This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

On selecting an optimal wavelet for detecting singularities in traffic and vehicular data

Zuduo Zheng[1*] and Simon Washington[1]

[1]Queensland University of Technology, Brisbane QLD 4001 Australia

Abstract

Serving as a powerful tool for extracting localized variations in non-stationary signals, applications of wavelet transforms (WTs) in traffic engineering have been introduced; however, lacking in some important theoretical fundamentals. In particular, there is little guidance provided on selecting an appropriate WT across potential transport applications. This research described in this paper contributes uniquely to the literature by first describing a numerical experiment to demonstrate the shortcomings of commonly-used data processing techniques in traffic engineering (i.e., averaging, moving averaging, second-order difference, oblique cumulative curve, and short-time Fourier transform). It then mathematically describes WT's ability to detect singularities in traffic data. Next, selecting a suitable WT for a particular research topic in traffic engineering is discussed in detail by objectively and quantitatively comparing candidate wavelets' performances using a numerical experiment. Finally, based on several case studies using both loop detector data and vehicle trajectories, it is shown that selecting a suitable wavelet largely depends on the specific research topic, and that the Mexican hat wavelet generally gives a satisfactory performance in detecting singularities in traffic and vehicular data.

Keywords: Wavelet transform; Singularity detection; Traffic data analysis; the Mexican hat wavelet; Short-time Fourier transform; Oblique cumulative curve

1. Introduction

Irregular structures and transient phenomena (singularities hereafter) of a signal (curve, graph) often contain considerably rich information regarding the phenomenon being studied, as examples edge detection in image processing (Mallat and Zhong, 1992), vibration analysis in machine health monitoring (Peng and Chu, 2004), and irregular structure detection in physics (Argoul et al., 1988), etc. Considering applications in transport, singularities in traffic data may indicate activation or deactivation of a bottleneck, state changes of traffic flow, and occurrences of abnormal events; these in vehicular data may indicate the start of a driver accelerating or decelerating, or changing lanes. Thus, to correctly understand traffic flow characteristics and individual driving behavior, singularities in traffic data and/or vehicular data need to be accurately and reliably detected. Despite these rather straightforward examples, however, random and systematic noise contained in transport related data often makes detection of singularities extremely challenging.

[*] Corresponding author. 2 George St Brisbane QLD 4001 Australia. Tel.: +61 7 3138 9989; Fax: +61 7 3138 1170; Email address: zuduo.zheng@qut.edu.au.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

Obviously, there is a tradeoff between dampening noise's impact on the underlying signal and preserving authentic singularities in the data. Unfortunately, commonly-used data processing techniques in traffic engineering (e.g., average, moving average, second-order difference of cumulative summation, and oblique cumulative curve[1]) attenuate the impact of noise at the cost of distorting or totally losing the original information, and potentially important singularities in particular. Such practice has significant consequences, especially when the underlying process is subtle, e.g., a sudden reduction in roadway capacity. Findings using inferior methods might identify artifacts created by the methods themselves, leading researchers perhaps to conclusions contradictory to what the underlying data really suggest. Importantly, this matter has not received sufficient attention in the literature, and thus serves as a primary research motivation here.

Data encountered in transport often exhibit variations across time. As a time-frequency decomposition tool, the wavelet transform (WT) is particularly effective for extracting local information from non-stationary time-series by moving the wavelet location and squeezing or dilating the wavelet window. Such a time-scale representation of the original time-series data, which are often noisy and aperiodic, finds plentiful applications in fluid mechanics, engineering testing and monitoring, medicine, finance, geophysics, and network operations to name just a few (Addison, 2002). The WT has also been introduced to traffic engineering and intelligent transportation engineering. Combined with data mining techniques such as clustering, fuzzy logic, and neural networks, the WT has been adopted to investigate various traffic-related issues, such as automatic detection of freeway incidents (Adeli and Samant, 2000; Ghosh-Dastidar and Adeli, 2003; Karim and Adeli, 2002, 2003; Samant and Adeli, 2000, 2001), traffic features around freeway work zones (Adeli and Ghosh-Dastidar, 2004; Ghosh-Dastidar and Adeli, 2006), traffic flow forecasting (Boto-Giralda et al., 2010; Jiang and Adeli, 2005; Vlahogianni et al, 2007; Xie et al., 2007), and traffic pattern recognition (Jiang and Adeli, 2004; Vlahogianni et al., 2008). These pioneering studies have demonstrated the potential of WTs in analyzing non-stationary or noisy traffic data. Recently, Zheng et al. (2011a) demonstrated WT's capabilities of analyzing important features related to bottlenecks and traffic oscillations in a systematic manner. Furthermore, using WTs enabled the identification of the origins of stop-and-go driving and the measurement of microscopic features of wave propagation (Zheng et al., 2011b).

In applying WTs to solve practical problems, however, there are hundreds of different wavelets, and it is also quite straightforward to design a new wavelet for a particular research or practical question. As a consequence of the wide range of readily available and newly derived wavelets, researchers are free to select a WT without a reasoned justification or explanation. As a general rule, most WTs perform well if visual verification is satisfactory for the research purposes at hand. However, if more exacting precision is needed, such as microscopic features of traffic flow, selecting a different wavelet may produce significantly different results, suggesting that justification of wavelet selection is needed. Zheng et al. (2011a, 2011b) neither showed

[1] Although these methods are data de-noising and trend analysis techniques, they are often used essentially as singularity detection tools directly or indirectly in traffic flow characteristic analysis, see e.g., Mauch & Cassidy (2002) and Munoz & Daganzo (2003) among others.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

mathematically why WT is a powerful tool in detecting singularities in traffic and vehicular data nor discussed how to select a suitable wavelet. These two important topics are closely related. Without a sound understanding of the mathematical features of wavelets, selecting a good wavelet for a particular research application is problematic at best.
This paper aims to fill this gap and provide sufficient and clear guidance on how and why to prefer a particular wavelet to others. Towards this end, the remainder of this paper is organized as follows. Based on a numerical experiment, Section 2 demonstrates shortcomings of several data processing techniques that are widely used in traffic engineering to detect singularities. Section 3 provides a theoretical background of WT essential for understanding its capability for detecting singularities of non-stationary signals. Section 4 first presents criteria of selecting a suitable wavelet, then uses a numerical experiment to demonstrate how to select a suitable wavelet in the context of traffic engineering. Finally performance of the selected wavelet is further verified using several case studies. Section 5 discusses conclusions and future research.

2. Traffic data processing techniques

In this section the performances of popular techniques for analyzing traffic data in both the time and frequency domains are discussed and compared with that of WT using numerical simulation, with the intent to uniquely compare these techniques' performances and to underscore their advantages and disadvantages. Through such a comparative analysis, undesirable consequences of using these popular data processing techniques are demonstrated and superiority of WT is confirmed.

The use of numerical simulation enables the objective comparison of candidates' performances with the ground truth that is unattainable using field data (Cheng and Washington, 2005, Washington and Cheng, 2008). Using the same motivation, a numerical experiment is also employed in selecting a suitable wavelet, as is discussed later.

2.1 Numerical Experiment I (NE I)

To mimic vehicle counts in rush hours collected at a loop detector downstream of a bottleneck, a time series is randomly generated, representing a sample of traffic data with a mean of 12 vehicles and a standard deviation of 5 vehicles. The simulation period is 3 hours with a time resolution of 20 s (so the average flow is 2160 vehicles per hour). In total, 540 data points are generated. Note that the simulated data are stochastic and exhibit white noise properties except for a non-zero mean, as shown in Figure 1.

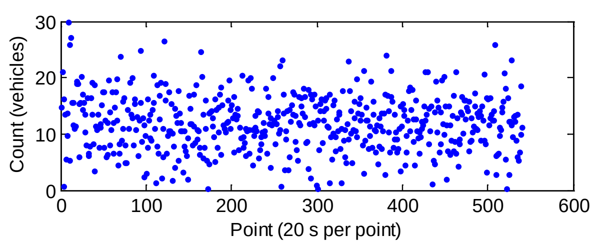


Figure 1 The vehicle counts from NE I

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

2.2 Averaging

To dampen or attenuate statistical noise in the traffic data, the simplest and one of the most commonly used techniques in the time domain is to aggregate data over a certain time interval (e.g., 5 or 15 minutes), which may be sufficiently effective to reveal long-term trends in traffic patterns. However, the shortcoming of this technique should be apparent: fine resolution information will be smoothed out and inaccurate or distorted information may be obtained because of the decreased time resolution. For example, depending on the starting point of averaging and the window size, data belonging to a single event may be divided into two or more groups and aggregated in different ways.

