Shock Waves in Highway Traffic: Macroscopic and Microscopic Investigation with Wavelet Transform

İsmail ŞAHİN¹
S. Şeyma KUŞAKCI²
Gökçe AYDIN³

ABSTRACT
Shock waves in highway traffic define the boundary conditions between different traffic states. Understanding of formation and propagation of shock waves is helpful for analyzing the features of traffic congestion and its consequences; moreover, for developing novel highway traffic operation techniques and management strategies. Formation and propagation of shock waves were analyzed with wavelet transform technique using two different sets of data, namely macroscopic (mean flow speeds) and microscopic (vehicle trajectories). Estimated propagation speeds of shock waves in this study are consistent with those reported in the literature. Shock wave speeds could be used to calibrate the fundamental diagram of a traffic stream and also to estimate flow speeds (journey times) over a highway stretch.

Keywords: Highway traffic flow, vehicle trajectories, non-stationary time series, shock waves, wavelet transform.

1. INTRODUCTION
Traffic flows constitutes simultaneous movements of individual vehicles. Traffic flow exhibits variations with respect to traffic demand, geometric properties of the highway on which the flow takes place, traffic composition, driver behavior, scheme of traffic management, and weather conditions. The flow, speed and density quantities, which are defined as the derived variables of traffic flow, change with respect to these factors. These derived variables have functional relationships in pairs, hence they affect each other. Any mathematical model based on these bivariate relations has to be compliant with the field observations. Field observations are carried out for both developing new mathematical models to describe the traffic flow and validating the existing models. Traffic flow models are useful tools for highway planning and operation. For this reason, experimental (empirical) studies that support the mathematical models for traffic flow, have an important role in traffic engineering.

1 Yıldız Technical University, İstanbul, Turkey - sahin@yildiz.edu.tr
2 Yıldız Technical University, İstanbul, Turkey - skusakci@yildiz.edu.tr
3 Yıldız Technical University, İstanbul, Turkey - elbe80@gmail.com
† Published in Teknik Dergi Vol. 27, No. 2 April 2016, pp: 7429-7452
At the present time, field measurements in traffic flow can be done either by hand or using various technological devices. Measurements by hand are done when the necessary technological support does not exist or the special conditions related to the research necessitate to do so. Measurement by hand has the flexibility to comply with the existing conditions, but the process of data collection is troublesome. Technological data collection relies on the sensors (detectors). By this method, various flow data can be continuously measured. The measured data can be used for real time operation applications, such as intelligent transportation systems. Alternatively, they can be stored and benefited from for the future transportation planning and traffic flow modeling studies.

The data related to the traffic flow can be collected from the field in the form of time series. Thus, flow, speed and occupancy (unitless form of density) values are measured with respect to time. Measurement intervals are chosen considering the purpose of the study. It can be any interval, for example, from 5 seconds to 60 minutes. By examining the traffic flow data, which is in the form of time series, time-wise or spatial numerical values for traffic flow can be obtained. For instance, the variation of traffic demand at the entrance of a highway, artery or an intersection with respect to time can be measured. The location, instant of activation and capacity of a bottleneck can be determined by using the time series data of traffic flow. The formation, propagation and properties (i.e., changes in amplitude and propagation speed) of kinematic waves within traffic flow can be examined by the help of time series. Time and position data in the trajectories of vehicles that constitute the traffic flow contains prosperous information about the car-following and lane-changing behaviors of drivers. Various methods can be used for extracting the properties of traffic flows from the time series. Statistics and signal processing form the basis of widely used techniques. In [1] and [2], it is shown that, using the scaled curves depicting the change of the cumulative number of arriving vehicles with respect to time, some properties related to the traffic flow can be obtained. In another study [3], curves drawn by computing the 5–min moving averages of the cumulative number of arriving vehicles measured at consecutive stations were used to examine the formation, propagation, growing, and reduction of the oscillations within traffic flow.

This study is about how to extract some properties of traffic flow from time series using the wavelet transform technique. Among the forerunning applications of wavelet transform in traffic engineering, there is the book written by Adeli and Karim, named “Wavelets in Intelligent Transportation Systems” [4]. The authors, in their study, used the wavelet transform as an incident detection means, by combining it with fuzzy logic and artificial neural networks. The applications of wavelet transform related to the examination of bottlenecks in highway traffic, the variations of the flow regime, and the analysis of the oscillations in traffic, is important to show that the technique actually works [5]. Zheng and Washington [6] performed a comprehensive study to choose the most convenient wavelet types to detect the abrupt changes in traffic flow and vehicle trajectory data.

