchor to $i-1$ is converted to one segment. Then the point i becomes the new anchor and the same process is repeated.

Bottom-Up algorithm creates finest segment approximation at the initial step. That is, a pair of adjoin points is considered as one segment generating $n/2$ segments of time series, where n is number of data points. The algorithm continues merging the segments with the lowest error/cost until the termination criteria is met.

The termination criteria for *Bottom-Up* and *Top-Down* algorithms can be *error*, *maximum error* and *number of segments* whereas, for *Sliding Window* algorithm it is *error*. The complexity of *Top-Down*, *Sliding Window* and *Bottom-Up* algorithm is $O(n^2K)$, $O(Ln)$ and $O(Ln)$, respectively. Where K , L and n are number of segments, average segment length and number of data points, respectively.

3. Study Site and Data

This study focuses on an urban signalised arterial corridor- Old Cleveland Road (OCL), Brisbane. OCL is one of the critical corridors connecting South East suburbs to Brisbane CBD (refer to **Figure 1**). Peak period was defined for a link based on data collected by Bluetooth Scanners BT 10750 (upstream scanner) and BT 10680 (downstream scanner). The length of the link is 1.82 km. The analysis included working days travel time data containing 11 months data (June 2015-May 2016). Working days are obtained by excluding the data of days that are corresponding to weekends, public holidays, and school holidays from the Bluetooth database. In this paper, only morning peak periods for the inbound traffic are analysed. The analysis can be repeated for the other peak periods (evening and school pick-up times).

Figure 1: Study Site

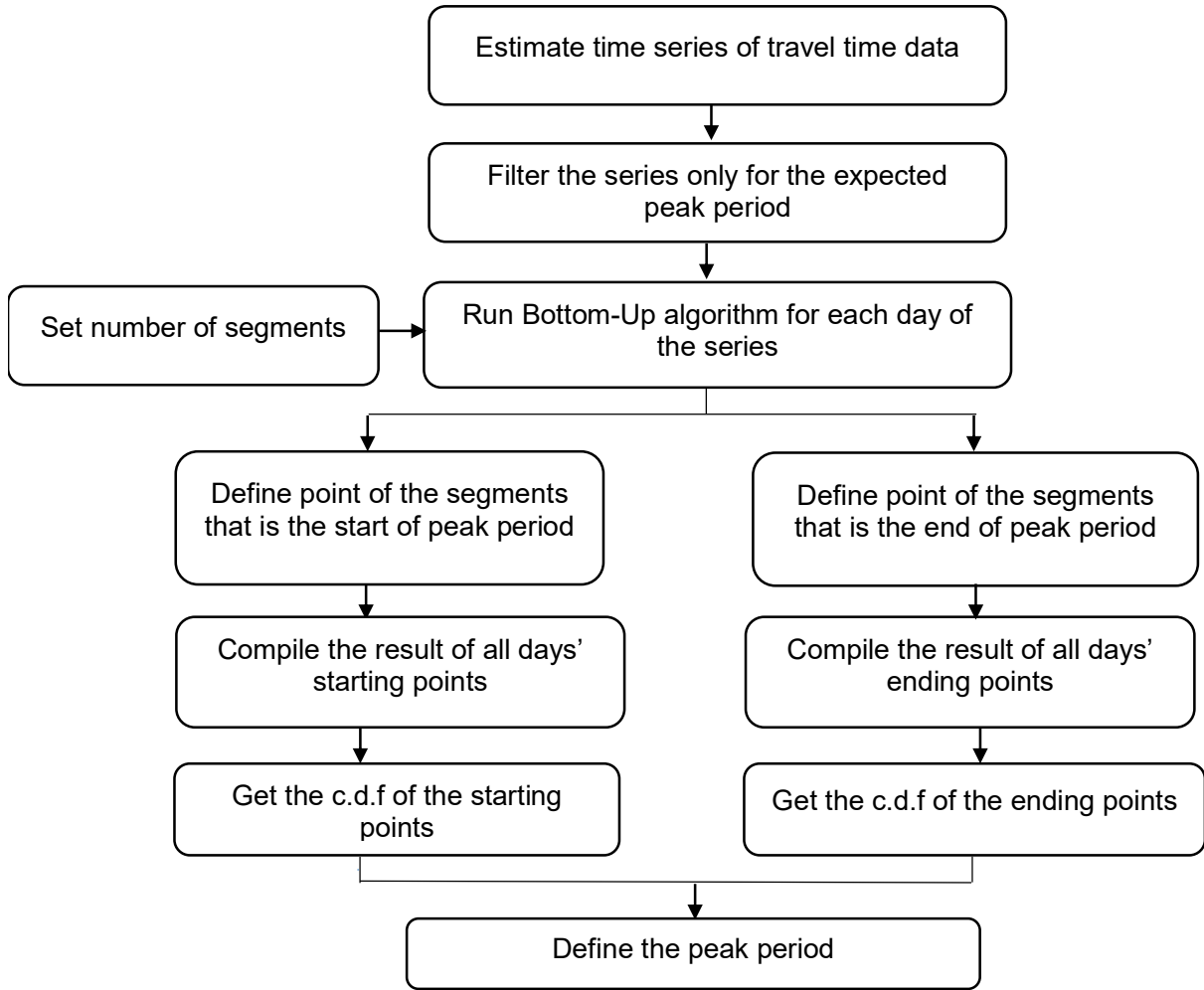


4. Proposed Framework

Here, we present the proposed framework that consists of the following steps (refer to **Figure 2**.)