The vehicle counts in NE I are averaged for each 5 minute period as shown in Figure 2. Based on this figure, one might conclude that near point *a*, the lane capacity dropped from 15 vehicles per 20 s to 12 vehicles per 20 s. In other words, one might claim that the capacity of this bottleneck dropped by 20 percent. Recall, however, these data are simulated and randomly generated—thus no authentic capacity reduction exists.

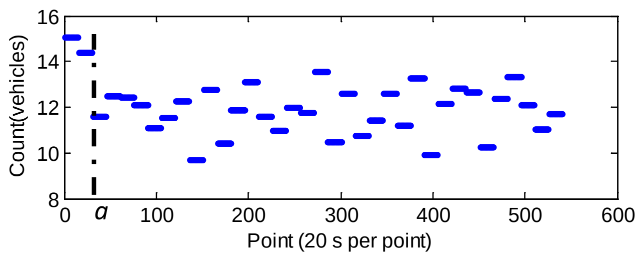


Figure 2 The averaged vehicle counts

2.3 Moving average

To preserve the time resolution, a moving average and its variants (e.g., weighted moving average) are often used to process time series traffic data. Generally, a moving average out-performs simple aggregation. However, a moving average suffers from some of the same issues confronted by simple averaging. Much of the most interesting information contained in the traffic data, i.e., singularities, is attenuated when a moving average is applied. Figure 3 demonstrates that a clear pattern appears once the vehicle counts in NE I are averaged using a 5-minute moving window. Note that for simplicity, points within the window are equally weighted.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

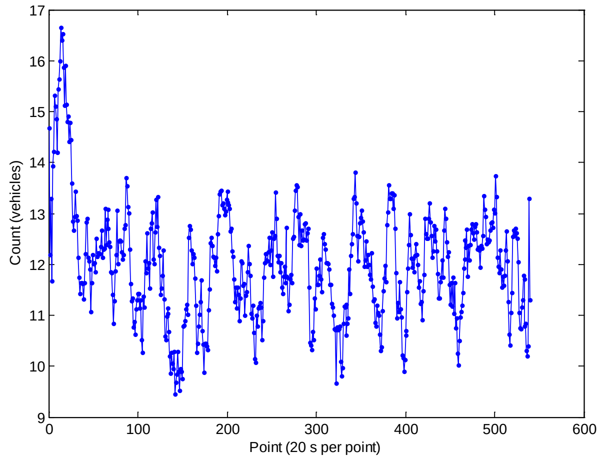


Figure 3 The moving-averaged vehicle counts

2.4 Second-order difference of cumulative data

To quantify traffic oscillations, researchers typically take the second-order difference of cumulative data sequence (e.g. vehicle counts, time-mean speeds) with a moving time window (Mauch and Cassidy, 2002, Ahn and Cassidy, 2007) to remove noise in the data and to de-trend longer-term changes, as defined in Eq. (1). This method is criticized for being sensitive to the choice of window length; with an inappropriate one, periodic oscillations may be dampened, or Gaussian white noise may be distorted into a periodically oscillating sequence (Li et al., 2010). This shortcoming is confirmed using the simulated data from NE I. For example, with the window size $2l = 50$, Figure 4 displays clear periodic patterns, which is further confirmed by the scalogram analysis[2], with scalograms of the original vehicle counts and the outcome of this method shown in figures 5(a)-(b), respectively. Note that the Mexican hat wavelet (MH) is used in the scalogram analysis for the reason discussed later. In Figure 5 (a), energy of the original data is almost uniformly distributed in the time-scale space except for the energy spikes at the boundaries due to the "boundary effect"[3]. In contrast, based on the energy spikes in Figure 5 (b), one may falsely claim that certain scales (frequencies) are prominent between points 100 and 300.

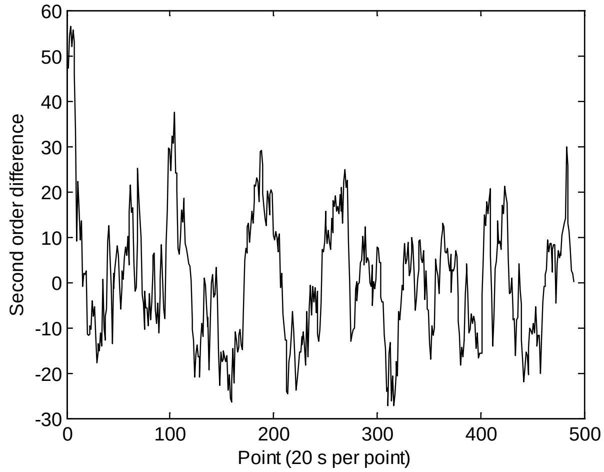
$$\breve{x}_j(l) = X_j - \frac{1}{2}(X_{j+l} + X_{j-l}) \qquad (1)$$

[2] Scalogram in the time scale space is equivalent to spectrogram in the time frequency space, communicating the time scale localization based on the energy density distribution. Note that the wavelet energy is defined as $E(a,b) = |T(a,b)|^2 = \left|w(a)\int_{-\infty}^{\infty} x(t)\,\psi(\frac{t-b}{a})dt\right|^2$, where $T(a,b)$ is the WT coefficient at (a,b) and a is a scale parameter that governs the dilation and contraction of the wavelet, and b is a translation parameter that governs the movement of the wavelet along the time dimension. $x(t)$ is a continuous signal; $\psi(t)$ is a wavelet. $w(a)$ is a weighting function and typically set to $1/\sqrt{a}$ to ensure that wavelets at all scales have the same energy. See Zheng et al. (2011a) for details.

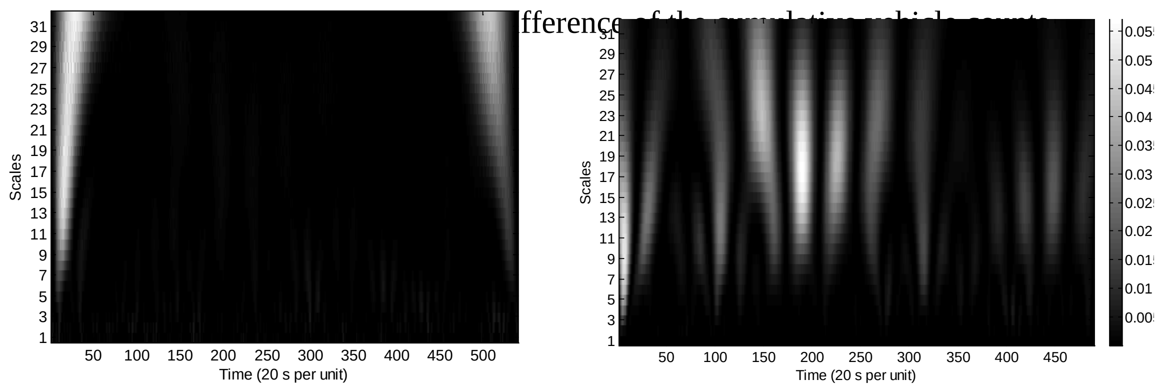
[3] The "boundary effect" is a common phenomenon in signal processing and characterized by large WT coefficients at both ends of the signal range. This occurs because the signal range is finite, and the value outside the signal range is assumed to be zero. Large WT coefficients are obtained at the boundaries where the signal shifts from zero to an actual non-zero value. The "boundary effect" also appears in other examples discussed later in this paper.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

Where $X_j = \sum_{i=1}^{j} x_i$; i and j are positive integers; l is half of the window length.



fference



(a) (b)

Figure 5 The scalogram analysis

(a) Scalogram of the original simulated data; (b) Scalogram of the second-order differenced data

2.5 Oblique cumulative curve

Oblique cumulative curve is another widely used analysis method intended to attenuate noise and reveal the underlying data trend. As defined in Eq. (2), an oblique curve is constructed by taking the difference between the cumulative measurement at time t, and a background reduction (Munoz and Daganzo, 2003). This method is simple to use and often reveals interesting patterns that are invisible through plotting of the original data and as a result is gaining in popularity.

$$\breve{x}(t) = X(t) - X_0 \times (t - t_0) \qquad (2)$$

where $X(t) = \sum_{j=t_0}^{t} x(j)$; X_0 is a scaling factor and t_0 is the starting time.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

In addition to the issues (e.g., subjective and labor intensive) identified previously (Zheng et al., 2011a), this method purports to provide a superior visual glimpse of the underlying phenomenon without providing solid theoretical support for such a claim. Thus, the soundness of its output requires additional exposition. The oblique curve of the data from NE I (Figure 6) shows different slopes for segments *ab*, *bc*, and *cd*, which implies different traffic flow rates. Based on curves like this one, researchers (Cassidy and Bertini, 1999) often measure characteristics of the bottleneck, e.g., capacity reductions, and activation and deactivation times. Intuitively, such different slopes represent random rather than structural differences in these data, as was the case previously. Its poor performance in changing traffic states is discussed later.