In the second section of this study, the basic properties of non–stationary time series and wavelet transform are explained. In the third section, within the scope of applying the wavelet transform to macroscopic time series data of traffic flow on a highway section and microscopic time series data of vehicle trajectories, the results especially about the formation and propagation of the shock waves are presented. In the fourth section, some considerations about how to use the kinematic waves in the traffic flow models, within the
context of the fundamental diagram of traffic flow, are presented. In the last section, the importance of empirical studies in traffic engineering is mentioned, some general considerations about the applications of the wavelet transform method to the time series of traffic flow are given, and some suggestions for future studies are presented.

2. NON–STATIONARY TIME SERIES AND WAVELET TRANSFORM

Time series of traffic flow (for example, speed–time series) contains noise, especially in high resolution (e.g., measurements in every couple of seconds). When time resolution is low, for example, when the spot speeds of vehicles passing through a highway cross section are measured with 5–minute intervals and an average of the measurements in every interval is calculated, the variation due to different vehicle speeds measured (i.e., noise) is depressed and a unique average value, representing the whole interval, is determined. However, in this case, some information contained within the variation of speeds may be buried into the average and lost. The real–life time series are mostly non–stationary, due to the fact that their statistical properties change with time. The macroscopic quantities of traffic flow, namely, flow, speed and occupancy, and the microscopic quantities obtained from vehicle movements (paths), namely, position, speed and acceleration, exhibit abrupt changes (jumps) in time and thus form non–stationary time series. In recent years, to obtain both the time and frequency properties of such series, the wavelet transform technique has started to be used. In this section, the non-stationary properties of time series and their analysis in time–frequency domain with the wavelet transform technique are explained.

2.1. Non–Stationary Time Series

The derived variables representing traffic flow (volume/flow, speed, density, etc.) are random variables. When they are measured at a definite instant, or their measurements are aggregated with short time intervals, they form a time series. Every measurement is obtained by the occurrence of the traffic variable in a stochastic process. In this context, the stationarity property of a time series representing a process depends on the constancy of its statistical parameters, such as mean and standard deviation, over time [7]. Traffic data generally has the non–stationarity property (like the existence of abrupt jumps in measurements). The variation in traffic demand and bottleneck capacity, merge and diverge sections, car-following and lane-changing behaviors of drivers, stop and go movements due to the oscillations in traffic flows altogether affect the traffic flow and convert it to a non–stationary process [5]. Various techniques have been developed for depressing the non–stationary time series to be stationary (moving average, exponential smoothing, etc.) [8]. However, this can lead to the loss of data/information by suppressing some properties of traffic flow. For this reason, it may be necessary to work with “raw” non–stationary data. In Figure 1, a synthetically produced speed–time series of the traffic flow on a highway cross section is depicted. Within the observation period, it is measured that, there is one abrupt decrease and then an abrupt increase in speed. This type of speed variation is frequently observed in highway sections with bottlenecks. Small circles in the figure represent the average speeds measured in the fixed-length of time intervals. Cumulative mean and standard deviation graphs of the time series are also shown in the figure. Because the cumulative mean and standard deviation of the speed–time series in Figure 1 change with
time, the time process of speed is non–stationary. The cumulative mean and standard deviation of the time series are calculated by Equations (1) and (2), respectively [9]. In the equations, the cumulative values are calculated in incremental manner step by step from the first value of the observation \( n = 1 \) to the last value \( n = N \), given that \( x_i \) is the speed measured at time/interval \( t \) and \( n \) is the sample size. In the following section, the analysis of non–stationary time series in the time–frequency domain by wavelet transform technique is explained.

\[
\bar{x} = \frac{\sum_{t=1}^{n} x_t}{n} \quad \text{for} \quad n = 1, 2, \ldots, N \tag{1}
\]

\[
s = \sqrt{\frac{\sum_{t=1}^{n} (x_t - \bar{x})^2}{n-1}} \quad \text{for} \quad n = 2, 3, \ldots, N \tag{2}
\]

2.2. Wavelet Transform

The way of interpretation of engineers for the signals has been renewed by the Fourier transform. However, by this method, it is impossible to obtain the signal’s properties that
change with time by resolving it into its basic trigonometric functions. Fourier basic functions are frequency based; hence, they lack information about time. This transform technique can provide whether there is an abrupt change in the signal or not, but it cannot provide when this change actually takes place. Although the short-time Fourier transform aims at filling this gap, the resolution obtained is limited [10]. Moreover, Fourier transform is sensitive to the noise in the signal and it does not produce good results in nonlinear problems. Therefore, other methods have been sought after for time–frequency analysis. Thus, the wavelet transform, which can form proper basic functions in both time and frequency domain, and allow to perform analyses in various resolutions, has been developed. With this technique, abrupt changes in time series can be detected together with time of occurrence.

