Accepted Manuscript

Title: Monitoring road traffic congestion using a macroscopic traffic model and a statistical monitoring scheme

Author: Abdelhafid Zeroual Fouzi Harrou Ying Sun Nadhir Messai

PII: S2210-6707(17)30281-0
DOI: <http://dx.doi.org/doi:10.1016/j.scs.2017.08.018>
Reference: SCS 737

To appear in:

Received date: 17-3-2017
Revised date: 8-8-2017

Please cite this article as: Abdelhafid Zeroual, Fouzi Harrou, Ying Sun, Nadhir Messai, Monitoring road traffic congestion using a macroscopic traffic model and a statistical monitoring scheme, *Sustainable Cities and Society* (2017), <http://dx.doi.org/10.1016/j.scs.2017.08.018>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Monitoring Road Traffic Congestion Using a Macroscopic Traffic Model and a Statistical Monitoring Scheme

Abdelhafid Zeroual^{a,b}, Fouzi Harrou^c, Ying Sun^c, Nadhir Messai^b

^aLAIG Laboratory, University of 08 May 1945, Guelma 24000, Algeria

^bCRESTiC-URCA UFR SEN, University of Reims Champagne-Ardenne, Moulin de la Housse, BP 1039 51687 Reims Cede x 2, France

^cKing Abdullah University of Science and Technology (KAUST)

Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia

E-mail: fouzi.harrou@kaust.edu.sa

Abstract

Monitoring vehicle traffic flow plays a central role in enhancing traffic management, transportation safety and cost savings. In this paper, we propose an innovative approach for detection of traffic congestion. Specifically, we combine the flexibility and simplicity of a piecewise switched linear (PWSL) macroscopic traffic model and the greater capacity of the exponentially-weighted moving average (EWMA) monitoring chart. Macroscopic models, which have few, easily calibrated parameters, are employed to describe a free traffic flow at the macroscopic level. Then, we apply the EWMA monitoring chart to the uncorrelated residuals obtained from the constructed PWSL model to detect congested situations. In this strategy, wavelet-based multiscale filtering of data has been used before the application of the EWMA scheme to improve further the robustness of this method to measurement noise and reduce the false alarms due to modeling errors. The performance of the PWSL-EWMA approach is successfully tested on traffic data from the three lane highway portion of the Interstate 210 (I-210) highway of the west of California and the four lane highway portion of the State Route 60 (SR60) highway from the east of California, provided by the Caltrans Performance Measurement System (PeMS). Results show the ability of the PWSL-EWMA approach to monitor vehicle traffic, confirming the promising application of this statistical tool to the supervision of traffic flow congestion.

Keywords: Traffic congestion, Macroscopic traffic model, Statistical monitoring, Quality control chart.

1. Introduction

The management of vehicle traffic congestion in public road networks is becoming a key factor for economic growth. Indeed, road traffic congestion hampers economic growth: it has a profound impact on travel time, causing delays that result in the significant financial loss of billions of dollars spent on extra hours of travel and fuel consumption (Barria and Thajchayapong, 2011; Williams and Guin, 2007). Automotive industry growth, which has improved the quality of vehicles and significantly increased their number, has also increased the environmental impact and traffic congestion in many parts of the world. It has been reported in (Muneer et al., 2011) that the automobile sector

globally consumes about 18 million barrels of oil per day. To give an indication of the impact of traffic congestion, a survey carried out by Texas A&M Transportation Institute (TTI) showed that traffic congestion in the United States accounted for over \$120 billion in fuel and time expenditures in 2013 (Chen et al., 2014). By 2020, traffic congestion is predicted to increase these costs to \$199 billion (Chen et al., 2014). Traffic congestion, which can result in traffic accidents, lost time, hampered mobility, traffic delays and even serious safety problems, is often difficult to avoid. Therefore, traffic flow monitoring is imperative to improve traffic management, transportation safety and cost savings. Of course, the reliable detection of traffic congestion also provides useful information that can help people take the necessary precautions to avoid crowded roads, thus reducing travel time, fuel consumption and pollution. In this paper, the focus is on the statistical detection of traffic congestion.

A good description of traffic dynamics is necessary for the detection of traffic congestion. Increased attention to traffic flow monitoring has led to the development of several mathematical modeling approaches for describing traffic flow, which can be grouped into two main classes: microscopic and macroscopic (Burghout et al., 2006). Microscopic approaches focus on the behavior of the vehicle-conductor pair and its interaction with neighboring vehicles. However, in such approaches, a large number of parameters are required to suitably describe the traffic dynamics, leading to a high computational cost. On the other hand, macroscopic modeling approaches are based on a hydrodynamic theory that captures traffic dynamics as a continuous flow. The main advantage of macroscopic approaches is that less parameters are used than in microscopic models, resulting in lower computational and time costs (Gning et al., 2011). Furthermore, the macroscopic approach derives a mathematical model in closed form to model the traffic variation. Macroscopic models have attracted the attention of engineers and practitioners because of their mathematical and systematic characteristics. Hence, macroscopic modeling approaches are adopted in this paper; we focus particularly on the cell transmission model (CTM), which is a discretized version of the first order macroscopic model (Daganzo, 1994).

Over the last two decades, several extensions of CTM modeling have been developed to meet various requirements in practical use. Daganzo (1999) developed the lagged CTM by introducing lags into the densities, which makes the CTM suitable for inhomogeneous highways and intersections. Muralidharan et al. (2009) proposed the link-node CTM to deal with highways that include subnetworks, in which the road segments are represented by links and the network junctions, by nodes. Chen et al. (2010) developed the location-specific CTM to describe some complex traffic phenomena, by defining various shapes of fundamental diagrams (FD) for each cell and taking into account some specific variables, like sensor location and geometric features. Muñoz et al. (2006) adapted a switched-mode model (SMM) to bypass the problem of nonlinearity in the CTM by using of a triangular FD form and switching between five sets of differential equations. Lemarchand et al. (2010) proposed a robust model that extended the SMM model

to deal with the uncertain speed of congestion waves for highway traffic density estimation. Morbidi et al. (2014) extended the CTM to take into account parametric uncertainties. Here, the piecewise switched linear (PWSL) model, which is an enhanced version of SMM (Zeroual et al., 2017), will be used as the modeling framework. This model is suitable for modeling hybrid systems, and is able to cover all possible traffic situations respecting the propagation of congestion waves. Moreover, the main advantage of the PWSL model is its ability to be modular for several configurations of road partitioning and junction types.

Due to the increasing need for traffic monitoring, several techniques aimed at detecting road traffic congestion have been proposed, including the multilayer feed-forward neural network (Chang, 1992), the Markov model and back propagation neural network (Yu et al., 2016), the probability neural network, the constructive probability neural network (Ritchie and Abdulhai, 1997; Jin et al., 2001) and the support vector machine (Cheu et al., 2003). The main drawback of these machine learning methods is that they require long computational training times, which make them less attractive for online applications. Several vision-based approaches have been proposed in literature for traffic congestion monitoring (Ramalingam and Varsani, 2016; Mehboob et al., 2016; Yu and Zuo, 2015; Pongpaibool et al., 2007). Pongpaibool et al. (2007) proposed an approach for a vision-based road-traffic evaluation system using manually tuned fuzzy logic and adaptive neuro-fuzzy techniques. Also, in (Mehboob et al., 2016) a fuzzy logic-based traffic event detection procedure using road surveillance videos is proposed. Other researchers focused in using spatiotemporal analysis of GPS data to detect road traffic congestions and incidents (D'Andrea and Marcelloni, 2016; Kuang et al., 2015; Yong-chuan et al., 2011).

While several anomaly detection techniques have been proposed for traffic monitoring (D'Andrea and Marcelloni, 2016; Kinoshita et al., 2015), statistical monitoring charts have not been widely used for monitoring in traffic congestion until recently. Kuang et al. (2015) proposed to use wavelet transform and principal component analysis for detecting anomalous traffic events in urban regions based on taxi GPS data. This paper presents an efficient scheme for detecting traffic congestion based on statistical monitoring charts. Specifically, we propose a statistical strategy that combines the flexibility and simplicity of a PWSL model with the greater ability to detect incipient changes of the exponentially-weighted moving average (EWMA) chart. In this framework, the PWSL model is used to capture traffic dynamics and describe traffic flow evolution. The residuals obtained from the PWSL model are then used as the input data for the EWMA chart. Note that the main advantage of EWMA chart is that it can be easily implemented in real time because of the low computational cost, which is not the case in classifiers based methods (the classifier algorithms are performed offline rather than online). A decision can be made for each new sample by comparing the value of the EWMA decision statistic with the value of the threshold. A congestion is declared if the EWMA statistic exceeds the threshold. However, the presence of measurement noise and modeling errors increase the rate of

false alarms. Therefore, to further improve the quality of congestion detection, wavelet multiscale filtering is used to filter the residuals obtained from the PWSL model, which helps decrease the false alarm rate. We apply the proposed PWSL-EWMA approach to real traffic data from a section of the Interstate 210 (I-210) freeway in California and a portion of the SR60 highway which is a four-lane highway in east of California, provided by the Caltrans Performance Measurement System (PeMS). Results show the ability of the new PWSL-EWMA scheme to detect road traffic congestion.