- Step-1 : Time series of travel time estimation
- Step-2 : Filter time series for the expected duration
- Step-3 : Apply time series segmentation on each day
- Step-4 : Develop database for start and end of peak periods for different days
- Step-5 : Evaluate the distribution

Figure 2: Framework to identify peak period.



where c.d.f is cumulative distribution function

Step-1: Time series estimation

The time series estimation depends on the type of the retrieved traffic data. This study utilizes travel time data from Bluetooth MAC scanners (BMS) and here we provide the necessary details of the travel time estimation from BMS. However, to apply the framework, the data is not limited to BMS.

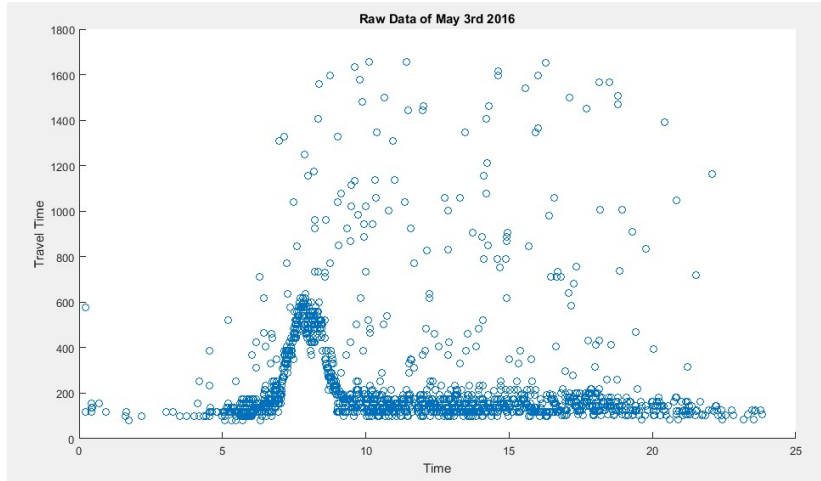
BMS provides the time when a vehicle is observed at the scanner location. Travel time of a vehicle (v) from an upstream scanner ($BMS_{u/s}$) to a downstream scanner ($BMS_{d/s}$) is the time difference between the times when the vehicle is observed at the scanner locations. Here, we consider the exit-to-exit travel time (tt_v) which is the time gap when the vehicle is last observed at $BMS_{d/s}$ ($t_{v,d/s}$) and $BMS_{u/s}$ ($t_{v,u/s}$)

$$tt_v = t_{v,d/s} - t_{v,u/s}$$

The above matching provides raw travel time data points, which can include significant noise due to reasons such as vehicle has taken different path to travel between the locations; vehicle was observed at upstream but it was missed to be captured at downstream during its travel and later that day it was observed again at downstream etc.

Figure 3 illustrates the data points from individual vehicles travel time from Old Cleveland Road, Brisbane for 3rd May 2016. The data noise is easily visible.

Figure 3: Example of raw travel time data



The raw travel time data is filtered to reduce the anomalies. We apply Median Absolute Deviation (MAD) (Gather and Fried 2004) filter. This method removes outliers by comparing a data point with its neighbouring observations within 10 minutes intervals. A window of 5 minutes before and 5 minutes after is considered for each minute. Points are identified as outliers if they are outside threshold value as defined below.

$$UBV = median + \sigma f$$

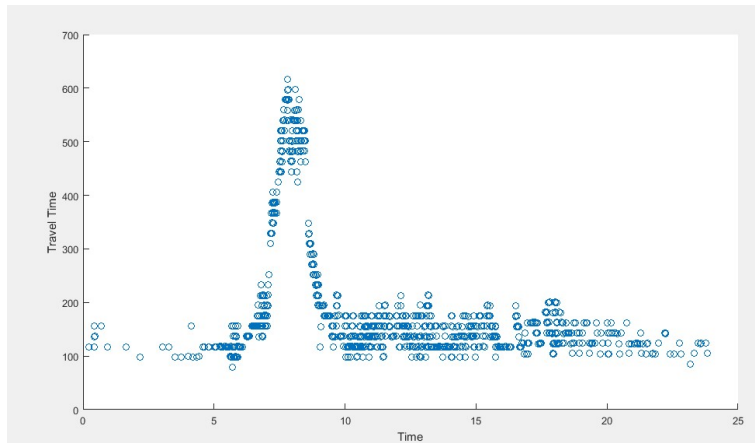
$$LBV = median - \sigma f$$

where UBV is the Upper Bound Value, LBV is the Lower Bound Value, and σ is the standard deviation of MAD that can be approximated as $\sigma = 1.4826 \times MAD$ for a normally distributed data. MAD is defined as:

$$MAD = median(|X_i - median(X)|)$$

The value σf represents the scatter of the sample, whereas the parameter f is a scale factor that defines the gap between UBV and LBV . The value of $f = 2$ is chosen for this study as suggested by Kieu, Bhaskar and Chung (2015). Note: Brisbane City Council and Qld Department of Transport and Mains Road also use $f=2$ for filtering the BMS data. **Figure 4** illustrated the filtered data, where the MAD filter is applied on the data presented on **Figure 3**.