A wavelet, $\psi(t)$, is a real or complex mathematical function. It can convert the continuous time series into various scaling components. The real wavelets, used in this study, satisfy two basic conditions as follows [5]:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0$$  \hspace{1cm} (3)

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$$  \hspace{1cm} (4)

Equation (3) shows that the wavelet function must have zero mean. Equation (4), which represents the wavelet energy, implies that the magnitude of the energy is finite.

Wavelet transform is a linear transform. It separates the signal $x(t)$ into parts by means of elementary functions. These elementary functions are formed by the expansion (in the frequency/scale domain) and translation (in the time domain) of the mother wavelet function, $\psi(t)$. The wavelet transform coefficient (output) of a continuous signal $x(t)$ is called continuous wavelet transform and defined by Equation (5) below:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-b}{a}\right) dt$$  \hspace{1cm} (5)

The scale parameter $a$, which represents the periodic (or harmonic) nature of the signal, is inversely proportional to the frequency, and it controls the dilation or contraction of the wavelet. The parameter $b$, which provides movements of the wavelet in the time domain, is the shift/translation parameter. The $1/\sqrt{a}$ coefficient at the beginning of the equation ensures that the wavelets in all scales have the same energy. When the parameter values are $a = 1$ and $b = 0$, $\psi(t)$ is called the mother wavelet. The wavelet coefficients $W(a, b)$ represent the order of similitude between the dilated/translated mother wavelet (at time $b$ and scale $a$) and the original signal.

There are various wavelet functions within the wavelet family (for example, Haar, Daubechies, Gauss, Morlet, Coiflets, Mexican hat, etc.) Most of these wavelets produce similar results for various types of signal. Zheng and Washington [6] applied some of this
family’s members onto the transform of speed–time traffic flow time series data. The authors selected the wavelet, which detects the time when abrupt changes in speed occur with least amount of error, and whose produced energy contains the highest level of information (in other words, having the smallest Shannon entropy). After the tests performed, they determined that the most convenient wavelet is the Mexican hat. An example of Mexican hat wavelet is depicted in Figure 2. This name is given to this wavelet because its shape is similar to the traditional Mexican hat. The Mexican hat wavelets are used for the transforms in this study.

Equation (6) is the mathematical expression of Mexican hat wavelet. One can see that, mother Mexican hat is the second derivative of \( e^{\frac{t^2}{2}} \) Gauss function (for \( a = 1 \) and \( b = 0 \)) [11].

\[
\psi\left(\frac{t-b}{a}\right) = -a^2 \frac{d^2}{dt^2} e^{\frac{(t-b)^2}{2a^2}} = \left[1 - \left(\frac{t-b}{a}\right)^2\right] e^{\frac{(t-b)^2}{2a^2}}
\]

For example, the wavelet transform coefficient of \( v(t) \) speed–time series, which itself is a continuous signal function, is obtained as in (7):

\[
W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} v(t) \left[1 - \left(\frac{t-b}{a}\right)^2\right] e^{-\frac{(t-b)^2}{2a^2}} dt
\]
Where, the average wavelet-based energy at time $b$ is calculated using the wavelet transform coefficients computed for different scales (8):

$$E_b = \frac{1}{\text{max}(a)} \int |W(a,b)|^2 \, da$$

Where, $\text{max}(a)$ is the highest scale chosen. This highest scale is chosen as the highest value where the general appearance of the energy is maintained without distortion. Translation value $b$ is determined such that, depending on the time resolution of the original signal, the energy is computed in all time steps. Energy is a unitless measure. Any abrupt change in the speed signal within time causes an abrupt increase in temporal distribution of the wavelet-based energy calculated by Equation (8). Thus, by tracing the energy distribution, start and end of the traffic queues, and abrupt speed changes due the passage of oscillation waves can be detected [12]. A general representation of the transform operation with Mexican hat wavelet is depicted in Figure 3, which was inspired by reference [13].