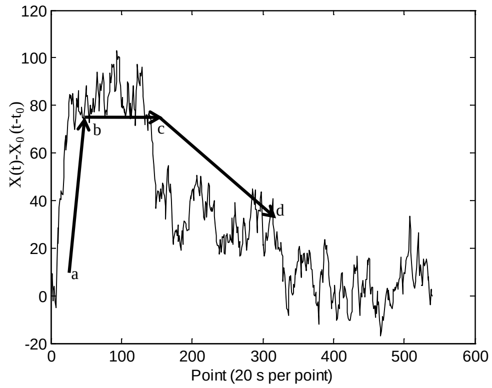


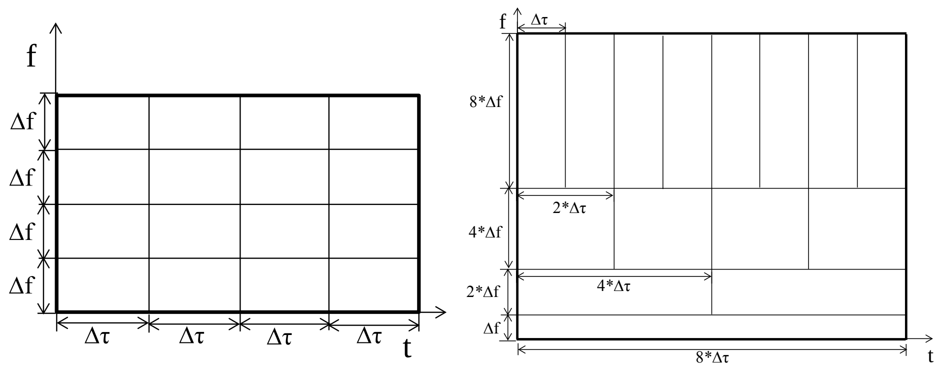
Figure 6 The oblique cumulative curve of the simulated data from NE I

2.6 Short-time Fourier transform

Techniques discussed previously are from the time domain. Techniques from the frequency domain are also used by researchers in traffic engineering to process data and extract useful information. Among them, a short-time Fourier transform (STFT) is frequently used. The STFT is designed to remedy a serious issue suffered by classic Fourier transform (FT) in that FT can only extract frequency information globally. Inaccurate frequencies are identified if FT is applied to data that contain local variations, as is the case for most traffic data. Stated simply, STFT can be regarded as a series of FTs: a signal is divided into several segments and FT is implemented within each segment. To improve its performance, neighboring segments often partially overlap.

However, time resolutions and frequency resolutions of STFT cannot be chosen arbitrarily. According to the Heisenberg uncertainty principle (Gröchenig, 2001), the product of time and frequency resolutions is lower-bounded by $\Delta t * \Delta f \geq 1/(4\pi)$. As a result, STFT's time and frequency resolutions must be implemented as demonstrated in Figure 7 (a). This figure shows that once the window size is selected, its resolution is fixed, which provides the historical motivation of developing WT (Addison, 2002).

When STFT is applied to the simulated data from NE I, the spectrogram is shown in Figure 8. This figure reveals several energy spikes, which should not exist.



(a) (b)

Figure 7 Time and frequency resolution diagrams
(a) STFT; (b)WT

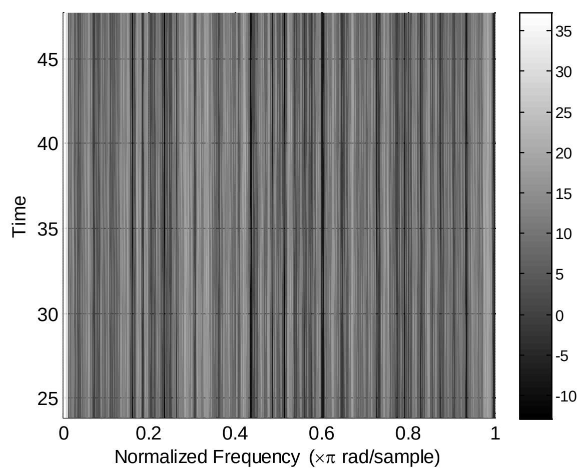


Figure 8 The spectrogram of the simulated data from NE I using STFT

Studies have been implemented to formally compare performance of STFT with that of WT. The superiority of WT's capability in extracting accurate localized information from non-stationary signals is generally confirmed (Addison, 2002).

2.7 Wavelet transform

In contrast to STFT, WT can be applied at different frequency and time resolutions (although it is also constrained by the Heisenberg Uncertainty Principle (Gröchenig, 2001)), as demonstrated in Figure 7 (b). This flexibility makes WT a powerful tool for accurately detecting local variations of non-stationary signals.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

When WT is used to analyze the simulated data from NE I, a uniform distribution of the energy is obtained, except for the energy spikes at the boundaries as shown in Figure 5(a). This uniform energy distribution is consistent with randomly generated synthetic data, unlike the energy spikes found using the previously described methods. The theoretical background on WT and how to select a suitable wavelet are now discussed.

Note that a sensitivity analysis has been implemented to investigate the robustness and consistency of the data processing methods by generating eight datasets with different standard deviations (STD) (i.e., STD=1, 2, …,8. STD is unlikely to be over 8 since it is in rush hours and traffic demand is relatively stable). The mechanism used to generate these datasets is the same as in NE I except that a different STD is used. Then the data processing techniques discussed in this paper were applied to analyze each of the datasets. Distorted and inconsistent representations were found from all the techniques except WT. More specifically, for commonly-used data processing techniques, a similarly distorted representation was generated for each dataset when the STD was less than 5, whereas distorted representation varied when the STD was above 5. In contrast, the scalogram of WT for each dataset consistently showed the randomness of the data series. This additional analysis further highlights the inappropriateness of these commonly-used data processing techniques and the robustness of WT in detecting the underlying structure of data series contaminated by noise. Furthermore, these techniques' performances in detecting singular points from changing traffic states are also evaluated and compared later (see 4.2) and the same conclusion is obtained.

3. Detecting singularities using wavelets: Theoretical background

Detecting and measuring singularities are well studied in mathematics. The local regularity of a function $f(x)$ is often measured using Lipschitz exponents. A function $f(x)$ is said to be Lipschitz α at x_0 if and only if the following condition is met:

$$|f(x_0 + h) - P_n(h)| \leq A|h|^\alpha \text{ if } h < h_0$$

where n is a positive integer; A and h_0 are positive constants; $n \leq \alpha \leq n + 1$; $P_n(h)$ is a polynomial of order n at h. Lipschitz exponents of $f(x)$ over an interval can be similarly defined (Mallat and Hwang, 1992).

Unlike the asymptotic decay of $f(x)$'s FT, which can be used to measure global Lipschitz regularity of $f(x)$ but is not able to describe its local regularity, the asymptotic decay of $f(x)$'s WT is effective in measuring the local Lipschitz regularity at a point x_0 or at an interval (Mallat and Hwang, 1992). Before mathematical features of WT are discussed, two important concepts need to be introduced: vanishing moments and compact support.

A wavelet $\psi(x)$ is said to have n vanishing moments if and only if it satisfies

$$\int_{-\infty}^{+\infty} x^k \psi(x) dx = 0$$

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

where $k = 1, 2, \dots, n-1$.

Support of a function $f(x)$ is a widely used concept in mathematical analysis. The support of $f(x)$ is the set of points where $f(x) \neq 0$. A wavelet $\psi(x)$ is said to have a compact support in X if and only if its support is a compact subset of X (Mallat, 2009). Loosely speaking, $\psi(x)$ with a compact support is zero at infinity and negative infinity. Based on the definition, $\psi(x)$ with a compact support of size K produces K high-amplitude coefficients at each scale. As a consequence, the number of high-amplitude coefficients can be reduced by minimizing the support size of $\psi(x)$.