The following section briefly reviews macroscopic traffic modeling and the improved piecewise macroscopic traffic model. Section 3 introduces the EWMA chart and its use in anomaly detection. Section 5 reviews the PWSL-based approach and how it can be merged with the EWMA chart for traffic congestion detection. In Section 6, the performances of the proposed methods are illustrated in a practical application. Finally, Section 8 concludes the paper.

2. PWSL macroscopic traffic model

2.1. Macroscopic traffic modeling

The macroscopic traffic approach is one of the most reputed modeling approaches in traffic flow modeling. The main benefit of a macroscopic traffic model is its capability to describe a traffic flow dynamic based on the hydrodynamic theory of modeling. Generally, three macroscopic models can be identified: first order (Lighthill and Whitham, 1955; Richards, 1956), second order (Payne, 1971; Whitham, 1974) and higher order macroscopic models (Aw and Rascle, 2000; Ross, 1988; Zhang, 1998). Daganzo (1995) demonstrated the limitations of the second and higher order models, which appear mainly as negative values of the estimated traffic speed under some conditions. However, the first order macroscopic model, or the LWR model (Lighthill-Whitham-Richards model), has been widely used in different traffic applications (Muralidharan et al., 2009; Morbidi et al., 2014). It is one of the most used traffic models in the literature. The LWR model describes traffic flow as a set of partial differential equations (PDE), given in Equations (1), (2) and (3), which aggregate the various macroscopic traffic parameters, such as the density, the flow and the flow speed:

$$\frac{\partial \rho_i(x, t)}{\partial t} + \frac{\partial q_i(x, t)}{\partial x} = 0, \quad (1)$$

$$q_i(x, t) = \rho_i(x, t) v_i, \quad (2)$$

$$q_i(x, t) = f(\rho_i(x, t)), \quad (3)$$

where, ρ_i is the density in a given road segment (the number of vehicles per unit of length), q_i is the flow (the number of vehicles entering the road per unit of time), v_i is the flow speed, x denotes the spatial measurement of the considered

link, t denotes the time and f is the Fundamental Diagram (FD) (Kerner, 2011) (Figure 1).

Figure 1: Schematic of the triangular form of the fundamental diagram.

The FD, a nonlinear, is an empirical curve that depicts the flow-density relationship evolution. In Figure 1, $\rho_{c,i}$ is the critical density, at which the flow takes its maximum value, $Q_{M,i}$ is the maximal flow, $\rho_{j,i}$ is the maximal density or traffic jam density, w_i is the wave speed congestion and $v_{f,i}$ is the free-flow speed.

The CTM (Daganzo, 1994) is an extended version of the LWR model that uses the Godunov schema (Godunov, 1959) method, which is introduced to solve the LWR equations. Furthermore, it is a simplified version that discretizes the LWR into space and temporal area by taking in consideration the unidirectionality of roads. In the CTM, the road is divided into segments or cells of equal length, Δx_i , and number of lanes; see Figure 2. The occupancy of vehicles in the CTM evolves through the following differential equation:

$$n_i(k+1) = n_i(k) + \Delta T (q_i(k) - q_{i+1}(k)), \quad (4)$$

where $n_i(k)$ is the occupancy of cell i (the number of vehicles in Δx_i at a time k), $q_i(k)$ is the flow (the number of vehicles moving from cell $i-1$ to cell i at an interval time $[k, k+1]$) and ΔT is the step time. The flow of the cell i is given by the following nonlinear equation (Daganzo, 1994):

$$q_i(k) = \min \left(n_{i-1}(k), Q_{M,i}, \frac{w_i}{v_{f,i}} (n_{j,i} - n_i(k)) \right), \quad (5)$$

where $Q_{M,i}$ is the maximal flow that can travel from the upstream cell to the downstream cell, w_i is the wave congestion speed, $v_{f,i}$ is the free-flow speed, $n_{j,i}$ is the capacity of the cell i , and $n_{i-1}(k)$ is the number of vehicles that can leave the cell $i-1$. Note that the term $\frac{w_i}{v_{f,i}} (n_{j,i} - n_i(k))$ corresponds to the number of vehicles that can be received by the cell i .

Figure 2: Spatial road segment partitions.

2.2. PWSL model

Based on the CTM, a new version has been proposed in (Zeroual et al., 2017). This modified version is the piecewise switched linear (PWSL) model, which has the nature of piecewise linear hybrid systems. Hence, the PWSL model presents the advantages of accurate modeling of the traffic flow dynamics, modularity for a desired number of cells and the ability to consider more than one congestion wave. In addition, this model can be used in several applications, such as incident detection, traffic estimation and traffic control. It approximates the boundary condition (Morbidi et al., 2014; Canudas-De-Wit et al., 2012) by using traffic data measurements at the upstream and the downstream boundaries of each section, and is able to decouple the traffic flow dynamics of a given section.

For the sake of simplicity and without loss of generality, this section assumes that only one congestion wave can appear at a time in the considered stretch of road. Thus, if we consider a highway partitioned into N cells and two possible traffic statuses (free or congested), one obtains 2^N possible modes. Moreover, by considering normal traffic behavior, one obtains $2 \times N$ modes.

In order to represent the discrete dynamics using the traffic variables, we accept that:

- The cell i is considered free (F) if $(\rho_i(k)) \leq (\rho_{c,i})$.
- The cell i is considered congested (C) if $(\rho_i(k)) > (\rho_{c,i})$.

As shown in Table 1, in a road section of N cells, when a bottleneck is created, the congestion wave propagates from the jam cell to its upstream cells. The congestion status of cells begins at the cell N and propagates cell by cell (Mode 2 to Mode N) until the section is fully congested (Mode $N + 1$). The same evolution will be followed when the traffic flow is released: the free-flow status will propagate from the cell N to the upstream cells, cell by cell (Mode $N + 2$ to the Mode $2N$) until the free status is fully propagated throughout the whole section (Mode 1).

Having addressed the discrete representation of the PWSL traffic model, let us now present the modeling of the traffic flow dynamic in each mode. It supposes that the cell lengths can be non-uniforms and respect the Courant-Friedrichs-Lewy condition given by equation (6), which is necessary for the convergence of the CTM solutions with the LWR solutions (Muñoz et al., 2006):

$$\Delta T v_{f,i} \leq \Delta x_i. \quad (6)$$

Table 1: The discrete modes of normal traffic behavior.

Modes	Cells					
	1	2	...	$N-2$	$N-1$	N
1	F	F	...	F	F	F
2	F	F	...	F	F	C
3	F	F	...	F	C	C
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$
N	F	C	...	C	C	C
$N+1$	C	C	...	C	C	C
$N+2$	C	C	...	C	C	F
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$
$2N$	C	F	...	F	F	F

The density of each cell i is given by the following vehicle conservation law:

$$\rho_i(k+1) = \rho_i(k) + \frac{\Delta T}{\Delta x_i} (q_{i,in}(k) - q_{i,out}(k)), \quad (7)$$

where $q_{i,in}(k)$ and $q_{i,out}(k)$ are respectively the total entering and leaving flows of cell i over the interval $[k, k+1)$, including the mainline flow, the on-ramp and the off-ramp, if they exist. Δx_i and ΔT are the length of cell i and the step time, respectively. Finally, the flow can be obtained by

$$q_i(k) = \min(S_{i-1}(k), R_i(k)), \quad (8)$$

$$S_{i-1}(k) = \min(v_{i-1}\rho_{i-1}(k), Q_{M,i-1}) \quad (9)$$

$$R_i(k) = \min(Q_{M,i}, w_i(\rho_{j,i} - \rho_i(k))) \quad (10)$$

where $S_{i-1}(k)$ represents the supply function (9), which is the maximal flow that can be supplied by the cell $i-1$ under free flow conditions, and $R_i(k)$ is the demand function (10), which is the maximal flow that can be received by cell i under the congestion conditions.