Figure 3. Mexican hat wavelet transform and obtained maximum coefficients graphic
3. WAVELET TRANSFORM APPLICATIONS

3.1. Shock Waves Forming at Highway Bottlenecks – Macroscopic Scale

The studied stretch and data

Within the scope of this study, the data collected by the sensors (detectors) at the Asia–Europe direction of Fatih Sultan Mehmet Bridge crossing the Bosphorus strait and at the approaching sections of TEM (Trans European Motorway) passage through Istanbul urban area, are examined using the wavelet transform method. These macroscopic scale data of traffic flow have been used in a M.Sc. thesis prepared in the Yıldız Technical University. In this thesis, the efficiency of lane-borrowing from the opposite direction in the evening peak hours has been examined for various implementation scenarios [14]. In this implementation, the leftmost lane in Asia–Europe direction is borrowed and allocated to the Europe–Asia direction, to meet the peak traffic demand. The data used were collected on February 21, 2008 at the evening peak hours between approximately 16:10 and 21:10, in the direction from Asia to Europe. The sky was clear and there was no precipitation on that day. The sensors (devices for data collection) are the RTMS (Remote Traffic Microwave Sensor) radar detectors operated by the Istanbul Metropolitan Municipality. These detectors collect the average speed and total traffic volume data for all travel lanes at the highway cross section under investigation, with 2-minute resolution. A sketch of the studied stretch of the highway, together with RTMS numbers and locations (km), is shown in Figure 4. In the sketch, the first RTMS 5 in the downstream direction is assumed to be the starting point, and the calculated RTMS km values are written within the parentheses accordingly.

![Figure 4. Studied stretch and RTMS locations](image)

This lane-borrowing implementation provides additional capacity for the Europe–Asia direction in the evening peak hours. However, as a result of this regulation, long vehicle queues in the Asia–Europe direction are formed due to the reduction in the number of lanes. As shown in the application below, queuing effect is first sensed by RTMS 92, located just upstream of the entrance of the bridge. Then, the interface between free flow condition and congestion condition propagates in the upstream direction (opposite to traffic flow) as a shock wave. When a shock wave (interface) arrives at a cross section, the flow speed at this cross section drops abruptly. The times of those abrupt speed changes on the highway cross sections, where RTMS’s are located, can be detected by the wavelet transform.

For the application of wavelet transform, a wavelet transform toolbox of a mathematical package (Wavelet Toolbox) is used. The speed–time series recorded by the RTMS is introduced as a signal function to this package. The package automatically performs the
wavelet transform for the specific type of wavelet (Mexican hat in this study) defined by the user, using the largest scale ($a_{max}$) again specified by the user; then it reports the wavelet coefficients matrix and local maximum coefficient curves. The screenshot of the wavelet transform of speeds measured by RTMS 63 is shown in Figure 5. There are three graphics in this screenshot. The uppermost graphic shows the speed–time series signal function introduced to the package, the intermediate graphic shows the contour curves formed by using the coefficient matrix (lighter areas indicate maximum coefficients), and the undermost graphic shows the local maximum coefficient lines.

Figure 5. Screenshot of the wavelet transform toolbox for the data of RTMS 63

**Wavelet transform application: macroscopic scale**

The length of the speed–time series used in this study is 300 minutes. Since they are collected with 2-minute intervals, there are approximately 150 data points in the time series (Figure 6a). In every interval, an average of spot speeds measured for the vehicles passing through the radar station in the interval is calculated to represent the speed of the interval. It was realized in our early trials with the wavelet transform that a large number of data points are required to produce good transform results. For this reason, the number of data points in the speed–time series has artificially been increased, in such a way that the general properties of the data is not harmed. It was decided that a 30–second resolution produced a
Figure 6. (a) Raw speed data measured by RTMS 4; (b) Speed data with 30-second resolution derived from raw speed data; (c) 30-second resolution speed data with noise added.

sufficient amount of data points. For this data manipulation, the flow speed is assumed to be constant in the 2-two-minute intervals and every interval is divided into four 30–second intervals. Thus, every two-minute interval is converted into four data intervals having the same speed (Figure 6b). When the original speed data is aggregated by taking the average
in 2–minute intervals, the series containing noise in reality is straightened. To make the straightened data more realistic, noise with a SNR (signal–to–noise ratio) of 25 is added to the data set with increased resolution (Figure 6c). Indeed, for the wavelet transform technique, increasing the resolution is much more effective than adding noise. Moreover, there is barely any difference between the transforms with or without noise. However, we increased the resolution of and added noise to the speed-time series data used in this section of the study.