Theoretically, the Lipschitz regularity of $f(x)$ either at a point x_0 or over an interval can be precisely estimated from the asymptotic decay of its WT if $\psi(x)$ has enough vanishing moments with a compact support (Mallat and Hwang, 1992). However, it is difficult to use this property to directly detect local singularities and characterize the Lipschitz exponents due to enormous computational cost. To remedy this issue, researchers have developed an effective method for detecting local singularities by combining the modulus and the phase information of WT (Mallat and Hwang, 1992). Without delving into mathematical details, a simple example depicted in Figure 9 is used to demonstrate this basic idea.

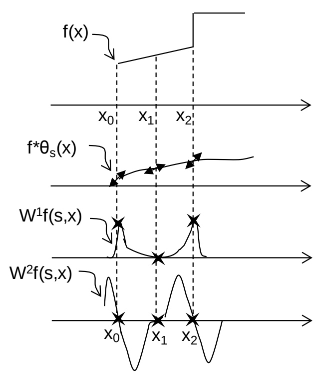


Figure 9 Illustration of using WT to detect singularities
(Adapted from Mallat and Hwang (1992))

In this figure, $f(x)$ is a piece-wise linear function with two singularities at x_0 and x_2. $\theta(x)$ is a smoothing function and $\theta_s(x) = \left(\frac{1}{s}\right) * \theta(\frac{x}{s})$. Two wavelets are defined as $\psi^1(x) = \frac{d\theta(x)}{dx}$ and $\psi^2(x) = \frac{d^2\theta(x)}{dx^2}$, respectively. Their corresponding WTs of $f(x)$ are defined as $W^1 f(s,x) = f * \psi_s^1(x)$ and $W^2 f(s,x) = f * \psi_s^2(x)$. This figure clearly reveals that the local regularities of $f(x)$ can be well characterized by either the zero-crossings of $W^2 f(s,x)$ or the local maxima of $W^1 f(s,x)$. However, $x_0 \sim x_2$ all correspond to zero-crossings in $W^2 f(s,x)$. Slow variations at $x_1$and sharp variations at x_0 and x_2 cannot be distinguished using $W^2 f(s,x)$. To accomplish this, modulus maxima of a WT is used to measure local regularities of a signal. Specifically, the nearest local maxima at different scales are connected to form local maxima lines. Isolated local maxima are excluded because they are caused by noise. Then, accurate locations of singularities

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

are identified using the corresponding local maxima line at the finest scale (i.e., scale 1). It has been mathematically proved that the Lipschitz exponents of the signal singularities are characterized by these modulus maxima (Mallat and Zhong, 1992).

Completeness of the wavelet modulus maxima has also been studied. Although it has been proven that a unique function cannot be constructed from the wavelet modulus maxima, Mallat and Zhong (1992) developed a numerical algorithm with the capability of reconstructing a close approximation of the original signal.

The theories briefly discussed above lay the foundation for using WT to detect singularities and/or remove noise in a signal. Note that denoising traffic data using WT is not discussed in this paper.

4. Wavelet selection for detecting singularities

4.1 Selection criteria

Although several widely known wavelets (e.g., Haar, Daubechies, MH, Morlet, Coiflets, etc.) are near-optimal and provide similar results for a wide variety of signals (Donoho, 1993), selecting different wavelets may have significant influences on analysis results. An inappropriate wavelet cannot accurately extract underlying information as demonstrated by examples discussed later in this paper, which highlights the importance of selecting a suitable wavelet for detecting singularities in traffic and vehicular data.

Several factors need to be considered in selecting a wavelet, e.g., orthogonality, symmetry, support size and vanishing moments. Among them, support size and vanishing moments play key roles.

The number of vanishing moments of a wavelet is determined by the maximum Lipschitz exponents we want to estimate. For example, a wavelet with one vanishing moment cannot detect the singularities in the derivative of a function. Obviously, wavelets with more vanishing moments are able to measure higher Lipschitz exponents. However, the number of maxima lines is also increased accordingly, which is not desirable because the complexity and the computational cost are increased. Therefore, a good candidate wavelet should 1) have as few vanishing moments as possible; yet 2) have enough vanishing moments to detect the highest Lipschitz exponents of interest. In traffic engineering, we are often only interested in detecting discontinuities and peaks that have Lipschitz exponents not greater than two because the maximum derivative of interest is normally two (i.e., acceleration of a vehicle), whereas wavelets with two vanishing moments are theoretically sufficient. Another important factor in selecting a suitable wavelet is the support size. Roughly speaking, a wavelet with a smaller support size is more efficient in detecting singularities and denoising data. It can be mathematically proven (Daubechies, 1992) that the minimum achievable support size of $\psi(x)$ with p vanishing moments is $(2p-1)$. Note that not all wavelets have the minimum achievable support size,

when a particular wavelet is chosen, the analyst faces a trade-off between the number of vanishing moments and the support size.
In practice, more than one wavelet that satisfies the vanishing moments and the support size requirements often exists. Thus, researchers have developed various methods to quantitatively compare and select a wavelet (e.g., Energy and Shannon entropy based measures, similarity based measures, information theory based measures, and their combinations) (Addison, 2002, Yan, 2007). In this paper, Shannon entropy is utilized to select a wavelet because it is extensively studied and widely used (Briggs and Levine, 1997, Katul and Vidakovic, 1996). For a discrete probability distribution p_i ($\sum_1^n p_i = 1$), the Shannon entropy is defined as $S(p) = -\sum_1^n p_i log\,(p_i)$. The normalized energy distribution of WT is used as the value of p_i. Controlling for other factors, a smaller Shannon entropy yields better performance.

In summary, to objectively select a suitable wavelet for detecting singularities in traffic and vehicular data, one first considers theoretical properties of candidates (e.g., vanishing moments and support size) and then uses Shannon entropy to quantitatively compare them. A final decision is made based on this comparison and locations of local maxima lines. This approach is now demonstrated using a numerical experiment, with the performance of the selected wavelet discussed in the context of several problems researchers in traffic engineering typically face.

4.2 Numerical Experiment II (NE II)

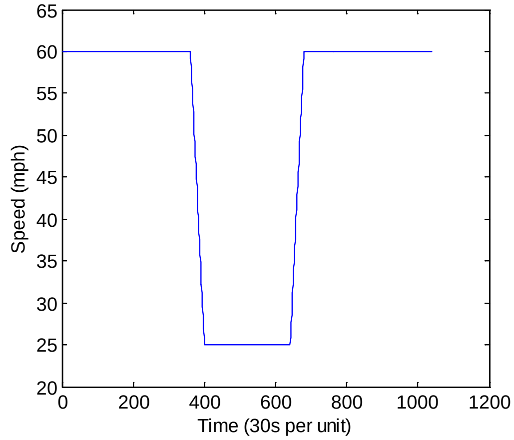
To demonstrate and compare the performance of different wavelets for detecting singularities in traffic data, a numerical experiment is designed. As previously mentioned, using numerical experiments supports the objective comparison of candidates' performances with the ground truth that is often unknown in real data. Using this approach high performing wavelets can be identified.

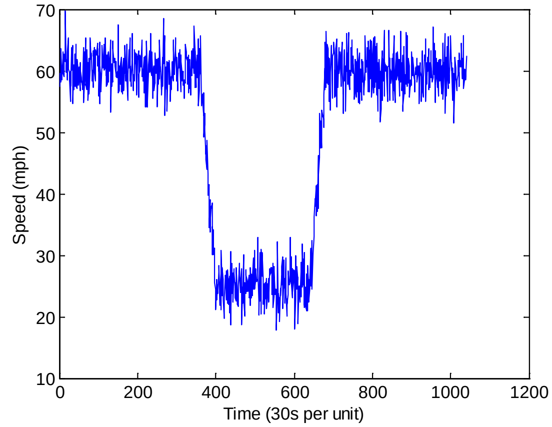
In this experiment, a speed time series is simulated to represent a typical speed profile collected at a loop detector. The data are simulated to progress through the following phases sequentially (number in the parentheses is the duration of each phase):

Free flow (3 hr) → Transition (20 min) → Congestion (2 hr) → Transition (20 min) → Free flow (3 hr)

Note that the transition represents a period in which traffic flow is transitioning from a free flow to a congested state and vice versa (Gazis, 2002; Munoz & Daganzo, 2003; Bertini & Ahn, 2010). Although the duration of a transition period varies, 20 minutes is typical (Munoz & Daganzo, 2003). Of keen interest are four singularities (i.e., A, B, C and D) in Figure 10 (a) because they are indicative of traffic state changes. Accurately identifying these state changes is of great importance to traffic operations. Unfortunately, such nicely behaved data are not readily available in reality due to traffic dynamics and noise. To make the simulated data more realistic, noise is added at a signal-to-noise ratio (SNR) of 25, as shown in Figure 10 (b).