The dynamic of the supply and demand function for cell 1 and cell N is given by Equations (11) and (12), respectively. cell 1:

$$\begin{cases} S_0 = q_{in} & \text{if cell 1 is free,} \\ R_1(k) = \min(Q_{M,1}, w_1(\rho_{j,1} - \rho_1(k))) & \text{if cell 1 is not free.} \end{cases} \quad (11)$$

cell N :

$$\begin{cases} S_{N-1}(k) = \min(v_{N-1}\rho_{N-1}(k), Q_{M,N-1}) & \text{if cell } N \text{ is free,} \\ R_N = q_{out} & \text{if cell } N \text{ is not free.} \end{cases} \quad (12)$$

After the presentation of the boundary condition, let us now address the dynamics of the flow according to the status of the traffic in the inner cells. In general, three principal connection types may be found in a given road segment: simple (without ramps), merge (with on-ramp) and diverge (with off-ramp). In this paper, to simplify the analysis, we will only deal with the road segments that have simple connections. A cell i has a simple connection if there are no ramps at its junctions with the upstream and downstream cells (Figure 3).

Figure 3: Simple connection.

In its general form, the dynamic of PWSL in each mode or subsystems is given by the following formulation:

$$\rho(k+1) = A_s \rho(k) + B_s u(k) + D_s \quad (13)$$

$$D_s = B_{jam,s} \rho_j + B_{Q,s} Q_M \quad (14)$$

where: $s = \{1, \dots, 2N\}$ indicates the mode of the system, $\rho(k) = [\rho_1(k), \dots, \rho_N(k)]^T$ is the density state vector, $u(k) = [q_{in}(k), q_{ON,2}(k), \dots, q_{ON,N-1}(k), q_{out}(k)]^T$ is the entering flow vector (input, output and on-ramp flow), $Q_M = [Q_{M,1}, \dots, Q_{M,N}]^T$ is the maximal flow vector, $\rho_j = [\rho_{j,1}, \dots, \rho_{j,N}]$ the jam density vector, and $(\rho(k), u(k), Q_M, \rho_j) \in \mathbb{R}^{N \times 1}$, $A_s \in \mathbb{R}^{N \times N}$, $B_s \in \mathbb{R}^{N \times N}$ and $D_s \in \mathbb{R}^{N \times N}$ are known matrices.

The selection parameters are, α , γ , δ , ζ and θ , which are introduced in order to facilitate the generalization of the model to any possible traffic status. The values of these selection parameters are given in Table 2 and Table 3.

$$A_s = \begin{bmatrix} \left(1 - \frac{\Delta T}{\Delta x_1} \alpha_1\right) & \frac{\Delta T}{\Delta x_1} \zeta_2 & 0 & \cdots & 0 \\ \frac{\Delta T}{\Delta x_2} \delta_1 & \left(1 - \frac{\Delta T}{\Delta x_2} \alpha_2\right) & \frac{\Delta T}{\Delta x_2} \zeta_3 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \frac{\Delta T}{\Delta x_{N-1}} \delta_{N-2} & \left(1 - \frac{\Delta T}{\Delta x_{N-1}} \alpha_{N-1}\right) & \frac{\Delta T}{\Delta x_{N-1}} \zeta_N \\ 0 & \cdots & 0 & \frac{\Delta T}{\Delta x_N} \delta_{N-1} & \left(1 - \frac{\Delta T}{\Delta x_N} \alpha_N\right) \end{bmatrix} \quad (15)$$

$$B_s = \begin{bmatrix} \frac{\Delta T}{\Delta x_1} \gamma_1 & \frac{\Delta T}{\Delta x_1} \gamma_2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{\Delta T}{\Delta x_{N-2}} \gamma_{N-2} & \frac{\Delta T}{\Delta x_{N-2}} \gamma_{N-1} & \\ & & 0 & \frac{\Delta T}{\Delta x_{N-1}} \gamma_{N-1} & 0 \\ 0 & \cdots & & 0 & \frac{\Delta T}{\Delta x_N} \gamma_N \end{bmatrix} \quad (16)$$

$$B_{jam,s} = \begin{bmatrix} \frac{\Delta T}{\Delta x_1} \theta_1 & -\frac{\Delta T}{\Delta x_1} \theta_2 & 0 & \cdots & 0 \\ 0 & \frac{\Delta T}{\Delta x_2} \theta_2 & -\frac{\Delta T}{\Delta x_2} \theta_3 & \ddots & \vdots \\ & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & & \frac{\Delta T}{\Delta x_{N-1}} \theta_{N-1} & -\frac{\Delta T}{\Delta x_{N-1}} \theta_N \\ 0 & \cdots & & 0 & \frac{\Delta T}{\Delta x_N} \theta_N \end{bmatrix} \quad (17)$$

$$B_{Q,s} = [b_{m,n}] \quad (18)$$

Table 2: Values of the selection variables according to the traffic status of cells i and its adjacent cells (F: Free status, C: Congested status).

Cells			Variable Selections				
cell $i-1$	cell i	cell $i+1$	α_i	δ_{i-1}	ζ_{i+1}	θ_i	θ_{i+1}
F	F	F	$v_{f,i}$	$v_{f,i-1}$	0	0	0
F	F	C	$v_{f,i}$	$v_{f,i-1}$	0	0	0
F	C	C	0	$v_{f,i-1}$	w_{i+1}	0	w_{i+1}
C	C	C	w_i	0	w_{i+1}	w_i	w_{i+1}
C	C	F	w_i	0	0	w_i	0
C	F	F	$v_{f,i}$	0	0	0	0

Moreover, the selection parameter γ_i is introduced in order to determine the effect of the entering flow vector on the density dynamic. The values of this parameter for the boundary cells are summarized in Table 3 and by $\gamma_{i,i \in [2, N-1]} = 0$.

Table 3: Values of the boundary cells parameters.

Cell	Status	Variable Values
1	F	$\gamma_1 = 1$
	C	$\gamma_1 = 0$
N	F	$\gamma_N = 0$
	C	$\gamma_N = -1$

Regarding the matrix D_s given by Equation 14, one remarks that $B_{jam,s}$ appears only in the congestion status. Moreover, $B_{Q,s}$ also appears in the upstream and downstream cells that form the boundary between the free and congested subsections, in the direction of the free wave propagation. Thus, by considering that the free wave is situated at the boundary of the cells i and $i + 1$, the elements of the matrix $B_{Q,s}$, given by Equation 18, can be written in the following form:

$$\begin{cases} b_{m,n} = -\frac{\Delta T}{\Delta x_i} & \text{if } m = i \wedge n = i + 1 \\ b_{m,n} = \frac{\Delta T}{\Delta x_{i+1}} & \text{if } m = i + 1 \wedge n = i + 1 \\ b_{m,n} = 0 & \text{if otherwise} \end{cases}$$

For more details about the PWSL model, refer to Zeroual et al. (2017)

3. Univariate statistical control charts

In manufacturing industries, statistical quality control has been traditionally used for monitoring and controlling product quality. Statistical process control charts can provide early warnings of abnormal changes in system operations, helping operators to identify the onset of potential abnormal events. These statistical charts include Shewhart, cumulative sum (CUSUM) and EWMA charts. Univariate statistical methods, such as the Shewhart chart and EWMA, have been widely used to monitor industrial processes for many years (Montgomery, 2007). Herein, we want to monitor traffic flow in an efficient manner via PWSL models, by applying statistical process control charts to traffic flow management. These methods are briefly introduced here.

3.1. Shewhart monitoring chart

Anomalies that occurs in a monitored process are detected by checking whether the current measurements are statistically different from the measurements that are known a priori to be faultless (i.e., measurements without anomalies). In a Shewhart chart, a sequence of samples (denoted by x_i) are plotted against time (Montgomery, 2007). Upper and lower control limits, which are denoted respectively by, UCL and LCL, for the samples are established around the

process mean (μ), based on the three-sigma rule, i.e.,

$$UCL, LCL = \mu_0 \pm 3\sigma_0,$$

where σ_0 is the standard deviation of the anomaly-free data computed when the process is running under healthy conditions. Whenever the most recently measured point or a consecutive sequence of points is outside the control limits, an abnormal condition is encountered and the source of the problem can be diagnosed.