In Figure 7, the speed–time series graphs (signals) belonging to the six RTMS’s examined are shown at the top and the local maximum coefficient lines obtained from the wavelet transform are shown at the bottom. There are no abrupt speed changes detected in the most downstream RTMS 153 and in the most upstream RTMS 5 stations where measured speeds remain in the order of 100 km/h (see the uppermost and undermost graphics in Figure 7). Wavelet transform coefficient graphs belonging to those radar stations have only two maximum coefficient lines, emerging from the bottom corners of the graphs (two ends of the signal). This is known as “boundary effect” in the literature. The signal introduced is of finite length and, values outside the boundaries are simply counted as zero. For this reason, at the endpoints of the signal, there is a jump from zero to the actual value of the signal. This leads to big wavelet coefficients obtained at these points. The points where the maximum coefficient lines intersect with the horizontal time axis (at \(a = 1\) scale) is either the instant that an abrupt speed change starts or the instant that speed becomes stationary (stable) again after the abrupt change [6]. These times mark the transition between free–flow and forced–flow states of traffic. Inspecting speed–time series graphs from the other RTMS stations indicate that the local maximum points at the smallest scale coincide speed change instants quite accurately. This information extracted from the data is valuable for traffic engineering and an important input for the decision-making process in highway operation. This result shows that the wavelet transform can be used for catching the formation of shock waves or interfaces between different flow states, and for calculating the propagation speeds of them.

The times of speed changes in the RTMS stations determined by the wavelet transform (i.e., the times when speed first starts to drop and when speed becomes almost constant) are given in Table 1. The time when speed first starts to drop and subsequently the time when speed becomes almost constant are obtained from the wavelet transform. A scatter plot in which the vertical axis is the location of the radar stations in the downstream direction (assuming that RTMS 5 is the first) and the horizontal axis is the passage times of the shock waves through the stations is drawn. Hence, the slope of the trend (regression) lines drawn through the data points is the average speed of the shock waves (kinematic waves). The corresponding scatter plots are depicted in Figure 8. In the figure, the filled circles (●) represent the points where speed first starts to drop at the respective radar station and the filled triangles (▲) represent the points where speed becomes almost constant. Both trend lines have high coefficient of determination (R²) values as shown on the plot. The slope of the lines is equal to the average speed of the shock waves (A negative speed indicates that the shock wave propagates in the upstream direction against traffic flow). The estimated shock wave propagation speeds for the instants when the speed first starts to drop and for the instants when the speed becomes almost constant are –11.61 and –10.73 km/h, respectively. These speeds are consistent with those found in [5] (–11.78 and –11.71 km/h,
Shock Waves in Highway Traffic: Macroscopic and Microscopic Investigation...

respectively). The shock waves formed in traffic flow can affect the macroscopic quantities of the flow, and the shock wave propagation speed itself is also a macroscopic quantity.

Figure 7. Speed–time series and maximum coefficient lines for RTMS's examined

**NOTE:** In the speed–time series signal (located at the top), the vertical axis is speed (km/h) and the horizontal axis is time (t x 30 s); in the maximum coefficient graph (located at the bottom), the vertical axis is scale (a) and the horizontal axis is time (t x 30 s).
Figure 7. (cont'd) Speed–time series and maximum coefficient lines for RTMS’s examined

NOTE: In the speed–time series signal (located at the top), the vertical axis is speed (km/h) and the horizontal axis is time (t x 30 s); in the maximum coefficient graph (located at the bottom), the vertical axis is scale (a) and the horizontal axis is time (t x 30 s).
Table 1. Times of speed changes in the RTMS stations

<table>
<thead>
<tr>
<th>RTMS No</th>
<th>Time when speed first starts to drop</th>
<th>Time when speed is almost constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>153</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>92</td>
<td>17:21:15</td>
<td>17:27:15</td>
</tr>
<tr>
<td>61</td>
<td>17:44:03</td>
<td>17:46:03</td>
</tr>
<tr>
<td>63</td>
<td>17:49:23</td>
<td>17:57:22</td>
</tr>
<tr>
<td>95</td>
<td>18:04:10</td>
<td>18:16:10</td>
</tr>
<tr>
<td>73</td>
<td>18:05:10</td>
<td>18:19:10</td>
</tr>
<tr>
<td>4</td>
<td>18:20:36</td>
<td>18:24:36</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 8. Scatter plot of locations of RTMS’s versus times of speed changes, and trend lines

3.2. Oscillations in Traffic Flow and Shock Waves – Microscopic Scale

Oscillations in traffic flow occur due to recurrence of deceleration and then acceleration processes of the flow. By using the macroscopic properties of the flow, the properties of the oscillations in traffic (for example formation and propagation) can be examined [3]. Oscillations can also be investigated, in macroscopic scale, by the help of vehicle trajectories. Thanks to the advances in image processing techniques, some high resolution vehicle trajectory data have recently been put into service for researchers [15]. In this section, the wavelet transforms of position–time data obtained from the vehicle trajectories are carried out. Thus, the speed changing times of vehicles, the reasons of the formation of shock waves and the propagation speeds of the waves are examined.
The studied stretch and data

Vehicle trajectory data used in this section are put into use of researchers through the internet by the USA’s Federal Highway Administration, within the scope of a program called NGSIM (Next Generation Simulation) [15]. The vehicle trajectory data for this program is collected from the US–101 highway located in Los Angeles, California. The sketch of a 640 m section in the southern direction is given in Figure 9. The resolution for the trajectory data is 10 records per second and they are collected between 07:50–08:35 on June 15, 2005.