Figure 4: Example of filtered data



The individual vehicle time series is aggregated into 5 minutes aggregated time series. Here, an average travel time for every 5 minutes interval is estimated. Reconstructing the data into

5 minutes interval, some empty intervals are found due to zero sample sizes, therefore linear interpolation was used to fill the gap of the time series.

Bluetooth sample sizes vary from time to time, there are intervals that contain zero or small samples. For treating interval with small sample size, data smoothing is applied. Local regression with weighted method is implemented to smooth the data. MATLAB Toolbox of 'rlowess' was used to apply the weighted local regression method, which employs weighted linear least squares and a 1st degree polynomial model. The regression allocates lower weight to outliers in the regression and zero weight to data outside six mean absolute deviations.

Step-2: Filter time series for expected duration

The last step provides time series for the entire day. By observing data visually over different days, peak period can be firstly assumed within some range of time. The range should be long enough, for example 3 to 4 hours, to prevent the true peak period from being excluded from the range. For instance, period from 6:00 am to 10:00 am can be assumed as the initial start for the morning peak period.

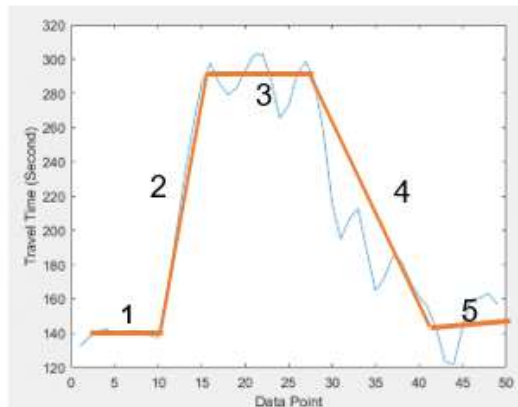
For time series segmentation, the data for the study peak period is filtered and the remaining steps are applied.

Step-3: Apply time series segmentation on each day

As discussed on previous chapter, the algorithm's stopping criteria can be specified based on one of these properties: number of segments, maximum error for each segment or total error of all segments. This study specifies *number of segments* as a target of the function, and peak periods can be specified accordingly.

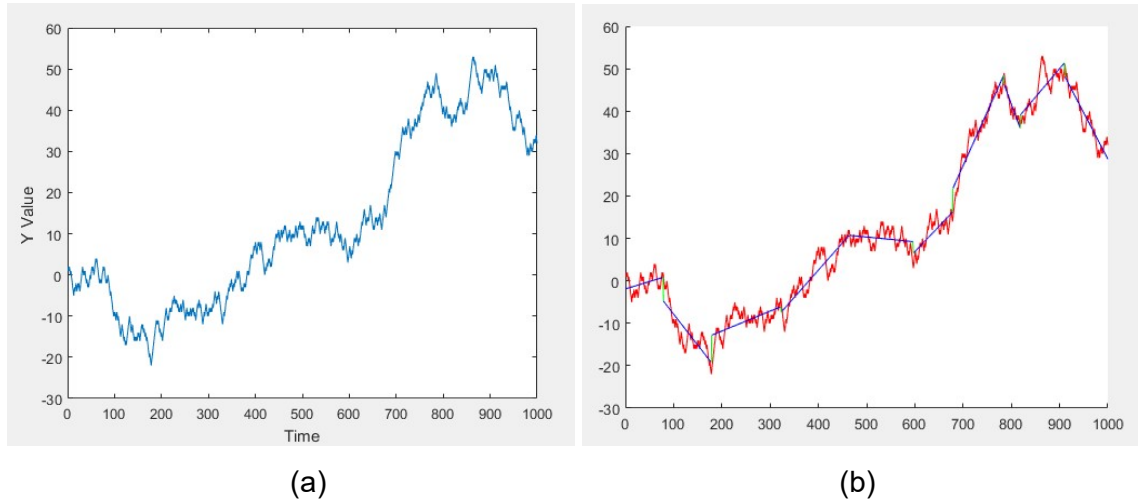
Number of segments can be specified based on travel time profile. Traffic regimes for travel time profile is expected to include free flow, congestion (queue) build up, congested (queue), congestion (queue) dissipation, and return to free flow condition. Number of segments as 4 or 5 are viable options. Refer to **Figure 5**, where 5 segments are defined for the time series, which seems to be a better representation. Nevertheless, both 4 and 5 as number of segments are tested and 5 provides better results. Details are provided in Section 5.1.

Figure 5: Travel time profile



Considering 5 as number of segments, the peak period is defined from the start (left point) of segment-2 to the end (right point) of segment-4. Refer to **Figure 5** for an example.

This study adopts *Bottom-Up* algorithm to split the time series, because it starts from the finest approximation to get the small targeted segment. In that manner, many iterations would have been executed before it is terminated. The pseudocode of Bottom-Up algorithm and the examples of the segmentation are presented on **Appendix A** and **Figure 6**, respectively.