3.2. Cumulative sum (CUSUM) charts

The history data obtained before current observation contains useful information for anomaly detection, but, the Shewhart charts do not make use of such information at all. For this reason, they are effective only for detecting large anomalies. Unlike Shewhart chart, CUSUM chart corporate all the information from past samples in addition to current samples in the decision procedure. The CUSUM charts have a good capacity to detect small shifts in the process mean due to an extensive memory of the process (Page, 1961). CUSUM chart monitors cumulative sums of the deviations of the sample rather than plotting a sample itself. The CUSUM statistic (s_t) is defined as the following (Montgomery, 2007):

$$s_t = \sum_{j=1}^n (x_j - \mu_0), \quad (19)$$

where t denotes the current time point, s_t is the cumulative sum of all samples, including the most recent, and μ_0 is the targeted process mean. The CUSUM statistic can be computed recursively using the following formula (Montgomery, 2007):

$$s_t = (x_t - \mu_0) + s_{t-1}. \quad (20)$$

A one-sided CUSUM statistic is computed using the following equation (Montgomery, 2007):

$$s_t = \sum_{j=1}^t \left[\bar{x}_j - (\mu_0 + k) \right], \quad (21)$$

where k is a parameter used as a reference to detect changes in the process mean. If s_t becomes negative, then the CUSUM statistic is set to zero. An out-of-control process is defined by s_t exceeding the decision interval, which is another parameter needed for the CUSUM charts to function. The parameters k and h are defined as $k = \frac{\Delta}{2}$, and $h = \frac{d\Delta}{2}$, respectively, where $d = \left(\frac{2}{\delta^2} \right) \ln \left(\frac{1-\beta}{\alpha} \right)$, $\delta = \frac{\Delta}{\sigma_x}$, σ_x is the standard deviation of the average of the process variable (x) being monitored, α and β are probabilities, and Δ is the size of the shift in the mean that needs to be detected. In practice, Montgomery recommends using a value of 4σ or 5σ for h (Montgomery, 2007). This choice

would provide a reasonable detection for a shift of 1σ in the process mean. Numerous variations of the CUSUM exist; for more details see (Montgomery, 2007). Here, we use Shewhart and CUSUM charts as a benchmark for abnormal traffic detection. In the next section, we briefly describe the EWMA chart and its use in anomaly detection.

3.3. EWMA monitoring chart

In this subsection, we briefly introduce the basic idea of the EWMA chart and its properties. For a more detailed discussion of the EWMA chart see Montgomery (2007); Harrou et al. (2013b). This chart is constructed based on the exponential weighting of the available observations, a design that provides improved sensitivity to small changes in the mean of a multivariate process. The EWMA control scheme was first introduced by Roberts (Lucas and Saccucci, 1990), and has been extensively used in time series analysis. The EWMA monitoring chart is also widely used as an anomaly-detection technique by scientists and engineers in various disciplines (Harrou and Nounou, 2014; Morton et al., 2001; Harrou et al., 2015; Kadri et al., 2016; Montgomery, 2007). Assume that $\{x_1, x_2, \dots, x_n\}$ are individual observations collected from a monitored process. The expression for the EWMA is (Montgomery, 2007):

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1}, \quad t = 0, \dots, n. \quad (22)$$

The starting value z_0 is usually set to be the mean of the anomaly-free data, μ_0 . The forgetting parameter $\lambda \in (0, 1]$ determines how fast EWMA forgets the data history. We can see that if λ is small, more weight is assigned to past observations and the chart is tuned to efficiently detect small changes in the process mean. On the other hand, if λ is large, more weight is assigned to the current observations, and the chart is more suitable for detecting large shifts (Montgomery, 2007; Kadri et al., 2016). When $\lambda = 1$, the EWMA is equal to the most recent observation, x_t , and provides the same results as the Shewhart chart. As λ approaches zero, EWMA approximates the CUSUM criteria, which gives equal weight to the current and historical observations.

Under anomaly-free conditions, the standard deviation of z_t is defined as

$$\sigma_{z_t} = \sigma_0 \sqrt{\frac{\lambda}{(2 - \lambda)} [1 - (1 - \lambda)^{2t}]}, \quad (23)$$

where σ_0 is the standard deviation of the anomaly-free or preliminary data set. The upper and lower control limits of the EWMA chart for detecting a mean shift are:

$$UCL, LCL = \mu_0 \pm L\sigma_{z_t}, \quad (24)$$

where L is a multiplier of the EWMA standard deviation σ_{z_t} . The parameters L and λ need to be set carefully (Montgomery, 2007; Kadri et al., 2016). L is usually specified in practice to be 3, which corresponds to a false alarm rate of 0.27%. If z_t is within the interval [LCL, UCL], then we conclude that the process is under control up to time point t . Otherwise, the process is considered out of control.

4. Wavelet-based multiscale filtering

Wavelet analysis has been shown to represent data with multiscale properties, efficiently separating deterministic and stochastic features (Bakshi, 1998). Herein, we introduce multiscale representation of data and describe their advantages when applied to anomaly detection techniques.

4.1. Wavelet transform

Wavelets represent a family of basis functions that can be expressed as the following localized in both time and frequency (Bakshi, 1998):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (25)$$

where a represents the dilation parameter, b is the translation parameter and $\psi(t)$ is the mother wavelet (Gao and Yan, 2010). Both these parameters are commonly discretized dyadically as $a = 2^m$, $b = 2^m k$, $(m, k) \in \mathbb{Z}^2$, and the family of wavelets can be represented as $\psi_{mn}(t) = 2^{-\frac{m}{2}} \psi(2^{-m}t - m)$. Here, $\psi(t)$ is the mother wavelet and m and k are the respective dilation and translation parameters. Different families of basis functions are created based on their convolution with different filters, such as the Haar scaling function and the Daubechies filters (Gao and Yan, 2010; Zhou et al., 2006).

The Discrete Wavelet transform (DWT) is used to decompose the original signal into a combination of approximation, A , and detail coefficients, D . Specifically, a set of scaling functions $\phi_{j,k}(t) = \sqrt{2^{-j}} \phi(2^{-j}t - k)$, $k \in \mathbb{Z}$ and a set of wavelet functions $\psi_{j,k}(t) = \sqrt{2^{-j}} \psi(2^{-j}t - k)$, $j = 1, \dots, J$, $k \in \mathbb{Z}$, which are respectively associated with low pass filter H and high pass filter G , are used to approximate the signal at multiple resolutions. Where the coarsest scale, J , usually termed the decomposition level. Any signal can be represented by a summation of all scaled and detailed signals as follows (Gao and Yan, 2010):

$$x(t) = \underbrace{\sum_{k=1}^{n2^{-J}} a_{Jk} \phi_{Jk}(t)}_{A_J(t)} + \sum_{j=1}^J \underbrace{\sum_{k=1}^{n2^{-j}} d_{jk} \psi_{jk}(t)}_{D_j(t)}. \quad (26)$$

where j, k, J and n represent the dilation parameter, translation parameter, number of scales, and number observations in the original signal, respectively (Strang, 1989; Daubechies, 1988). d_{jk} and a_{jk} are respectively the scaling and the wavelet coefficients, and $A_j(t)$ and $D_j(t)$, ($j = 1, 2, \dots, J$) represent the approximated signal and the detail signals, respectively. The procedure of decomposing a signal via DWT is illustrated in Figure 4.

Figure 4: Multilevel representation of a signal using DWT.

4.2. Separating noise feature

Two important applications: data compression and data denoising can be achieved through wavelet multiscale decomposition. One of the biggest advantages of multiscale representation is its capacity to distinguish measurement noise from useful data features, by applying low and high pass filters to the data during multiscale decomposition. This allows the separation of features at different resolutions or frequencies, which makes multiscale representation a better tool for filtering or denoising noisy data than traditional linear filters, like the mean filter and the EWMA filter. Despite their popularity, linear filters rely on defining a frequency threshold above where all features are treated as measurement noise. The ability of multiscale representation to separate noise has been used not only to improve data filtering, but also to improve the prediction accuracy of several empirical modeling methods and the accuracy of state estimators.

A noisy signal is filtered by a three-step method (Donoho et al., 1995):

- (1) Apply wavelet transform to decompose the noisy signal into the time-frequency domain
- (2) Threshold the detail coefficient and remove coefficients under the selected threshold.
- (3) Transform back into the original domain the thresholded coefficients to obtain a filtered signal.

5. Combining PWSL model with EWMA scheme (PWSL-EWMA)

In this paper, we exploit the advantages of PWSL modeling with those of the univariate EWMA monitoring chart, which should result in improved monitoring of traffic flow. The basic idea is to build a PWSL model based on the free-flow data (i.e., uncongested data). This model can then be used to monitor traffic evolution and to detect any abnormal traffic flow or congested situations. Towards this end, the EWMA chart is then used to evaluate residuals obtained from PWSL model. Residuals represent the difference between the measured traffic density and the predicted traffic density obtained from the PWSL model. Indeed, under free-flow conditions, the residuals are close to zero despite measurement noise and errors, but they significantly deviate from zero when traffic is congested. In this work, the residuals are used as congestion indicators. Defining the residual vector $\mathbf{e} = [e_1, \dots, e_t, \dots, e_n]$, as

$$\tilde{e}_t = y_t - \hat{y}_t, \quad t \in [1, n] \quad (27)$$

where y_t and $\hat{y}_t$ are the measured traffic density and the predicted traffic density obtained from the PWSL model, respectively. Then, the EWMA decision function can be computed using the residuals as follows:

$$z_t = \lambda e_t + (1 - \lambda)z_{t-1}, \quad t \in [1, n]. \quad (28)$$

The implementation of the developed monitoring method is comprised of two stages: offline modeling and online monitoring. In the offline modeling phase, PWSL is performed on the free-flow data (training data), enabling us to obtain a reference PWSL model. Then, the anomaly detection procedure is executed by using the reference PWSL model with EWMA chart in the online monitoring phase. The proposed PWSL-EWMA approach is schematically summarized as shown in Table 4, which is schematically represented in Figure 5.