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.





(a) (b)

Figure 10 NE II: a typical speed profile collected by a loop detector
(a) Without noise; (b) Noise is added at SNR of 25

Considering the requirements on the vanishing moments and the support size discussed previously, wavelets considered in this paper include Haar, the Daubechies family (D(n), n is the order of the Daubechies wavelet, similar notations are also used for other wavelet families), MH, the Coiflets family (Coif(n)), the Symlets family (Sym(n)), Meyer, Morlet, and the Gaussian family (Gaus(n)) because they are well studied and widely used in other fields (Addison, 2002, Daubechies, 1992, Mallat, 2009). Their mathematical definitions are included in the Appendix. Note that although some of the candidates do not have a compact support, e.g., MH, Gaus(n), Meyer, and Morlet, they have a small range of effective support size (see the Appendix). In addition, whilst wavelets with two vanishing moments should be sufficient for detecting abrupt changes in traffic data (as described previously), considering noise in the field data, the wavelet candidate pool is extended to include ones with three or four vanishing moments.

We also assume that the duration of traffic oscillations (stop-and-go driving) is around a couple of minutes, e.g., 4~5 minutes (Laval & Leclercq, 2010; Li et al., 2010), which is about 0.0039 Hz. Thus, 0.0039 Hz is set as the lowest central frequency to which the wavelet candidates should be decomposed[4]. The corresponding maximum decomposition scale and the Shannon entropy of each candidate are summarized in Table 1.

Table 1 Max decomposition scales, Shannon entropies and performances of the candidates
('-' means that result is not provided due to significant error, which is often above 10)

[4] For issues without sufficient prior knowledge, a pragmatic approach to identifying a reasonable lowest central frequency for implementing WT decomposition is through trial and error. More specifically, WT can be implemented up to a scale that is obviously too big. Then a series of statistical tests along with Monte Carlo simulations can be utilized to test whether the energy output at a particular scale is significantly different from the one from a white or red noise (see Torrence & Compo (1998)). Moreover, the energy output at a particular scale should not be dominated by the "boundary effect"(see footnote 3).

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

Name	Max scale	Shannon	A	B	C	D	Average error*
			360	400	640	680	
MH	64	14.2	362	400	648	680	2.5
Gaus2	77	14.3	362	400	631	678	3.25
Gaus3	103	14.7	378	407	647	707	14.75
D2	170	15.3	363	407	647	685	5.5
Sym2	170	15.3	363	407	647	685	5.5
Gaus4	128	15.3	362	402	628	678	4.5
Sym4	185	15.6	-	-	-	-	-
Coif2	185	15.6	-	-	-	-	-
Haar	256	15.7	-	-	-	-	-
D4	185	15.8	-	-	-	-	-
Coif1	205	15.8	-	-	-	-	-
Meyer	175	15.8	-	-	-	-	-
D3	205	15.9	-	-	-	-	-
Sym3	205	15.9	-	-	-	-	-
Morlet	210	16.2	-	-	-	-	-
Gaus1	51	12.7	-	-	-	-	-

* $average\ error = \sum_{i=1}^{4} |x_i - x_{ti}| / 4$; where x_i is a detected singular point and x_{it} is a true singular point.

If Shannon entropy measures are solely relied upon and theoretical properties ignored, a Gaus1 would emerge as the best wavelet. However, as depicted in Figure 11(a), four state changes are not successfully detected by the local maxima lines produced by Gaus1. Specifically, only two useful maxima lines are produced, because Gaus1 has only one vanishing moment. Applying the same criteria, Haar is also excluded from any further comparison, although its Shannon entropy is not the largest. Morlet has the biggest Shannon entropy. Not surprisingly, its performance on this numerical experiment is poor. Figure 11(b) shows that the four turning points detected by Morlet are 295, 387, 589, and 678, which are significantly different from the ground truth (360, 400, 640, 680). For the same reason, other candidates with high Shannon entropy measures (i.e., Sym3, D3, Meyer, Coif1, D4, Coif2, and Sym4) are excluded.

The remaining candidates can detect these state changing points with differing degrees of success, as shown in Table 1. Gaus3 has small Shannon entropy but a high average error in detecting the changing points. The average error is calculated as the average absolute difference between the measured points and the ground truth. Gaus4 performs well; however, two extra maxima lines are produced in Figure 11(c), which is due to Gaus4 having four vanishing moments. As shown in figures 11(d)-(e), D2, Sym2 and Gaus2 also perform well, but confounding maxima lines appear, too. Note that D2 and Sym2 essentially have the same performance, which is due to their almost identical mathematical features (see the Appendix).

Among the candidates, MH reveals the best performance. As shown in the middle of Figure 11(f), four energy spikes that correspond to four singularities in the original signal are present in the energy contour. Furthermore, the locations of these points are accurately captured by the maxima lines and no confounding maxima lines are produced, which make it possible to automate this process, e.g., Zheng et al. (2011a) proposed a method using the average wavelet-based energy to automatically detect abrupt speed changes.

A sensitivity analysis was implemented by repeating the above procedure to signals with different SNRs (i.e., SNR=10, 15, 20, or 30) and similar conclusions were obtained.

To further demonstrate the superiority of WT (MH in particular), the popular methods discussed in NE I with a stationary traffic state was also applied to NE II with a changing traffic state. Unsatisfactory performances of these popular methods were observed. As an example, Figure 12 is the outcome produced by using the oblique cumulative curve, which clearly shows that this method only detected two singular points (i.e., 373 and 670) and thus completely failed to detect the two transition periods. Similar observations were found for other methods.

Next, to further verify MH's performance, it is used to study several well-known research topics in traffic engineering using real data.

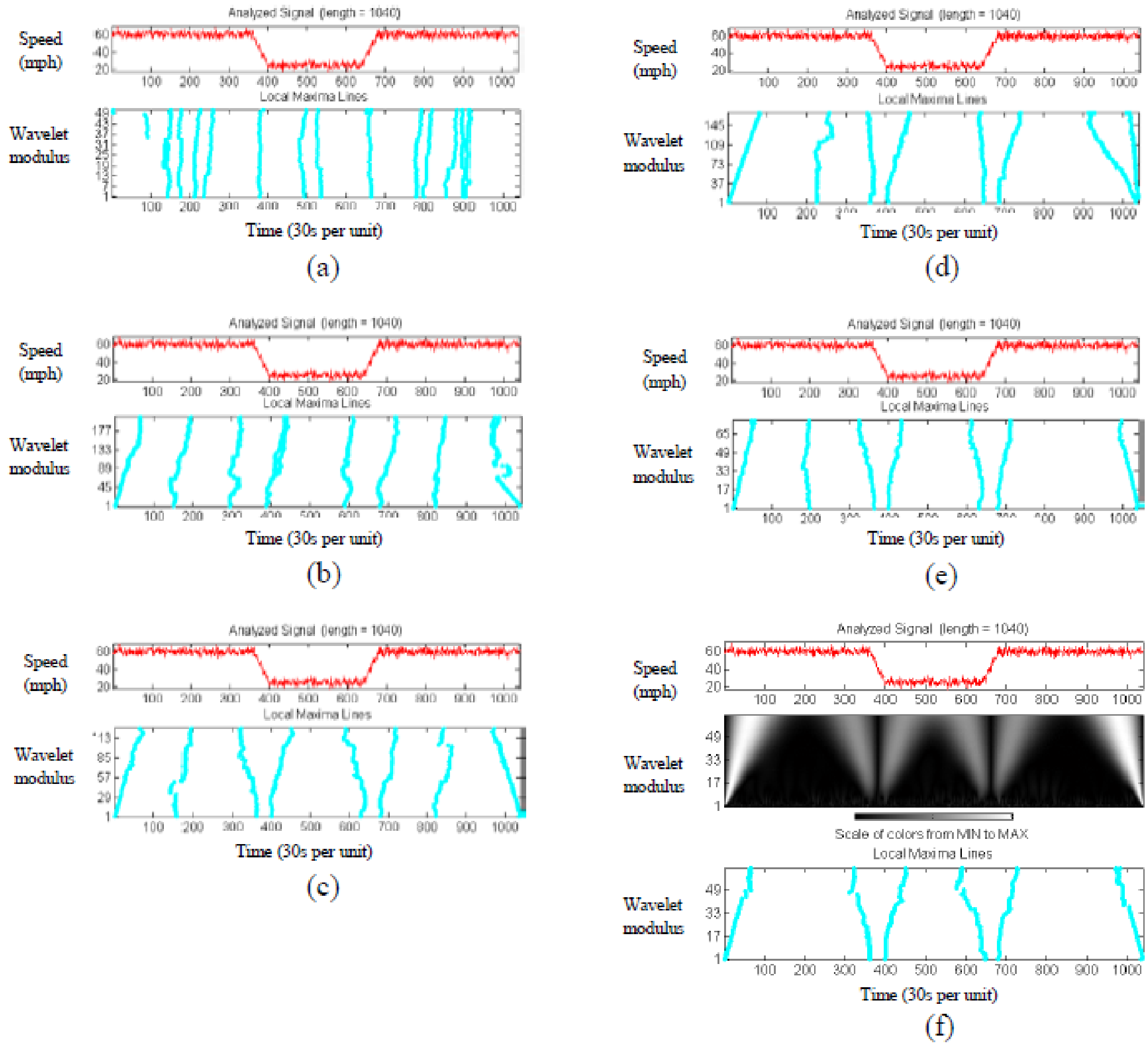
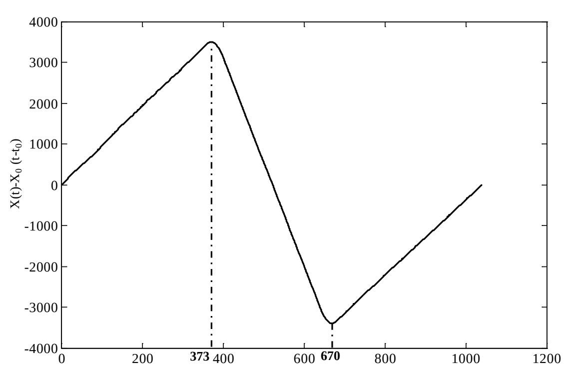