Figure 6: (a) Sample time series (b) segmentation result


A random walk time series data, as illustrated on **Figure 6 (a)**, is split into ten sections using Bottom-Up algorithm. **Figure 6 (b)** shows the segmentation output. The blue line and the green line represent the piecewise linear representation and the boundary of each segment respectively. Each segment has property of its left point (lx) (start point) and right point (rx) (end point) with regards to x -coordinate, as presented on **Table 1**.

Table 1. Segments' properties

Segment	lx	rx
1	1	78
2	79	178
3	179	322
4	323	464
5	465	596
6	597	678
7	679	784
8	785	818
9	819	910
10	911	1000

Step-4: Develop database for start and end of peak period

The previous task generates segments' properties indicating left and right points of each segment. Expected travel time profile, as shown on **Figure 5**, signifies that left point of segment 2 corresponds to the start point of peak period, whereas right point of segment 4 corresponds to the end point of peak period. After segmenting process is done for the entire days, database for both starting and ending point of peak period are established.

Step-5: Peak period identification

Here, we show how peak period can be defined from the database gathered from previous step. To identify peak period, it is important to pick a reasonable point of the starting and ending time. To this end, cumulative distribution functions ($F(x)$) of both the start and the end of peak period are independently constructed.

$$F(x) = P_r(X \leq x)$$

Where X is the start (or end) of peak period. $F(x)$ denotes the probability of that the value of X will be less or equal to x . The function of $F(x)$ is monotonic, $F(-\infty) = 0$ and $F(\infty) = 1$.

The next section provides details of the application of $F(x)$.

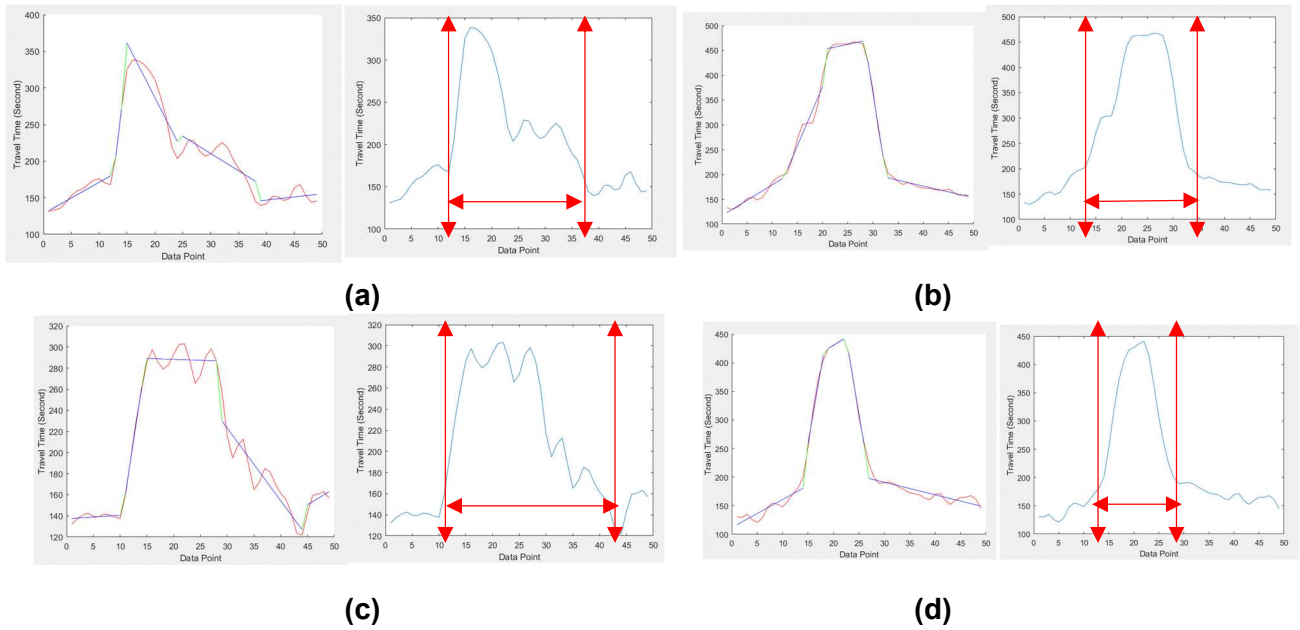
5. Result and Discussion

5.1. Peak Period Identification

Number of segments as 5:

Considering 5 as the number of segments, **Figure 7** illustrates 4 cases of segmentation for the expected morning peak period, which is 6:00 am to 10:00 am. The x-coordinate that includes 49 data points represents 5 minutes interval of duration from 6:00 am to 10:00 am. The pictures of the left side are the segmentation algorithm output, with red line, green lines and blue line represent travel time profile, boundary of segments, and piecewise linear estimation respectively. Meanwhile, the right picture presents the travel time profile with the area between straight red lines indicates the peak period. The peak period is constructed according to the left point of segment 2 and right point of segment 4.