Table 4: PWSL-based EWMA traffic congestion monitoring procedure.

Step	Action
1.	Given: <ul style="list-style-type: none"> • A training congestion-free data set that represents the normal traffic flow and a testing data set (possibly congested data), • The parameters of the EWMA control scheme: smoothing parameter λ and the control limit width L,
2.	Build the PWSL model using the training congestion-free data <ul style="list-style-type: none"> • Compute the the residuals using the built PWSL model, • Compute the control limits of the EWMA monitoring scheme'
3.	Test the new data <ul style="list-style-type: none"> • Compute the residuals using the built PWSL model, • Compute the EWMA decision statistic,
4.	Check for traffic congestions <ul style="list-style-type: none"> • Declare a traffic congestion when the EWMA decision function exceeds the control limits previously computed using the training data.

Figure 5: A flowchart of PWSL-EWMA traffic monitoring scheme.

However, the presence of measurement errors (noise) in the data and model uncertainties degrade the quality of anomaly detection techniques. The objective is to use a wavelet-based multiscale representation of the data (Ganesan et al., 2004) to further enhance the effectiveness of the proposed method. Indeed, one of the advantages of multiscale representation is its capacity to distinguish measurement noise from useful data features (Madakyaru et al., 2017; Harrou et al., 2013a), by applying low and high pass filters to the data during the multiscale decomposition. This allows the separation of features at different resolutions or frequencies, which makes multiscale representation a better tool for filtering or denoising noisy data than traditional linear filters, e.g., the mean filter and the exponentially weighted moving average filter. In the proposed PWSL-EWMA monitoring approach, multiscale data prefiltering will be used before applying the EWMA chart, in order to enhance the robustness of this method against measurement noise (Figure 5).

6. Simulation results

6.1. Data description and model calibration results

6.1.1. Geographical description

To assess the performance of the proposed PWSL-EWMA method, we consider two road sections, the first is a 1.5-mile stretch of I-210, a three-lane highway in southwestern California, which starts from the I/O Mills passed by Indian Hill ends in Mountain city. The second is a 1.73 miles stretch of SR60, which is a four-lane highway in east of California. This begins from San Antonio and ends in Grove City, including an ON and Off ramps in the Euclid point. We choose these segments for its interesting congestion periods that can help us test the evolution of our model, and also because they are equipped with the necessary magnetic loops embedded in the mainline highway. Respecting the Courant-Friedrichs-Lewy condition, the first segment has been partitioned into 3 cells (Figure 6) and the second into 4 cells (Figure 7). In which, each cell is equipped with a vehicle detector station (VDS), which provides data that we can compare to the simulation results.

(a) Road map I210-W highway portion

Abs PM: Absolute Post Mile
ML: Mainline
VDS: Vehicle Detector Station

(b) Schematic representation of the I210-W portion

Figure 6: 1.5-mile segment of highway I-210.

(a) Road map SR60-E highway portion

Abs PM: Absolute Post Mile
ML: Mainline
VDS: Vehicle Detector Station

(b) Schematic representation of the SR60-E portion

Figure 7: 1.73-mile segment of highway SR60-E.

6.1.2. Data calibration results

Data calibration or fundamental diagram identification is a necessary step for each macroscopic model validation (Gomes and Horowitz, 2009). This step provides the density-flow relationship or FD parameters (Figure 8) which are the free flow speed, the critical density, congestion wave speed, the maximum flow and the jam density. For data calibration, we exploit the abundance of raw VDS data, obtained with a sampling period of five minutes, provided by the Caltrans Performance Measurement System (PeMS) database for the highways segments described above (Figure 6 and Figure 7). Using data collected over a period of 98 days (equivalent to 28225 data points) guarantees that there are sufficient data points corresponding to congested traffic status (Figure 8). For more accuracy, we assume that for each road section, each cell has its own FD and we take into consideration the detectors health (99% functionality during this period). Flow and speeds are observed by VDS and density is estimated through the following equation (Zeroual et al., 2015):

$$Density = \frac{Flow}{Speed}. \quad (29)$$

The following procedure is adopted for model calibration.

- Step 1** Plot scatter plot of density-flow data provided by the VDSs.
- Step 2** Compute the maximum value of the flow, $\bar{Q}_{max} = \max(q_i)$, and determine the density value corresponding to the $\bar{Q}_{max}$. Through this last, the data scatter-plot are separated into two approximate regions (free and congested), and then the application of a first order least squares on each region is performed.
- Step 3** Border the congested area least square line by two parallel regression lines on the upper and lower bounds. This permits to limit and monitor the direction of the congestion area.
- Step 4** Identify the critical density, ρ_c , by projecting the intersection point between the upper bound of the congested area and the horizontal line of the value $\bar{Q}_{max}$.
- Step 5** Approximate the mean maximum flow Q_M with the mean of flow values corresponding to the critical density with a desired approximation.
- Step 6** Fit both least squares lines (congested and free) until they cross the mean maximum flow.
- Step 7** Identify the free flow speed v_f (slope value of free flow regression line), the congested wave speed w (slope value of the congested regression line) and the jam density ρ_{jam} (intersection point of congested line and the axis of densities).

We apply this calibration methodology on the considered road portions to identify the triangular FD form for each cell and estimate the PWSL model parameters. Figure 8 shows an example of a triangular FD form approximation, obtained by applying the calibration algorithm proposed in (Zeroual et al., 2017) to traffic data collected over 98 days. For more details about this calibration methodology, please refer to (Zeroual et al., 2017, 2015)

Figure 8: Scatter plot of traffic data and FD parameter identification (each dot denotes data point (flow, density)).

6.2. Model validation

The constructed PWSL model is fitted to the free-flow training dataset from three different cells from I-210 dataset, and the goodness of fit is shown in Figure 9(a-c). Here, the performance of PWSL model is compared to the CTM. Figure 10 shows the observed and predicted traffic density using a CTM. From Figure 9, we can see that the two time series for cell 01, cell 02 and cell 03 are well-adjusted by the constructed PWSL models. From Figures 9 and 10, it can be seen a high prediction accuracy of the proposed PWSL model compared to CTM.

Figure 9: Plots of measured and predicted traffic density from I-210 dataset using a PWSL macroscopic traffic model for cell 01(a), cell 02 (b) and cell 03 (c) for December 03rd, 2014.

Figure 10: Plots of measured and predicted traffic density I-210 dataset using a CTM for cell 01(a), cell 02 (b) and cell 02 (c) for December 03, 2014.

Furthermore, to illustrate the quality of the PWSL-based models, the scatter plots of the measured vehicle density data and the predicted values from the constructed PWSL-based models as well as from CTM are presented in Figure 11 and Figure 12, respectively. It is clear that the measured data are well-fitted by the PWSL-based model.

Figure 11: Scatter plots of the measured traffic density data and the predicted data obtained from the PWSL-based models for three datasets of cell 01 (a), cell 02 (b) and cell 03 (c).

Figure 12: Scatter plots of the measured traffic density data and the predicted data obtained from the CTM-based models for three datasets of cell 01 (a), cell 02 (b) and cell 03 (c).

Now, the constructed PWSL model and CTM are fitted to the free-flow training datasets from four different cells from SR60 dataset, and the quality of fit is shown in Figure 13(a-d) and Figure 15(a-d), respectively. Scatter plots of the measured vehicle density data and predicted values from the PWSL-based models and CTM are shown in Figure 14 and Figure 14(a-d), respectively. These Figures testify again that the proposed model has well learned the structure of this data. The PWSL model showed a high prediction accuracy compared to CTM.

Figure 13: Plots of measured and predicted traffic density from SR60 dataset using a PWSL macroscopic traffic model for for four datasets of cell 01 (a), cell 02 (b), cell 03 (c) and cell 04 (d) for June 23, 2012.

Figure 14: Scatter plots of the measured traffic density data and the predicted data obtained from the PWSL-based models for cell 01 (a), cell 02 (b), cell 03 (c) and cell 04 (d).

Figure 15: Plots of measured and predicted traffic density from SR60 dataset using a CTM for cell 01 (a), cell 02 (b), cell 03 (c) and cell 04 (d) for June 23, 2012..

Figure 16: Scatter plots of the measured traffic density data and the predicted data obtained from the CTM-based models for two datasets of cell 01 (a), cell 02 (b), cell 03 (c) and cell 04 (d).

In addition, to evaluate the performance of the constructed PWSL model, three numerical criteria were used: r^2 , the mean absolute percent error (MAPE) and the root mean square error (RMSE). These were calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}, \quad (30)$$

$$MAPE = 100 \left[\frac{\sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{|y_i|} \right)}{n} \right], \quad (31)$$

$$r^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \text{mean}(Y))^2}, \quad (32)$$

where y_t is the measured values, $\hat{y}_t$ is the corresponding predicted values by the PWSL model and n is the number of samples.