Figure 11 Performances of the candidate wavelets
(a) Gaus1; (b) Morlet; (c) Gaus4; (d) D2 & Sym2; (e) Gaus2; (f) MH



This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

Figure 12 The oblique cumulative curve of the simulated data from NE II

4.3 Case studies

In this section, field data collected from loop detectors and video cameras are used to demonstrate that a well-selected wavelet is capable of detecting information-rich singularities in traffic and vehicular data. Conclusions derived from NE II are confirmed by these case studies. Note that in-depth analysis of the topics used in the case studies is on-going and beyond this study's scope.

4.3.1 Using loop detector data

A segment of US-101 data (from milepost (MP) 19.832 to MP 25.562) in Los Angeles, California is used in this first case study (see Figure 13). Vehicle counts and occupancies were collected by loop detectors and aggregated every 30 seconds (PeMS, 2008). Speed at each loop detector is not measured directly and estimated using Eq. (3) (Kockelman and Ma, 2007):

$$\hat{\hat{v}}_{l,t} = \left(\frac{C_{l,t}}{30/3600}\right) \times \left(\frac{\bar{l}_{l,t}+l_d}{5280 \times OCC_{l,t}}\right), \qquad (3)$$

Where $\hat{\hat{v}}_{l,t}$: is the estimated time-average speed at location *l* and time *t*, $C_{l,t}$: is the vehicle count, $OCC_{l,t}$: :is the occupancy (%) defined as the average percentage of time that the detectors are occupied by vehicles, $\bar{l}_{l,t}$:is the average vehicle length, and l_d: is the effective detection zone length.

The sum of $\bar{l}_{l,t}$ and l_d is the effective vehicle length, which is assumed to be 18 ft. The average effective vehicle length can change by time of day and tends to be bi-modal around effective lengths for passenger cars (around 20 ft) and larger vehicles (around 60 ft) (Kwon et al., 2003, Coifman, 2001). However, because our method is designed to detect considerable temporal changes in speed, the accuracy of the estimated effective vehicle length is not critical.

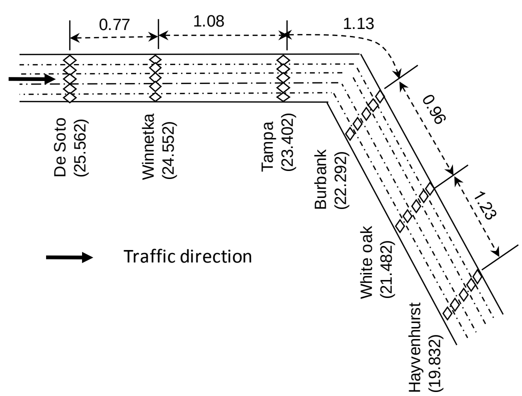


Figure 13 Schematic of US 101, MP 19.832 - 25.562 (Unit: miles)

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

(1) Detecting traffic state changing points

Detecting traffic state changing points is an important task in traffic operations because such information is critical for congestion management and for understanding other important questions, e.g., onset of traffic congestion, traffic wave propagation, and driving behaviors in different traffic states. MH is capable of accurately detecting these singularities. To illustrate, the speed time series in the first lane at MP 21.482, US-101, September 1 2009 and the output of its WT are shown in Figure 14. The figure reveals that four state-changing points are accurately located by the local maxima lines of WT. Specifically, onset of the congestion is around point A, and the congestion is fully developed around point B, and lasts until point C. Traffic returns to the normal condition near to point D.

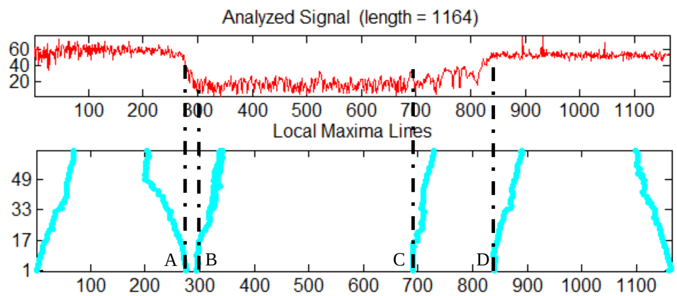
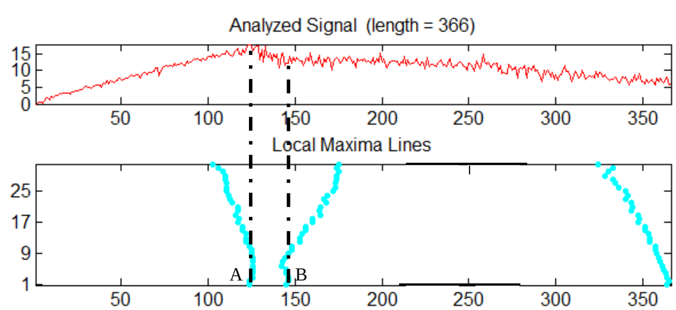


Figure 14 Detecting traffic state changing points using MH

(2) Detecting discontinuity of the FD diagram

An empirical fundamental diagram (FD) constructed by plotting vehicle count (or flow) and occupancy(or density) is potentially a powerful tool to study traffic phenomena, e.g., to analyze transitions (Munoz and Daganzo, 2003). The start of transition is identified when flow-occupancy pairs start to drift from the uncongested to the congested branches of the empirical FD. The end of transition is identified when the data points finally reach the congested branch. However, identifying the precise event times can be rather subjective or inaccurate because of noise in the data and reduced data resolution. To again illustrate, MH is implemented to detect the start and the end points of a transitioning from the free flow to the congested regime based on the FD diagram at MP 24.552, September 1 2009. Result is shown in Figure 15, which clearly reveals that the transitioning started around point A and ended around point B.



This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

Figure 15 Detecting discontinuity of FD using MH

4.3.2 Using vehicle trajectory data

This case study uses vehicle trajectory data collected as part of FHWA's Next Generation Simulation (NGSIM) program (NGSIM, 2010). The study site, shown in Figure 16, is a 2100-ft section on US-101 southbound in Los Angeles, California (note that the lane numbering is incremented from the left-most lane). Trajectory data were collected with the resolution of 10 records per second from 7:50 a.m. to 8:35 a.m. on June 15, 2005.