Figure 7: Segmentation results with 5 segments



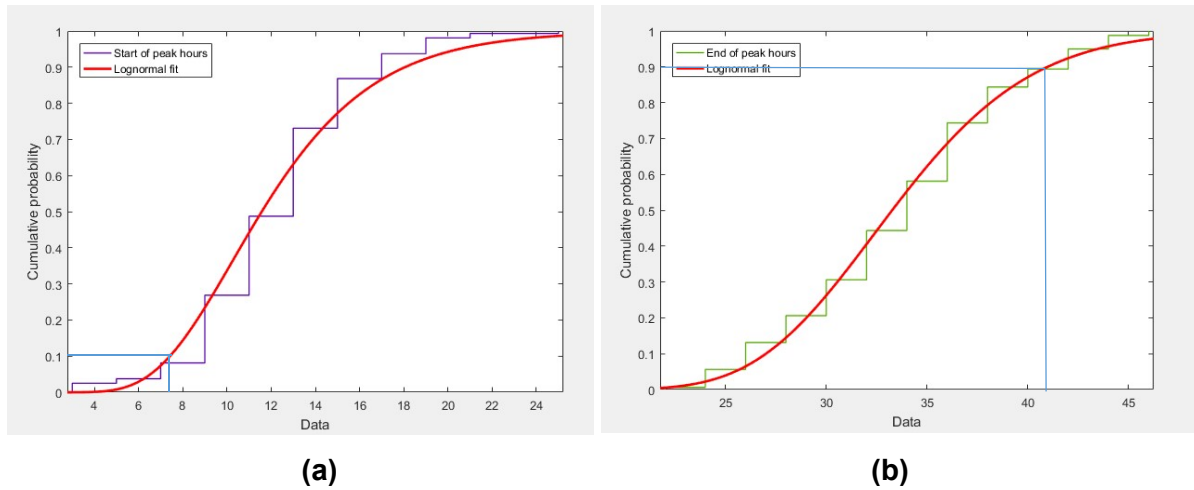
Clearly shown on the figure, the segmentation with 5 segments can distinct off-peak and peak periods for all the cases. To show an example of peak period identification, **Table 2** presents a segmentation result of travel time shown on **Figure 7 (c)**.

Table 2: Segments' properties

Segment	Start point	End point
1	1	10
2	11	14
3	15	28
4	29	44
5	45	49

Accordingly, data point number 11, equal to 6:50 am, is set to be starting point, and data point number 44, equal to 9:35 am, is specified to be ending point of peak period. Therefore, peak period for the particular example is from 6:50 am to 9:35 am. After implementing segmentation for the entire data that includes 160 working days, the $F(x)$ of both the start and the end of peak period can be derived, independently.

Figure 8: Cumulative Distribution Function (F_x) of (a) start and (b) end of peak period

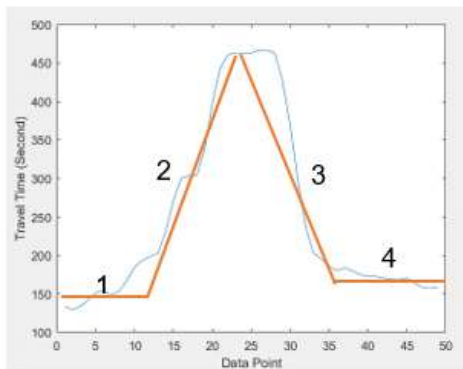


The $F(x)$ of the start and the end of peak period are presented in **Figure 8 (a)** and **(b)**, respectively. It is observed that it best fits the lognormal distribution. From the distribution, practitioners can easily define the peak period boundaries. For instance, if they are interested to define the peak period where 90% of the times the boundaries are correctly identified, then $F(x)$ as 0.1 for start and $F(x)$ as 0.9 for the end should be used. Considering these values for the case study site, the corresponding data points for start is 6:30 am (7th data point) and end is 9:20 am (41th data point). Therefore, the peak period for the site starts from 6:30 am and ends at 9:20 am, with total duration of 170 minutes.

Number of segments as 4:

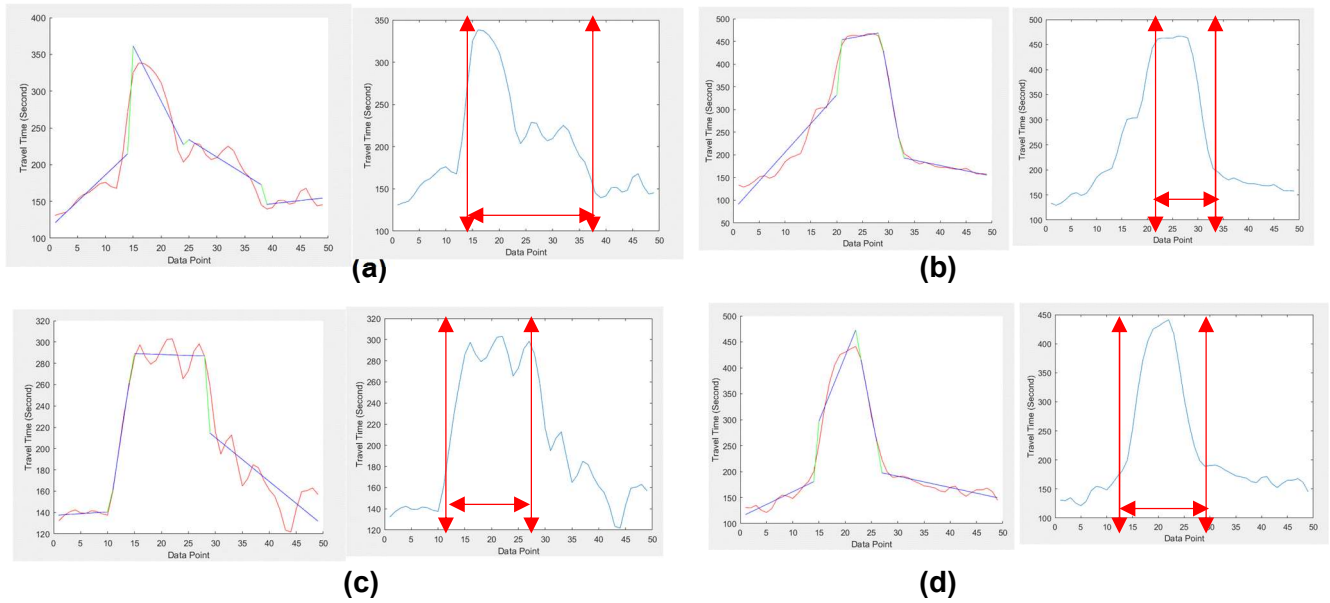
One can also expect a travel time profile as shown on **Figure 9** consisting of 4 segments i.e. free flow, queue build up, queue dissipation and free flow. In this case, peak period is defined from the left point of segment 2 to the right point of segment 4.