Tables 5 and 6 summarize the goodness of fit statistics obtained for the two models studied (PWSL and CTM) applied to two experimental datasets I-210 and SR60, respectively. The results in Tables 5 and 6 show that the PWSL models describe well the traffic density. It was also seen that PWSL models used in this study exhibited high r^2 values (above 0.95), indicating overall good predictive quality. According to r^2 , MAPE and RMSE values the PWSL model provided better results compared to the CTM. This demonstrates that the constructed PWSL models can be used for modeling traffic density effectively.

Table 5: Statistical validation metrics applied to data from Figure 9.

	PWSL (I-210)			CTM (I-210)		
	Cell 01	Cell 02	Cell 03	Cell 01	Cell 02	Cell 03
r^2	0.99	0.97	0.959	0.9282	0.90495	0.8803
RMSE (veh/mi)	0.25	0.88	1.6915	3.154	4.1187	4.9693
MAPE %	4.93	7.6	9.078	14.32	14.841	16.091

Table 6: Statistical validation metrics applied to data from Figure 15.

	PWSL (SR60)				CTM (SR60)			
	Cell 01	Cell 02	Cell 03	Cell 04	Cell 01	Cell 02	Cell 03	Cell 04
r^2	0.99	0.954	0.9522	0.9584	0.822	0.8491	0.9412	0.9396
RMSE (veh/mi)	0.3042	1.5928	1.649	7.1287	5.9878	5.2566	2.0288	8.5276
MAPE %	3.80	6.9575	7.671	1.9684	17.1841	14.633	8.7578	2.8572

6.3. Detection results

After the PWSL model is built from the free-flow traffic dataset (i.e., uncongested data), it can be used for monitoring traffic flow. Three case studies are presented here to assess the performance of the developed PWSL-EWMA monitoring approach. The traffic density data in the first case contains abrupt congestion, the data in the second case study is contaminated by intermittent congestion and the data for the third case study has gradual congestion.

To quantify the efficiency of the developed approach, we use two metrics: the false alarm rate (FAR) and the missed detection rate (MDR) (Harrou et al., 2016). The FAR is the number of normal observations that are wrongly judged as faulty (false alarms) over the total number of anomaly-free data. The MDR is the number of anomalies that are wrongly classified as normal (missed detections) over the total number of anomalies.

6.3.1. Abrupt congestion

In the first case study, an abrupt congestion is simulated by adding a small constant deviation to the raw data between samples 1600 and 1800. The magnitude of the deviation is equal to 10% of the total variation in the raw data. This could represent an incident such as stalled cars, road repairs, overturned vehicles, and bad weather. Monitoring results of PWSL-based Shewhart, CUSUM and EWMA charts, and CTM-EWMA chart based on unfiltered are shown in Figure 17(a-d). All the charts show signs of a congestion because the magnitude of this congestion in this case is quite large. Figure 17(a) shows that the PWSL-Shewhart is able to detect the congestion, but with several false alarms. We then apply the CUSUM chart with $k = 0.25$ and $h = 0.85$ and the EWMA chart with $\lambda = 0.25$ to the testing dataset. The CUSUM chart is shown in Figure 17(b), from which it can be seen that the CUSUM statistic violate clearly the control limit and thus the ability of this chart to detect this congestion, but it resulted in several false alarms (i.e., FAR=52.015%) and some missed detections (MDR=10.5%). Indeed, after the end of the abnormal conditions the CUSUM chart takes much time to return back to the nominal state and thus resulting in large number of false alarms. From Figure 17(d), it can be seen that the CTM-EWMA chart with $\lambda = 0.25$ can detect the abrupt congestion but with several false alarms (i.e., FAR=40.02%). On other hand, Figure 17(c) shows that PWSL-EWMA correctly detect this congestion with few false alarms (i.e., FAR=3.39%).

Figure 17: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the unfiltered residuals in the presence of abrupt congestion between samples 1600-1800 (first example).

It is well-known that filters have been widely used for reducing false alarms caused by noisy data or model errors. A wavelet-based multiscale representation of data has been used extensively in literature to enhance the effectiveness and robustness of fault detection strategies. The multiscale representation of data using wavelets is a powerful feature extraction tool that is well-suited to denoising time series data (Sheriff et al., 2014; Harrou et al., 2013a; Bakshi, 1998). The four monitoring charts when applied to filtered residuals are shown in Figure 18(a-d). Figure 18(a) reflect an improvement in the performance of the PWSL-based Shewhart chart when applied to filtered residuals rather than unfiltered residuals, reducing the false alarm rate from 5.33% to 1.24%. Figure 18(b) shows that the PWSL-CUSUM is able to detect the congestion, but with several false alarms. It can be seen from Figure 18(d) that the CTM-EWMA chart can indeed detect this fault, but with an expense of false alarms (i.e., FAR=39.05%). On the other hand, the PWSL-EWMA chart applied to the filtered residuals performs reasonably well (see Figure 18(c)). Results in this case study show that the PWSL-EWMA chart based on the filtered residuals outperformed the other charts in detection traffic congestion.

Figure 18: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the filtered residuals in the presence of abrupt congestion between samples 1600-1800 (first example).

In the second example, an abrupt congestion is simulated by adding a small constant deviation which is 5% of the total variation in the raw data between samples 3000 and 3500. Results of the four charts are demonstrated in Figure 19(a-d). Figure 19(a) shows that the Shewhart chart cannot correctly detect this fault (i.e., MDR=77%). The Shewhart chart is insensitive to this fault because it is designed to detect relatively moderate and large anomaly, while the magnitude of congestion in this case is small. This is mainly due to the fact that the Shewhart chart uses only the observed data at a particular instant to make a decision about the process performance, ignoring past data. CUSUM and EWMA charts, incorporates information from the entire process history, rather than just the most recent observations, so that they are more sensitive to small changes than Shewhart chart. The CUSUM chart is shown in Figure 19(b), from which it can be seen that the CUSUM chart detect this traffic congestion, but it resulted in several false alarms (i.e., FAR=13.25%). The performance of the CTM-EWMA chart is shown in Figure 19(d), from which it can be seen that this chart is able to detect this traffic congestion but with several false alarms (i.e., FAR=43.53%). From Figure 19(c), it can be seen that the EWMA chart based on unfiltered residuals is able to detect the abrupt congestion, but with some false alarms (false alarm rate of 5.83%) and few missed detections (MDR of 0.6%).

Figure 19: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the unfiltered residuals in the presence of abrupt congestion between samples 3000-3500 (second example).

The four monitoring charts based on the filtered residuals are shown in Figure 20(a-c). The results presented in Figure 20(a) reflect a slight improvement in the performance of the PWSL-based Shewhart chart when applied to filtered residuals rather than unfiltered residuals, minimally reducing the false alarm rate from 7.4% to 7.22%. CUSUM chart is able to detect this congestion but with several false alarms (see Figure 20(b)). Results of EWMA chart applied to the filtered residuals obtained from the CTM are shown in Figure 20(d). Figure 20(d) shows that the CTM-EWMA chart detected this congestion but on the expense of a lot of false alarms (i.e. FAR= 42.54%). The EWMA chart using multiscale filtering is shown in Figure 20(c), from which it can be seen that the EWMA statistic clearly exceeds the control limit, thus confirming the ability of this chart to detect the abnormal traffic flow, with fewer false alarms (FAR of 1.7%) and no missed detections. These results show an improvement of the studied charts applied to the filtered data over unfiltered data.

Figure 20: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the filtered residuals in the presence of abrupt congestion between samples 3000-3500 (second example).

6.3.2. Intermittent congestion

In the second case study, we introduce into the testing data a bias of amplitude 3% of the total variation in the testing data between samples 1600 and 1800, and a bias of 10% from sample 3000 to sample 3500. The monitoring results of the Shewhart chart based on unfiltered residuals are demonstrated in Figure 21(a). From Figure 21(a), it can be seen that the Shewhart chart is capable of detecting these moderate abnormal congestion but with some missed detections (MDR=20.28% and FAR=3.01%). Figure 21(b) shows the monitoring results of the PWSL-CUSUM chart based on unfiltered residuals. However, the CUSUM chart gave several false alarms, a FAR of 22.23%. Application of the CTM-EWMA chart to the testing data is shown in Figure 21(d). The CTM-EWMA statistic clearly violates the control limits and thus the ability of this chart to detect this intermittent congestion but with several false alarms (i.e., MDR=21.14% and FAR=39.75%). From Figure 21(c), it can be seen that the PWSL-EWMA chart can detect intermittent congestion, but with some false alarms and missed detections (MDR=15.85% and FAR=3.06%).