By means of the distribution of the wavelet energy in vehicle trajectory with the changes of vehicle speed, the trigger and propagation of deceleration (acceleration) waves can be detected [5]. In this section, it is shown that, using the position data of vehicle trajectories, instead of speed as in the literature, performing similar analyses are possible. Any abrupt change in vehicle position trend (amount of displacement or distance traveled with respect to a reference start point) produces an abrupt energy jump in the wavelet frame. Indeed, the slope of the tangent drawn to the trajectory curve at any point on it gives the spot speed of the vehicle at that particular time and position. Any change in vehicle speed, which can be considered as abrupt, is the main source of this energy jump. By tracking the energy jumps of the consecutive vehicles (related to their positions), the vehicle that initiated the deceleration/acceleration wave and the propagation speed of this wave can be determined.

In Figure 10, trajectories (space–time plots) of vehicles on the leftmost lane 1 of US–101 highway are depicted.

If the position–time series is introduced to the wavelet toolbox mentioned above as a signal function, the package automatically performs the wavelet transform for the wavelet type specified by the user (Mexican hat in this study) and for the largest scale specified by the user ($a_{\text{max}}$). (Since the resolution of NGSIM vehicle trajectory dataset is 0.1 s, $a_{\text{max}} = 32$ scale corresponds to a value of 3.2 seconds in the time domain and the wavelet window does not contain more than one acceleration–deceleration.) Using the temporal distribution of the wavelet-based energy of vehicle trajectories as in Equation (8), the sudden slope changes (i.e., instantaneous speeds) in vehicle trajectories can be determined. A temporal wavelet-based energy distribution of vehicle trajectory 539 is shown in Figure 11. This trajectory represents the position–time series. It can be seen that there is a correspondence between the hill points of the energy graph and the points where the slope of the trajectory changes. Thus, the hill (maximum) points of the temporal wavelet energy distribution mark the oscillations in vehicle trajectory (deceleration or stopping and acceleration or starting).
Based on this analysis, the locations of formations and the reasons of oscillations (or the kinematic waves/shock waves) together with their propagation speeds in traffic flow can be detected.

Figure 10. Trajectories of a group of vehicles in NGSIM records (US-101, lane 1)

Figure 11. Trajectory of vehicle 539 in the NGSIM records (the monotonically increasing curve with some oscillations) and the wavelet-based energy distribution (waving curve) (US-101, lane 1)
Wavelet transform application: microscopic Scale

If deceleration of a vehicle is represented by an abrupt decrease in the slope of its trajectory, this sudden trend change in trajectory produces a jump in energy in the wavelet window. Using the wavelet-based energy, acceleration/deceleration waves in traffic can be detected. It is possible to find the vehicle causing the energy jump, in other words, the source of the deceleration wave, by tracing the energy hills. The acceleration waves can also be traced similarly.

The trajectories of 22 vehicles are depicted in Figure 10. For the position–time series of these trajectories, the wavelet transform is first performed and then using the wavelet transform coefficients obtained, the temporal wavelet-based energy distribution graphs are drawn as in Figure 12. Each vehicle is represented by an energy graph and the graphs are displaced vertically by adding some constant values to them. (Since the trajectory lengths of vehicles 505 and 530 are too short to produce meaningful energy distributions, their energy graphs are depicted as short horizontal lines, just to show their positions in the graph.)

![Figure 12. Curves of temporal distributions of wavelet-based energy for each trajectory](image)

Using the wavelet energy distribution curves of vehicle trajectories in Figure 12, time and location values of the hill points (speed change points) are recorded. These records are then marked on the vehicle trajectories as shown in Figure 13. The points marked on the consecutive vehicles are joint successively to obtain trajectories of the acceleration and deceleration waves (shown by dashed arrows on the figure). In the time period of concern,
traffic is flowing under the forced flow conditions. Hence, the waves propagate in the upstream direction (opposite to traffic flow).