Figure 9: Expected travel time profile with 4 segments



To illustrate the segmentation result using only 4 segments, **Figure 10** is presented.

Figure 10: Segmentation results with 4 segments



The analysis with 4 segments is not able to identify peak period properly for the profiles that do not have single apparent peak, which is illustrated on **Figure 10 (a),(b) and (c)**. For example, the start of peak period in **Figure 10 (b)** is at the top of the travel time profile, which is not true, as the algorithm fails to consider the queue regime on the top of the profile. According to the results, consideration of 4 segments does not provide the result as expected.

5.2. Threshold Based Approach

Here, we apply the example of the existing peak period identification that use threshold values, such as demand or speed threshold. WSDOT's defines peak period as the period when the average speed is lower that 75% of posted speeds (Washington State Department of Transportation 2011) . This threshold was then applied into dataset considering the speed limit of 60 km/h is the free flow speed. Defining 75% of 60 km/h as a congested speed, the congested travel time of 146 seconds is acquired. The travel time is applied as the threshold to define the peak period, and 6 cases example of the results is presented on **Figure 11**.

Figure 11: Peak period identification using travel time threshold

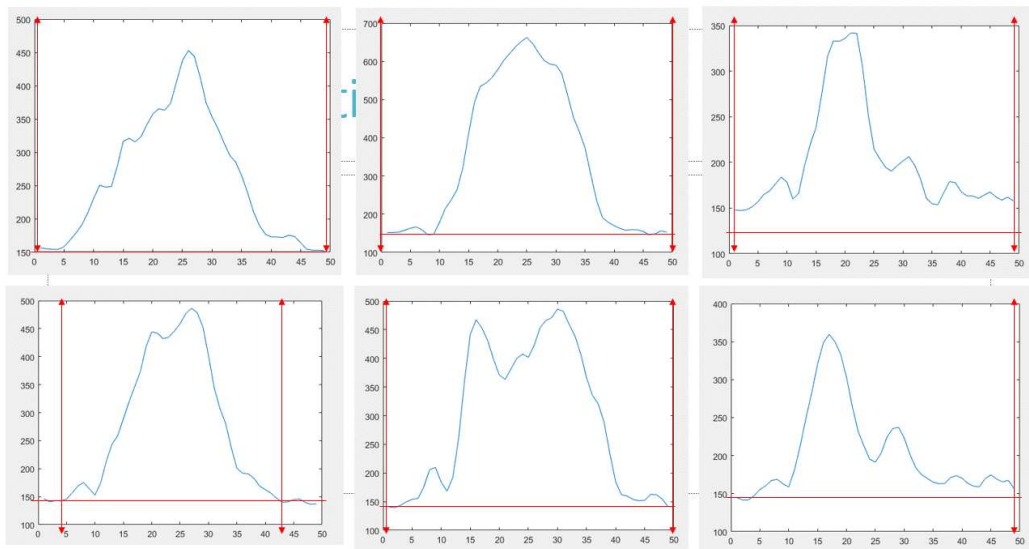


Figure 11 shows the peak duration within vertical red lines, whereas the horizontal red line is the travel time threshold. The method is not able to distinguish between peak and off-peak period as it is clearly shown on the figure that it considers off-peak period as peak period. Based on the result, consideration of speed threshold does not provide expected result as it overestimates peak period.