Figure 21: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the unfiltered residuals in the presence of intermittent congestion.

The PWSL-Shewhart chart based on filtered residuals has a much smaller false alarm rate ($FAR=4.57\%$ and $MDR=20.57\%$) than that of the PWSL-Shewhart method with unfiltered residuals (Figure 22(a)). Figure 22(b) shows that the PLS-MCUSUM chart is capable of detecting this intermittent congestion but with an expense of false alarms. Therefore, the Shewhart method based on filtered residuals is more suitable for detecting moderate congestions. Figure 22(d) shows that the CTM-EWMA chart is capable of detecting this intermittent congestion but with a lot of false alarms (i.e., $MDR=18.71\%$ and $FAR=39.33\%$). In Figure 22(c) we can see that detection performance is much enhanced when EWMA chart is applied to the filtered residuals ($MDR=14.71\%$ and $FAR=3.37\%$). This case study testifies again to the superiority of the proposed approach compared to PWSL-Shewhart, PWSL-CUSUM and CTM-EWMA approaches. The PWSL-EWMA chart is also able to capture congestion severity (amplitude), in addition to congestions. The larger the congestion is, the greater the amplitude of the EWMA statistic, compared to its amplitude under free traffic.

Figure 22: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the filtered residuals in the presence of intermittent congestion.

6.3.3. Gradual congestion

The aim of the third case study is to assess the potential of the proposed PWSL-EWMA monitoring approach to detect gradual congestion of vehicle traffic on a highway. To do so, a slow increase with a slope of 0.01 was added to the simulated test data starting at sample number 3000. Figure 23(a) shows that the Shewhart chart statistic, when applied to the unfiltered data, exceeds the control limit after the occurrence of congestion at sample number 3042, but with several false alarms (FAR=5.24%). The CUSUM and EWMA charts begins to increase linearly from the sample 3000th and exceed the control limits around the 3042 sample (see Figures 23(b)-(c)). In the CUSUM chart, k and h are chosen to be 0.25 and 0.85, respectively, and in the EWMA chart, λ is chosen to be 0.25. Figure 23(d) shows the monitoring results of CTM-EWMA chart. A signal is first detected at sample 3092. Therefore, fewer extra observations are needed for the PWSL-EWMA and CUSUM charts to detect this gradual congestion compared to the other charts.

Figure 23: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the unfiltered residuals in the presence of gradual congestion.

The Shewhart chart, when applied to the filtered residuals, gradually increases as the congestion slowly develops, and begins to violate the threshold value when the size of the congestion becomes sufficiently important that it is detected by this model (Figure 24(a)), with fewer false alarms ($FAR=0.8\%$). The PWSL-CUSUM chart is shown in Figure 24(b), which first flags the congestion at sample 3042 without false alarms. From Figure 24(c), it can be seen that the first signal of abnormal traffic flow is given by the PWSL-EWMA chart at sample number 3042, without any false alarms. Figure 24(d) shows that the CTM-EWMA detect the first signal at the 30109 observation but with several false alarms ($FAR=48.06\%$). Of course, this paper also demonstrates through simulated data that significant improvement in traffic congestion detection can be obtained by using the PWSL model when combined with well established statistical technique the EWMA chart. Also, the PWSL-EWMA method outperforms the PWSL-based Shewhart and CUSUM methods and CTM-EWMA approach by detecting all anomalies with a smaller number of false alarms. These results also show that applying the proposed approach on pre-filtered residuals slightly enhanced its detection ability by reducing a number of false alarms.

Figure 24: Monitoring results of PWSL-Shewhart (a), PWSL-CUSUM (b), PWSL-EWMA (c) and CTM-EWMA (d) charts when applied to the filtered residuals in the presence of gradual congestion.

7. Discussion

In this paper, the ability of PWSL model to accurately describe traffic flow has been studied through practical density traffic data. The accuracy of traffic flow modeling using PWSL model and CTM, which is one of the most well-known traffic models in the literature of highway traffic modelling, has also been compared. Results clearly demonstrates the significant advantages of PWSL model over CTM in providing an accurate description of traffic density. This is mainly due to the flexibility and simplicity of PWSL model to model traffic flow dynamics.

A good description of traffic dynamics may be a tool for facilitating detection of traffic congestion. After the model is identified, it is used to monitor traffic congestion. If a traffic congestion is detected, the drivers can choose alternative routes to avoid traffic jam and long waiting times. This study shows that the estimated model can be combined with the well-established techniques of statistical monitoring for more accurate congestion detection. Residuals, which are the difference between computed outputs from PWSL model or CTM and measured values, enabling us to monitor traffic congestion. Indeed, the PWSL-Shewhart charts use the observed residuals at the current time point alone for making decisions about the traffic status at the current time point. Results show that such control charts are effective for detecting relatively large congestion, and are inefficient to detect moderate traffic congestion. The ability to detect moderate levels of traffic congestion can be improved by using a chart based on the information contain in the entire process history such as CUSUM and EWMA. In particular, it is shown through simulation using real data that the proposed PWSL-EWMA method outperformed PWSL-Shewhart, PWSL-CUSUM and CTM-EWMA, and exhibited the highest accuracy.

Unfortunately, the effectiveness of model-based congestion detection approaches relies on the accuracy of the models used. Modeling errors and measurement noise can affect detection performance of the proposed PWSL-

EWMA monitoring chart. Because of collected data from sensors is inherently contaminated by noise, a filter can usually be designed for reducing measurement noise in data. Here, we used multiscale representation to pre-filter the residuals obtained from PWSL model, and then use the filtered residuals in congestion monitoring using control charts, such as EWMA chart. The reason behind the advantage of using the multiscale filtering is that multiscale representation (or filtering) provides an effective separation of features, which helps perform better fault detection. Furthermore, the application of monitoring charts on the filtered residuals via wavelet-based multiscale filter improve further the congestion detection accuracy. Of course, multiscale filtering of the residuals obtained from the PWSL model improves congestion detection performance of the PWSL-EWMA method by reducing the number of false alarms.

8. Conclusion

The problem of monitoring road traffic congestion is addressed in this paper. This paper offers a procedure based on a piecewise switched linear (PWSL) macroscopic traffic model and an EWMA monitoring chart for the detection of road traffic congestion. It is based on the calculation of residuals using the PWSL model. These residuals are then used as the input data for the EWMA chart to check the presence of abnormal traffic flow. The effectiveness and superiority of the PWSL-EWMA method was demonstrated on traffic data collected from a section of the Interstate 210 (I-210) freeway in California and a portion of the State Route 60 (SR60) highway which is a four lane highway in the east of California. Satisfactory detection results were obtained using the proposed method. The new scheme, PWSL-EWMA, was confirmed to be more powerful than the conventional PWSL-Shewhart and CUSUM methods. This study shows also that a better result is obtained when EWMA is applied to residuals filtered via wavelet-based multiscale filtering than when applied to unfiltered residuals, especially for moderate traffic congestion.

The following two important perspectives are obtained through this study.

- While the PWSL model can reasonably describe traffic flow, it is based on traffic density, which is a fundamental macroscopic characteristic of traffic flow, alone. One direction for future research can, therefore, be to consider models including several variables to enhance accuracy. Moreover, it would be useful to incorporate more data inputs such as traffic count, speed, density, vehicle classification, and meteorological data to further enhance the effectiveness of a congestion detection system.
- The occurrence of possible traffic congestion can be caused by traffic accidents, road repairs, overturned vehicles or bad weather. After detecting traffic congestion, it is very useful to provide more information about traffic congestion levels. Another direction for future research is to classify road traffic congestion levels based

on the magnitude of the residuals by using machine learning-based classification. Specifically, three classes can be defined: jam, heavy and light. Thus, the goal of this line of research will be to develop monitoring techniques which can detect and classify road traffic congestion levels. One way is to use detector to discriminate congestion-free data from congested data and then only the classification phase is executed only if there is congestion.

Acknowledgement

The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2015-CRG4-2582. We would like to thank the reviewers of this article for their insightful comments, which helped us to greatly improve its quality.