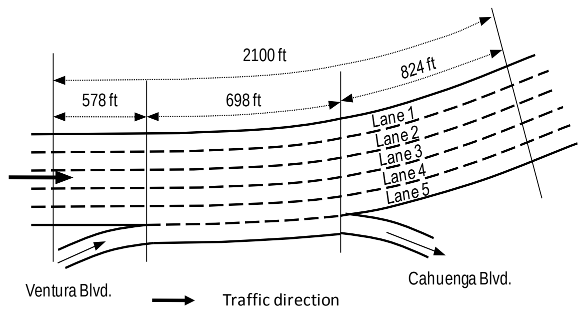


Figure 16 Schematic of a 2100 ft section of US-101

(1) Detecting start point of acceleration (deceleration) waves

As shown in Figure 17, the speed profile of vehicle 1753 (the top figure) is used to demonstrate the performance of MH in detecting start points of acceleration (deceleration) waves. The local maxima lines produced by MH are shown in the bottom figure. Figure 17 convincingly shows that each local maxima line at the finest scale (points A, B, C, D, and E) can be traced back to an abrupt change in the speed profile, signaling arrivals of acceleration (deceleration) waves.

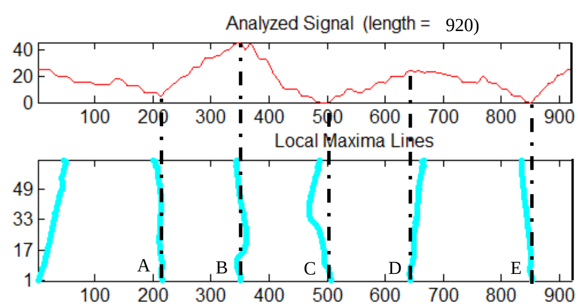


Figure 17 Detecting start point of acceleration (deceleration) waves using MH

(2) Determining start and end of vehicle lane-changing maneuvers (LCMs)

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

LCMs on freeways have received increased attention due to their negative impact on crashes (Zheng et al., 2010) and their macroscopic impacts on traffic streams such as capacity reduction, traffic oscillations, and delay (Zheng et al., 2011c). To study characteristics of LCM, the first task is to find a tool that can reliably determine the start and the end of a LCM. MH is such a tool. As shown in Figure 18, by visually checking the lateral trajectory of vehicle 373, we observe that the vehicle starts changing its lanes around point A and completes it around point B. Such information is successfully detected by the local maxima lines of its WT, despite the notable fluctuations contained in its lateral trajectory.

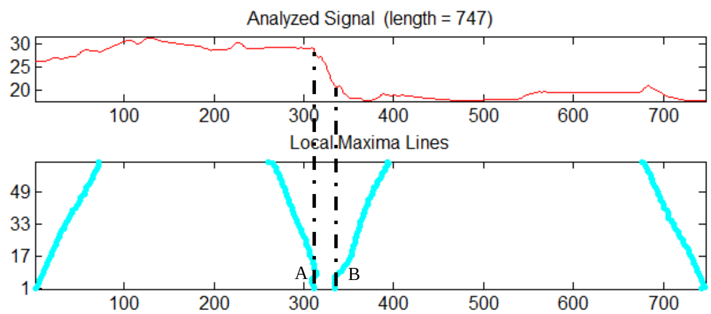


Figure 18 Determining start and end of LCM using MH

5. Conclusions

Based on two numerical experiments, several commonly-used data processing techniques are shown to be inappropriate for detecting singularities in traffic and vehicular data. A distorted representation of the original signal is likely to be obtained and incorrect conclusions may be drawn using these techniques. Particularly, singular points are falsely detected when traffic conditions are stable; however, true singular points in non-stationary traffic states are either inaccurately detected or completely ignored by these techniques. In contrast, WT has been demonstrated as a powerful tool to analyze complex traffic phenomena, e.g., activation and deactivation of a bottleneck and characteristics of traffic oscillations (Zheng et al., 2011b, Zheng et al., 2011a). However, neither theoretical background essential to WT's ability to detect singularities nor how to select a suitable wavelet is provided. This paper fills this gap in the literature.

It has been mathematically proven by others that the Lipschitz exponents of the signal irregularities are characterized by the modulus maxima produced by WT. Researchers have developed a numerical algorithm with the capability of reconstructing a close approximation of the original signal. These nice mathematical features indicate that local maxima lines can be used to effectively detect singularities in data. Specifically, the accurate locations of singularities can be identified using the corresponding local maxima line at the finest scale.

This paper addresses another important question: how to select a suitable wavelet among many candidates. In selecting a suitable wavelet for a particular research question, theoretical properties of candidates need to be first considered (e.g., vanishing moments and support size), then Shannon entropy and the result from numerical experiments can be utilized to quantitatively compare them. This approach is demonstrated using a numerical experiment and MH is selected as a suitable wavelet for detecting singularities in traffic and vehicular data. To further verify performance of MH, it is applied to detect traffic state changing points, transitioning period in FD, arrivals of acceleration (deceleration) waves, and the start and the end points of LCM. Good results are consistently obtained.

In further work, the authors are using the WT to investigate some other well-known traffic phenomena such as capacity drops e.g. see Hall and Agyemang-Duah, 1991), and hysteresis effects (e.g. see Treiterer and Myers, 1974)).

References

ADDISON, P. S. 2002. The illustrated wavelet transform handbook : introductory theory and applications in science, engineering, medicine and finance, Bristol, Institute of Physics Publishing.

ADELI, H., GHOSH-DASTIDAR, S. 2004. Mesoscopic-wavelet freeway work zone flow and congestion feature extraction model. Journal of Transportation Engineering 130 (1), 94–103.

ADELI, H., SAMANT, A. 2000. An adaptive conjugate gradient neural network-wavelet model for traffic incident detection. Computer-Aided Civil andInfrastructure Engineering 15 (4), 251–260.

AHN, S. & CASSIDY, M. J. Freeway traffic oscillations and vehicle lane-change maneuvers. In: ALLSOP, R. E., BELL, M. G. H. & HEYDECKER, B. G., eds. International Symposium of Transportation and Traffic Theory, 2007 Amsterdam. Elsevier, 691-710.

ARGOUL, F., ARNEODO, A., ELEZGARAY, J. & GRASSEAU, G. Wavelet analysis of fractal growth process. Proc. 4th EPS Liquid State Conf., 1988 Arcachon, France.

BERTINI, R., AHN, S. 2010. Characteristics of transitions in freeway traffic. Project report: OTREC-RR-10-12.

BOTO-GIRALDA, D., DÍAZ-PERNAS, F., GONZLEZ-ORTEGA, D., DÍEZ-HIGUERA, J., ANTÓN-RODRÍGUEZ, M., MARTÍNEZ-ZARZUELA, M., TORRE-DÍEZ, I. 2010. Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks. Computer-Aided Civil and Infrastructure Engineering 25 (7),530–545.

BRIGGS, W. M. & LEVINE, R. A. 1997. Wavelets and field forecast verification. Monthly Weather Review, 125, 1329-1341.

CASSIDY, M. J. & BERTINI, R. L. 1999. Some traffic features at freeway bottlenecks. Transportation Research Part B-Methodological, 33, 25-42.

CHENG, W. & WASHINGTON, S. P. 2005. Experimental evaluation of hotspot identification methods. Accident; analysis and prevention, 37, 870-81.

COIFMAN, B. 2001. Improved velocity estimation using single loop detectors. Transportation Research Part a-Policy and Practice, 35, 863-880.

DAUBECHIES, I. 1992. Ten lectures on wavelets, Philadelphia, Society for Industrial and Applied Mathematics.

DONOHO, D. L. 1993. Unconditional bases are optimal bases for data compression and for statistical estimation. Applied and Computational Harmonic Analysis, 1, 100-115.

GAZIS, D.C. 2002. Traffic theory. Kluwer Academic Publishers.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

GHOSH-DASTIDAR, S., ADELI, H. 2003. Wavelet-clustering-neural network model for freeway incident detection. Computer-Aided Civil and Infrastructure Engineering 18 (5), 325–338.

GHOSH-DASTIDAR, S., ADELI, H. 2006. Neural network-wavelet microsimulation model for delay and queue length estimation at freeway work zones. Journal of Transportation Engineering 132 (4), 331–341.

GRÖCHENIG, K. 2001. Foundations of time-frequency analysis, Boston, Birkhäuser.

HALL, F. L. & AGYEMANG-DUAH, K. 1991. Freeway capacity drop and the definition of capacity. Transportation Research Record, 1320.

JIANG, X., ADELI, H. 2004. Wavelet packet-autocorrelation function method for traffic flow pattern analysis. Computer-Aided Civil and Infrastructure Engineering 19 (5), 324–337.

JIANG, X., ADELI, H. 2005. Dynamic wavelet neural network model for traffic flow forecasting. Journal of Transportation Engineering 131 (10), 771–779.

KARIM, A., ADELI, H. 2002. Incident detection algorithm using wavelet energy representation of traffic patterns. Journal of Transportation Engineering 128(3), 232–242.