Various sources can cause the kinematic waves (shock waves) to form. Among the main reasons for occurrence of the shock waves in forced flow conditions are the driver behaviors in car-following and lane-changing. However, because of some other factors, kinematic waves can also form instantaneously [12]. The examination of vehicles trajectories under investigation and the wavelet-based energy curves reveals that the kinematic waves observed form due to similar triggering behaviors. In the following paragraph, the vehicle trajectories in Figure 14 (shown within the inset box) are examined, together with the emerging acceleration and deceleration waves.

Two examples of kinematic waves determined (one acceleration and one deceleration) are presented in Figure 14. (The vehicle trajectories are shown in the inset box at the bottom right part of the wavelet-based energy distribution graphs of the trajectories). The acceleration wave starts with the acceleration of vehicle 425. Vehicle 450 responds to this wave with some time lag, but the following vehicle 457 and the subsequent vehicles accommodate to this wave and increase their speeds. Delay in acceleration of vehicle 450 caused the distance between itself and the preceding vehicle 430 to increase. Vehicle 457 started acceleration (probably) by observing vehicle 430. It approached too close to vehicle 450, which was late accelerating; and soon after, vehicle 457 left the lane (the trajectory of vehicle 457 ended after the acceleration wave). It is found out that the main factor for this acceleration wave is the car-following mechanism. In the deceleration wave, the primary
role is again played by vehicle 450. The late response of vehicle 450 to the acceleration wave created a long gap downstream on lane 1 where it is traveling. This caused vehicles 437 and 451 to change their lanes and to enter lane 1, in front of vehicle 450. Vehicle 450 again responded late, this time to the deceleration of vehicle 451 but eventually slowed down. The driver of vehicle 465, which follows immediately vehicle 450, realized that vehicle 451 slowed down, and responds early to this maneuver by slowing down and increases its following distance slightly to vehicle 450. The deceleration wave transmitted from vehicle 451 to vehicle 465, continued to propagate in the upstream direction (opposite to traffic). The main factor triggering this deceleration wave is the lane–changing behavior of vehicles 437 and 451, causing the increase in vehicle density on lane 1.

![Figure 14. Acceleration and deceleration shock wave on vehicle trajectories and energy distribution curves](image)

The propagation speeds of acceleration and deceleration waves can be estimated by the regression using the relationship between the temporal distribution of wavelet-based energy and vehicle trajectories. The trajectories of the shock waves shown in Figure 13 are plotted in location–time domain and a regression analysis performed through the successive speed change points is shown in Figure 15. Linear regression equations determined for the shock wave trajectories and coefficients of determination ($R^2$), which show the goodness of fit of the regression line with data, are also shown on the figure. The calculated coefficients of determination vary between 0.7143 and 0.9850, indicating, in general, that there is a good fit of the regression line to the data. The slopes of the regression lines are the speeds of the kinematic waves/shock waves. Based on this study, the speeds of the shock waves

1991
Shock Waves in Highway Traffic: Macroscopic and Microscopic Investigation...

propagating in the upstream direction against traffic are between 2.72 m/s (9.79 km/h) and 6.68 m/s (24.05 km/h). The average propagation speed is 4.48 m/s (16.13 km/h). In similar studies in the literature, average propagation speeds are 18.3 km/h [16], between 14.07 km/h and 16.96 km/h [5], 12.87 km/h and 18.67 km/h [12]. The propagation speeds are in similar order; however, examination of the variations in propagation speeds is the subject of ongoing research.

Figure 15. Regression analysis through the points of temporal distribution of wavelet-based energy and propagation speeds of kinematic waves/shock waves