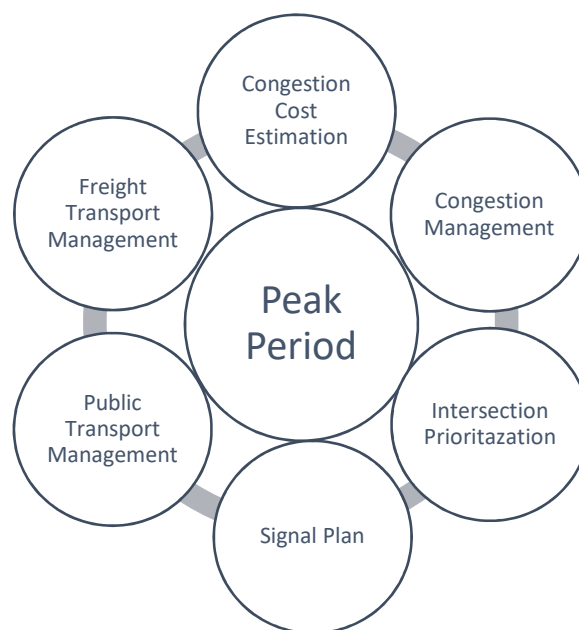
5.3. Application of Peak Traffic Period

Here, we discuss some applications in both traffic planning and operation that consider peak period to give example of the applicability of the presented framework. Peak period including morning and afternoon peak period has been utilized to calculate performances of key corridors by Brisbane City Council (BCC). The morning peak period for Transport Main Roads' corridor is from 6:00 am to 9:00 am, meanwhile the morning peak period for Brisbane City Council's corridor is from 7:00 am to 9:00 am (Brisbane City Council 2016a). BCC used the aforementioned peak period to calculate vehicle kilometres travelled and corridor average speed for key corridors presented in the performance report (Brisbane City Council 2016a). The Old Cleveland Road belongs to BCC's corridor, therefore the morning peak period is between 7:00 am and 9:00 am or 120 minutes, which is 50 minutes less than the peak period estimated in the previous section.

As published by the BCC, the Old Cleveland Road with 5.87 km in length has average speed of 31.53 km/hr in 2015 and 29.26 km/hr in 2016 for the morning peak (Brisbane City Council 2016b). To compare the said average speed with that during the estimated peak period, speed during the estimated period need to be derived by dividing the section length by the travel time. Then, the average speed of 31.4 km/h for the study site during peak period of 6:30 am-9:20 am is obtained. The proposed framework produces similar average speed with the data published by BCC indicating that the framework generates rationale result.

The peak period of the proposed framework is systematically defined from the data and is site dependent. It can further help practitioners in the applications of peak period, which the examples are illustrated on **Figure 12**.

Figure 12: Peak period application



The applications of peak period, as presented on **Figure 12**, are as follow:

1. Estimate the cost of congestion

The growth of traffic lead to elevated levels of congestion. Estimating the cost of congestion is important because it gives the information of the consequent impact of the urban congestion. There are some indicators of congestion used to calculate the cost (Litman 2003). The indicators that utilize peak period are:

- a) Travel time rate: ratio of peak period to free-flow travel time in recurrent condition,
- b) Travel time index: ratio of peak period to free-flow travel time in both recurrent and non-recurrent condition,
- c) Percent travel time in congestion: portion of vehicle/person during peak period condition,
- d) Congested Road Miles: portion of roadway miles congested during peak period
- e) Congested Time: the length of peak period
- f) Congested Lane Miles: the number of congested lane miles during peak period,
- g) Annual Delay Per Road User: extra travel time hours divided by peak period users,
- h) Average traffic speed: average travel speed during peak period.

The peak period that is defined systematically using the field data improves the accuracy of the estimated congestion cost.

2. Intersection prioritisation tool

The prioritisation of intersections requires ranking of intersections considering the congestion minutes per hour. The congestion minutes per hour differs from intersection to intersection. The proposed framework can help to decide the prioritisation, because it identifies the start and the end of peak period for different link or intersection.

3. Time of the day signal plan

Traffic signals that uses time of the day based signal plans are set considering peak and off-peak periods. Here, the signals timings are optimised offline and then fixed for different traffic condition. By having a properly defined peak period using the proposed framework, the controller can have a proper set up. Therefore, the proposed framework can be applied to improve signals at traffic junctions with regards to increase travel speed/travel time, increase veh/hr/lane capacity and reduce traffic crashes.

4. Congestion management

One of the crucial task in congestion management is to measure the congestion itself. Congestion, in term of delay, is measured for specific time i.e. peak period. If the peak period is not well defined, the congestion measurement will be underestimate or overestimate.

5. Public transport improvement

Peak period identification can help the identification of public transport frequency because it gives representation of the demand of public transport. This can be used to double check the analysis of the demand based on the number of passengers. Public transport fare determination can also take benefit of the peak period identification by which peak and off-peak period fare are regulated.

6. Freight transport management

Generally, freight vehicles have restricted access to urban arterials, this includes off-peak period access only. The shifting of urban freight distribution to off-peak period increases the utilization of road infrastructure and improve the goods distribution efficiency by avoiding extra cost due to congestion. This also avoid freight transports present in urban network in peak period in which recurrent congestion occurs. Improper identification of peak and off-peak periods would yield higher congestion

level because freight transports enter the busy urban network; the congestion slows good movement causing cost increment. Therefore, peak period needs to be defined appropriately to avoid this condition.

The proposed framework can identify the peak period more accurately compared to threshold-based approach. Over estimating peak period, such the threshold based approach generally does, may lead to several problems: e.g., setting a traffic signal cycle to a peak period within an off-peak period lead to ineffective traffic signal control. Therefore, the above-stated application can benefit from the proposed framework to get the efficient applications.