KARIM, A., ADELI, H. 2003. Fast automatic incident detection on urban and rural freeways using wavelet energy algorithm. Journal of Transportation Engineering 129 (1), 57–68.

KATUL, G. & VIDAKOVIC, B. 1996. The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. Boundary-Layer Meteorology, 77, 153-172.

KOCKELMAN, K. M. & MA, J. 2007. Freeway speeds and speed variations preceding crashes, within and across lanes. Journal of the Transportation Research Forum, 46, 43-62.

KWON, J., VARAIYA, P. & SKABARDONIS, A. 2003. Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation. Freeways, High-Occupancy Vehicle Systems, and Traffic Signal Systems, 106-117.

LAVAL, J.A. & LECLERCQ, L. Mechanism to describe stop-and-go waves: A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Philosophical Transactions of the Royal Society A*, 2010.

LI, X. P., OUYANG, Y. F. & PENG, F. 2010. Measurement and estimation of traffic oscillation properties. Transportation Research Part B-Methodological, 44, 1-14.

MALLAT, S. & HWANG, W. 1992. Singularity detection and processing with wavelets. IEEE Transaction on Information Theory, 38, 617-643.

MALLAT, S. & ZHONG, S. 1992. Characterization of signals from multiscale edges. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 710-732.

MALLAT, S. G. 2009. A wavelet tour of signal processing : the sparse way, Amsterdam ; Boston, Elsevier/Academic Press.

MAUCH, M. & CASSIDY, M. J. 2002. Freeway traffic oscillations: Observations and predictions. Transportation and Traffic Theory in the 21st Century, 653-673.

MUNOZ, J. C. & DAGANZO, C. F. 2003. Structure of the transition zone behind freeway queues. Transportation Science, 37, 312-329.

NGSIM. 2010. Next generation simulation [Online]. Available: http://ngsim-community.org/ [Accessed 08.08 2010].

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

PEMS. 2008. Freeway performance measurement system [Online]. Available: http://pems.eecs.berkeley.edu/ [Accessed 10.08 2008].

PENG, Z. K. & CHU, F. L. 2004. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mechanical Systems and Signal Processing, 18, 199-221.

SAMANT, A., ADELI, H. 2000. Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis. Computer-Aided Civil and Infrastructure Engineering 15 (4), 241–250.

SAMANT, A., ADELI, H., 2001. Enhancing neural network incident detection algorithms using wavelets. Computer-Aided Civil and Infrastructure Engineering 16 (4), 239–245.

TORRENCE, C., COMPO, G. 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society, 79, 61–78.

TREITERER, J. & MYERS, J. A. The hysteresis phenomenon in traffic flow. Proceedings of the 6th International Symposium on Transportation and Traffic Theory, 1974 Sydney. Reed Pty Ltd., 13-38.

VLAHOGIANNI E. I., GEROLIMINIS, N. AND SKABARDONIS, A. 2008. Empirical and Analytical Investigation of Traffic Flow Regimes and Transitions in Signalized Arterials, Journal of Transportation Engineering,134 (12): 512-522.

VLAHOGIANNI, E. I., KARLAFTIS, M. G., GOLIAS, J. C., AND STATHOPOULOS, A. 2007. A wavelet-based neural network approach for identifying transitional behavior in short-term traffic flow, Transportation Research Board 86th Annual Meeting, Washington, D.C.

WASHINGTON, S. & CHENG, W. 2008. New Criteria for Evaluating Methods of Identifying Hot Spots. Transportation Research Record: Journal of the Transportation Research Board, 2083, 76-85.

XIE, Y., ZHANG, Y., YE, Z., 2007. Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition. Computer-Aided Civil and Infrastructure Engineering 22 (5), 326–334.

YAN, R. 2007. Base wavelet selection criteria for non-stationary vibration analysis in bearing health diagnosis. Doctor, University of Massachusetts Amherst.

ZHENG, Z., AHN, S., CHEN, D. J. & LAVAL, J. 2011a. Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations. Transportation Research Part B-Methodological, 45, 372-384.

ZHENG, Z., AHN, S., CHEN, D. J. & LAVAL, J. 2011b. Freeway traffic oscillations: microscopic analysis of formations and propagations using wavelet transform. Transportation Research Part B-Methodological, 45(9), 1378-1388.

ZHENG, Z., AHN, S., CHEN, D. J. & LAVAL, J. 2011c. The impact of lane-changing on the immediate follower: anticipation, relaxation, and behavioral change. Submitted.

ZHENG, Z., AHN, S. & MONSERE, C. M. 2010. Impact of traffic oscillations on freeway crash occurrences. Accident Analysis and Prevention, 42, 626-636.

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

Appendix

A brief description of each wavelet used in this paper is provided in this appendix (Mallat, 2009, Daubechies, 1992).

1. Haar

Haar wavelet is the oldest and the simplest wavelet, as defined in Eq. (A.1).

$$\psi(t) = \begin{cases} 1 & 0 \le t < 0.5 \\ -1 & 0.5 \le t < 1 \\ 0 & otherwise \end{cases} \qquad \text{(A.1)}$$

Haar is orthogonal, symmetric, and not continuous. It is compactly supported with the support size of one. The number of vanishing moments is also one.

2. The Mexican hat

The Mexican hat is the second derivative of the Gaussian probability density function, as defined in Eq. (A.2).

$$\psi(t) = [1 - t^2]e^{-\frac{t^2}{2}} \qquad \text{(A.2)}$$

It is symmetric but not orthogonal with no compact support (its effective support size is [-5 5]) and two vanishing moments.

3. Gaussian family

The Gaussian family is derivatives of the Gaussian probability density function, as defined in Eq. (A.3).

$$gaus(x, n) = Cn * diff(\exp(-x^2), n) \qquad \text{(A.3)}$$

The Gaussian family wavelets are not orthogonal with no compact support (their effective support size is [-5 5]). When n is even, it is symmetric; otherwise, it is anti-symmetric.

4. Daubechies family

The Daubechies wavelet family is well known because it has the highest number of vanishing moments for a given support width. Specifically, for D(n), the vanishing moments is n and the support size is 2n-1.

It is orthogonal and asymmetric. Note that the Daubechies wavelets do not have explicit mathematical expression except for N=1, which is actually the Haar wavelet.

5. Coiflets family

This is a draft version. Please cite this paper as: Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C 25*, 18-33.

They are also constructed by Daubechies for an application in numerical analysis, at the request of Ronald Coifman. They are designed to yield the highest number of vanishing moments for both the base wavelet and the scaling function for a given support width. For Coif(n), the vanishing moments are 2n and the support size is 6n-1.

Coiflets family wavelets are orthogonal and near symmetric.

6. Meyer

The Meyer wavelet is regular orthogonal with no compact support (the effective support size is [-8 8]), as defined in Eq. (A.4).

$$\hat{\psi}(\xi) = \begin{cases} (2\pi)^{-\frac{1}{2}} e^{\frac{i\xi}{2}} \sin\left[\frac{\pi}{2} v\left(\frac{3}{2\pi}|\xi| - 1\right)\right], \frac{2\pi}{3} \le |\xi| \le \frac{4\pi}{3}, \\ (2\pi)^{-\frac{1}{2}} e^{\frac{i\xi}{2}} \cos\left[\frac{\pi}{2} v\left(\frac{3}{4\pi}|\xi| - 1\right)\right], \frac{4\pi}{3} \le |\xi| \le \frac{8\pi}{3}, \\ 0 \quad otherwise, \end{cases} \tag{A.4}$$

Where v is a function satisfying

$$v(x) = \begin{cases} 0, & x \le 0, \\ sin^2 \frac{\pi}{2} x, & 0 \le x \le 1, \\ 1, & x \ge 1. \end{cases}$$

With the additional property

$$v(x) + v(1 - x) = 1$$

7. Morlet

The Morlet wavelet is defined as in Eq. (A.5).

$$\psi(t) = c_\sigma \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2}(e^{i\sigma t} - \kappa_\sigma) \tag{A.5}$$

Where $\kappa_\sigma = e^{-\frac{1}{2}\sigma^2}$; $c_\sigma = (1 + e^{-\sigma^2} - 2e^{-\frac{3}{4}\sigma^2})^{-\frac{1}{2}}$

It is symmetric and non-orthogonal with no compact support (the effective support size is [-4 4]).

8. Symlets family

The symelets family wavelets are similar to the Daubechies wavelets except for better symmetry. They are compactly supported wavelets with least asymmetry and highest number of vanishing moments for a given support size.