4. SIMPLIFIED THEORY OF TRAFFIC FLOW AND SHOCK WAVES

The shock wave theory holds an important place in modeling of traffic flows. Simplified theory of traffic flow was proposed by Prof. Newell [17]. This theory is a simplified version of the original theory, proposed in 1955–56 and named as LWR (Lighthill and Whitham, 1955; Richards, 1956). The original theory assumes that traffic is a continuous medium. That is, vehicle movements in traffic can altogether be represented by models defined by the average behavior. Newell described his model as simplified partly because of the relationship he assumes to exist between the magnitude of traffic flow and travel time [18]. This relationship can vary along the highway but does not vary with time. These relationships are fully empirical; they are obtained from the field measurements. They are dependent mainly on highway geometry; however, they also have some dependency on environmental factors such as weather conditions and properties of traffic flow (i.e., ratio of heavy vehicles and behavior of their drivers).
For the continuous flow models, bivariate relationships defining the boundary conditions between various traffic variables can be used. However, the most widely used one is the relation between flow \((q)\) and density \((k)\). The simplified theory assumes that the relation between flow and density, known as the fundamental diagram, is in triangular form as shown in Figure 16. The left part of this relation represents the free-flow traffic and the right part represents the congested/queued traffic. According to the fundamental equation of traffic flow, speed \((u)\) is obtained by dividing flow by density \((q/k)\). For this reason, the simplified theory assigns the value of free-flow speed \((uf)\) to the left part as its slope. The positive constant slope on this side implies that changing \((q, k)\) states propagate in the downstream direction with vehicles in traffic. In the right part of the bivariate relation, flow decreases with increasing density. Propagation speed of the interfaces between stationary traffic states is given by the ratio of difference in flow to difference in density \(\left[(q_2 - q_1)/(k_2 - k_1)\right] = (\Delta q/\Delta k)\). It is assumed that the right part of the relation is also linear. That is, the interface between different congested traffic states propagates with a unique (constant) speed \((w)\). Based on these assumptions, it is possible to determine whether the traffic flow is in free-flow or forced-flow condition, depending on the direction of the propagation of kinematic waves. This triangular fundamental diagram, after being calibrated for capacity \((q_{\text{max}})\), jam density \((k_j)\), free flow speed \((uf)\), and propagation speed of the shock waves \((w)\) of the section under investigation, can be used, for example, to estimate traffic flow speeds (dashed arrow no. 2 in Figure 16) and travel times accordingly given the value of traffic flow (dashed arrow no. 1 in Figure 16).

![Figure 16. Triangular flow–density relationship](image)

5. DISCUSSIONS AND CONCLUSIONS

The applications of various data processing techniques are increasing together with the advances of data collection technologies and their widespread usage. The set of “big data” collected from the field using these technologies for developing and validating the models for highway planning and operation provides also unequalled opportunities for the development and validation of different traffic management strategies. The big data opportunity provided by technology, however, is bounded with the properties of data
collected (depending on their accuracy and level of detail). RTMS data used in this study have 2-minute resolution and NGSIM vehicle trajectories have 0.1-second resolution. Because of the limits of today’s radar technology, the measurement interval can only be shortened up to a certain length of time. Increased data resolution means increased volume of data and thus increased storage cost of data; however, this provides the opportunity to perform finer/more accurate analyses. Today’s loop detectors can provide macroscopic traffic data with intervals as short as 20 s. When investing in data collection technologies, a trade-off between quality and cost should be considered.

High resolution (for example 0.1 s) vehicle trajectory data, collected by image processing techniques for researchers, provide unequalled research opportunities for highway/traffic engineering studies. Today, other techniques such as Global Positioning Systems (GPS), can collect high resolution data related to vehicle movements (e.g., time, position, speed, etc.) The high resolution data is required to increase the functionality of Intelligent Transportation System (ITS) components. The performance of highway traffic planning and operation processes has the potential to increase with the collection and analysis of such data.

Various data analysis techniques have been adopted and used in the highway traffic engineering discipline. These techniques have advantages and disadvantages compared to each other. Enlarging the range of techniques used can provide new opportunities in traffic planning and operation. The wavelet transform technique used in this study is among the analysis techniques emerging in traffic engineering literature in recent years. It provides ease of application because of its mathematical foundations and usage flexibility. The wavelet transform technique, used for examination of formation and propagation properties of highway traffic shock waves (in both macroscopic and microscopic scales) in this study, provides promising results for the future. This technique facilitates the examination of bottlenecks in both recurrent (peak hour) and non-recurrent (incident) congestion conditions. Thus, it has the potential to support the efforts to increase the safety and efficiency in highway traffic management and operation.

With wavelet transform method, oscillations within traffic can be detected with consistent degrees. However, some partial variability in empirical results shows that the technique still needs some improvements. As a matter of traffic’s nature, this variability can also stem from the differences in driver behaviors. The oscillations in traffic flow cause problems with travel safety and resource usage efficiency. Studies for making the traffic flow steady will continue in the direction of developing both the analysis techniques and the models and algorithms. Estimation of traffic flow speeds and travel times based on the kinematic waves are also among the potential research topics.

Acknowledgements

We are grateful to İstanbul Metropolitan Municipality (İBB), Department of Transportation and especially Mr. Batuhan ALTUN, M.Sc. for their help providing the RTMS radar data. We are also grateful to the U.S. Ministry of Transportation, Federal Highway Administration (FHWA) for preparing the NGSIM (Next Generation Simulation) highway traffic data for public usage.
References


Shock Waves in Highway Traffic: Macroscopic and Microscopic Investigation...