6. Conclusion

This paper presented a procedure to estimate peak period in systematic manner. The framework includes travel time time-series mining using Bottom-Up algorithm as a segmentation method. The function's stopping criteria is number of segments that is required to be set to be 5. Accordingly, peak period can be defined from the left point of segment 2 to the right point of segment 4. Using real Bluetooth data from Brisbane network, this paper demonstrates the capability of the suggested method to identify the peak period. The results confirmed that the proposed framework is able to differentiate peak period from off-peak period.

7. Acknowledgement

The authors wish to thank Brisbane City Council for providing valuable data for this research.

8. References

- Abbott-Jard, Michael, Harpal Shah and Ashish Bhaskar. 2013. "Empirical evaluation of Bluetooth and Wifi scanning for road transport." In *Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia*, edited, 14.
- Bhaskar, Ashish and Edward Chung. 2013. "Fundamental understanding on the use of Bluetooth scanner as a complementary transport data." *Transportation Research Part C: Emerging Technologies* 37: 42-72.
<http://www.sciencedirect.com/science/article/pii/S0968090X13002027>. doi:
<http://dx.doi.org/10.1016/j.trc.2013.09.013>.
- Brisbane City Council. 2016a. *Greater Brisbane Key Corridor Performance Report January-June 2016*.
- Brisbane City Council. 2016b. "Greater Brisbane key corridors performance July-December 2016." Accessed 20/July. <https://www.brisbane.qld.gov.au/traffic-transport/traffic-management/greater-brisbane-key-corridors-performance-report/greater-brisbane-key-corridors-performance-july-december-2016>.
- Corpuz, Grace. 2006. "Analysis of peak hour travel using the Sydney household travel survey data." In *Australasian Transport Research Forum (ATRF), 29TH*, edited.
- Gather, Ursula and Roland Fried. 2004. "Methods and algorithms for robust filtering." In *COMPSTAT 2004—Proceedings in Computational Statistics*, edited, 159-170: Springer.
- Keogh, Eamonn, Selina Chu, David Hart and Michael Pazzani. 2004. "Segmenting time series: A survey and novel approach." *Data mining in time series databases* 57: 1-22.
- Kieu, Le Minh, Ashish Bhaskar and Edward Chung. 2015. "Empirical modelling of the relationship between bus and car speeds on signalised urban networks." *Transportation Planning and Technology* 38 (4): 465-482. doi: 10.1080/03081060.2015.1026104.
- Litman, Todd. 2003. *Transportation cost and benefit analysis: Techniques, estimates and implications*. <http://www.vtpi.org/tca/>.
- Washington State Department of Transportation. 2011. *WSDOT's Congestion Measurement Approach: Evaluating System Performance*.
https://www.wsdot.wa.gov/NR/rdonlyres/821BF63A-BB4C-49C5-AEDE-3F27CC671813/0/CongestionFolioForWeb_June2011.pdf.

Appendix A . Bottom-Up Algorithm's Pseudocode (adopted from Keogh et al. (2004))

```

function residuals=b_up(data,num_segments,forceplot)
assign 1:2: size(data)-1 to left_x
assign left_x+1= right_x
assign size(data) to right_x(end)
number_of_segments= length(left_x) %create finest approximation of segments
for i=1: number_of_segments
    assign left_x(i) to segment(i).lx
    assign right_x(i) to segment(i).rx
    assign inf to segment cost
end
for i=1:number_of_segments-1
    calculate the coefficient of polynomial curve fitting for segment(i).lx :segment(i).rx
    get the y-value of the approximation
    calculate cost of each segment;
end
while number of segment < num_segment
    get the value and position (i) of minimum cost
    if i > 1 & i < length(segment) -1
        calculate the coefficient of polynomial curve fitting for segment(i).lx :segment(i+2).rx
        get the y-value of the approximation
        assign segment(i).rx = segment(i+1).rx;
        segment(i+1) = []; %delete segment (i+1)
        i = i - 1;
        calculate the coefficient of polynomial curve fitting for segment(i).lx :segment(i+1).rx
        get the y-value of the approximation
        calculate cost of each segment;
    elseif i==1
        calculate the coefficient of polynomial curve fitting for segment(i).lx :segment(i+2).rx
        get the y-value of the approximation
        calculate cost of each segment;
        assign segment(i).rx = segment(i+1).rx;
        segment(i+1) = [];
    else
        assign segment(i).rx = segment(i+1).rx;
        segment(i).mc = inf;
        segment(i+1) = [];
        i = i - 1;
        calculate the coefficient of polynomial curve fitting for segment(i).lx :segment(i+2).rx
        get the y-value of the approximation
        calculate cost of each segment;
    end;
end;
residuals= [];
for i = 1 : length(segment)
    calculate the coefficient of polynomial curve fitting for segment(i).lx :segment(i).rx
    y-value = (coef(1)*( [segment(i).lx :segment(i).rx]' ))+coef(2);
    residuals = [ residuals ; sum((data([segment(i).lx :segment(i).rx])-(y-value)).^2)];
end;
if nargin > 2
    hold on
    plot data
    plot linear approximation for each segment
    save y-value of each segment(i).rx and segment(i).lx
    for i = 1 : length(segment) - 1
        plot line that connect end of segment (i) to the start of segment (i+1)
    end;
end;

```