References

- Aw, A., Rascle, M., Jan 2000. Resurrection of "Second Order" Models of Traffic Flow, *SIAM Journal on Applied Mathematics* 60 (3), 916–938.
- Bakshi, B., 1998. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal* 44 (7), 1596–1610.
- Barria, J., Thajchayapong, S., 2011. Detection and classification of traffic anomalies using microscopic traffic variables. *IEEE Transactions on Intelligent Transportation Systems* 12 (3), 695–704.
- Burghout, W., Koutsopoulos, H., Andreasson, I., 2006. A discrete-event mesoscopic traffic simulation model for hybrid traffic simulation. In: 2006 IEEE Intelligent Transportation Systems Conference. IEEE, pp. 1102–1107.
- Canudas-De-Wit, C., Ojeda, L. L., Kibangou, A. Y., 2012. Graph constrained-CTM observer design for the Grenoble south ring. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 197–202.
- Chang, E., 1992. A neural network approach to freeway incident detection. In: The 3rd International Conference on Vehicle Navigation and Information Systems, VNIS. IEEE, pp. 641–647.
- Chen, P.-T., Chen, F., Qian, Z., 2014. Road traffic congestion monitoring in social media with hinge-loss markov random fields. In: IEEE International Conference on Data Mining (ICDM). IEEE, pp. 80–89.
- Chen, X., Shi, Q., Li, L., Aug 2010. Location specific cell transmission model for freeway traffic. *Tsinghua Science and Technology* 15 (4), 475–480.
- Cheu, R., Srinivasan, D., Teh, E., 2003. Support vector machine models for freeway incident detection. In: Proceedings on Intelligent Transportation Systems. Vol. 1. IEEE, pp. 238–243.
- Daganzo, C., 1999. The lagged cell-transmission model. In: 14th International Symposium on Transportation and Traffic Theory.
- Daganzo, C. F., Aug 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological* 28 (4), 269–287.
- Daganzo, C. F., Aug 1995. Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B: Methodological* 29 (4), 277–286.
- D'Andrea, E., Marcelloni, F., 2016. Detection of traffic congestion and incidents from GPS trace analysis. *Expert Systems with Applications* 73 (1), 43–56.
- Daubechies, I., 1988. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics* 41 (7), 909–996.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G., Picard, D., 1995. Wavelet shrinkage: Asymptotia? *J. Roy. Stat. Soc. B* 57 (301).

- Ganesan, R., Das, T., Venkataraman, V., 2004. Wavelet-based multiscale statistical process monitoring: A literature review. *IIE transactions* 36 (9), 787–806.
- Gao, R., Yan, R., 2010. *Wavelets: Theory and applications for manufacturing*. Springer Science & Business Media.
- Gning, A., Mihaylova, L., Boel, R. K., jun 2011. Interval Macroscopic Models for Traffic Networks. *IEEE Transactions on Intelligent Transportation Systems* 12 (2), 523–536.
- Godunov, S. K., 1959. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik* 89 (3), 271–306.
- Gomes, G., Horowitz, R., 2009. Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity. *Transportation Research Board* (January), 1–14.
- Harrou, F., Nounou, M., 2014. Monitoring linear antenna arrays using an exponentially weighted moving average-based fault detection scheme. *Systems Science & Control Engineering: An Open Access Journal* 2 (1), 433–443.
- Harrou, F., Nounou, M., Nounou, H., 2013a. Enhanced monitoring using PCA-based GLR fault detection and multiscale filtering. In: *2013 IEEE Symposium on Computational Intelligence in Control and Automation (CICA)*. IEEE, pp. 1–8.
- Harrou, F., Nounou, M., Nounou, H., 2013b. A statistical fault detection strategy using PCA based EWMA control schemes. In: *9th Asian Control Conference (ASCC)*. IEEE, pp. 1–4.
- Harrou, F., Nounou, M., Nounou, H., Madakyaru, M., 2015. PLS-based EWMA fault detection strategy for process monitoring. *Journal of Loss Prevention in the Process Industries* 36, 108–119.
- Harrou, F., Sun, Y., Madakyaru, M., 2016. Kullback-leibler distance-based enhanced detection of incipient anomalies. *Journal of Loss Prevention in the Process Industries* 44, 73–87.
- Jin, X., Srinivasan, D., Cheu, R., 2001. Classification of freeway traffic patterns for incident detection using constructive probabilistic neural networks. *IEEE Transactions on Neural networks* 12 (5), 1173–1187.
- Kadri, F., Harrou, F., Chaabane, S., Sun, Y., Tahon, C., 2016. Seasonal arma-based spc charts for anomaly detection: Application to emergency department systems. *Neurocomputing* 173, 2102–2114.
- Kerner, B. S., 2011. Modern approaches to basic traffic modeling: Three-Phase Traffic Theory. *Transp. Res. Circular E-C* 149 (June), 22–42.
- Kinoshita, A., Takasu, A., Adachi, J., 2015. Real-time traffic incident detection using a probabilistic topic model. *Information Systems* 54, 169–188.
- Kuang, W., An, S., Jiang, H., 2015. Detecting traffic anomalies in urban areas using taxi GPS data. *Mathematical Problems in Engineering* 2015.
- Lemarchand, A., Koenig, D., Martinez, J. J., dec 2010. Robust design of a switched PI controller for an uncertain traffic model. In: *49th IEEE Conference on Decision and Control (CDC)*. IEEE, pp. 2149–2154.
- Lighthill, M. J., Whitham, G. B., may 1955. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 229 (1178), 317–345.
- Lucas, J., Saccucci, M., 1990. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics* 32 (1), 1–12.
- Madakyaru, M., Harrou, F., Sun, Y., 2017. Improved data-based fault detection strategy and application to distillation columns. *Process Safety and Environmental Protection* 107 (April 2017), 22–34.
- Mehboob, F., Abbas, M., Jiang, R., 2016. Traffic event detection from road surveillance videos based on fuzzy logic. In: *SAI Computing Conference (SAI)*, 2016. IEEE, pp. 188–194.
- Montgomery, D., 2007. *Introduction to statistical quality control*. John Wiley & Sons.
- Morbidi, F., Ojeda, L. L., Canudas de Wit, C., Bellicot, I., jun 2014. A new robust approach for highway traffic density estimation. In: *2014 European Control Conference (ECC)*. IEEE, pp. 2575–2580.
- Morton, P., Whitby, M., McLaws, M., Dobson, A., McElwain, S., Looke, D., Stackelroth, J., Sartor, A., 2001. The application of statistical process

- control charts to the detection and monitoring of hospital-acquired infections. *Journal of quality in clinical practice* 21 (4), 112–117.
- Muneer, T., Celik, A., Caliskan, N., 2011. Sustainable transport solution for a medium-sized town in turkey - A case study. *Sustainable Cities and Society* 1 (1), 29–37.
- Muñoz, L., Sun, X., Horowitz, R., Alvarez, L., 2006. Piecewise-Linearized Cell Transmission Model and Parameter Calibration Methodology.
- Muralidharan, A., Dervisoglu, G., Horowitz, R., 2009. Freeway traffic flow simulation using the Link Node Cell transmission model. In: 2009 American Control Conference. IEEE, pp. 2916–2921.
- Page, E., 1961. Cumulative sum charts. *Technometrics* 3 (1), 1–9.
- Payne, H. J., 1971. *Models of Freeway Traffic and Control*. Simulation Councils, Incorporated.
- Pongpaibool, P., Tangamchit, P., Noodwong, K., 2007. Evaluation of road traffic congestion using fuzzy techniques. In: IEEE Region 10 Conference, TENCON 2007. IEEE, pp. 1–4.
- Ramalingam, S., Varsani, V., 2016. Vehicle detection for traffic flow analysis. In: IEEE International Carnahan Conference on Security Technology. IEEE 978-1-5090.
- Richards, P. I., feb 1956. Shock Waves on the Highway. *Operations Research* 4 (1), 42–51.
- Ritchie, S., Abdulhai, B., 1997. Development testing and evaluation of advanced techniques for freeway incident detection. California Partners for Advanced Transit and Highways (PATH).
- Ross, P., dec 1988. Traffic dynamics. *Transportation Research Part B: Methodological* 22 (6), 421–435.
- Sheriff, M., Harrou, F., Nounou, M., 2014. Univariate process monitoring using multiscale shewhart charts. In: International Conference on Control, Decision and Information Technologies (CoDIT). IEEE, pp. 435–440.
- Strang, G., 1989. Wavelets and dilation equations: A brief introduction. *SIAM review* 31 (4), 614–627.
- Whitham, G., 1974. *Linear and nonlinear waves*, John Wiley Edition.
- Williams, B., Guin, A., 2007. Traffic management center use of incident detection algorithms: Findings of a nationwide survey. *IEEE Transactions on intelligent transportation systems* 8 (2), 351–358.
- Yong-chuan, Z., Xiao-qing, Z., Zhen-ting, C., 2011. Traffic congestion detection based on GPS floating-car data. *Procedia Engineering* 15, 5541–5546.
- Yu, J., Zuo, M., 2015. A video-based method for traffic flow detection of multi-lane road. In: Seventh International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE, pp. 68–71.
- Yu, X., Xiong, S., He, Y., Wong, W. E., Zhao, Y., 2016. Research on campus traffic congestion detection using bp neural network and markov model. *Journal of Information Security and Applications* 31, 54–60.
- Zeroual, A., Messai, N., Kechida, S., Hamdi, F., 2015. Calibration and Validation of a Switched Linear Macroscopic Traffic Model. In: 3rd International Conference on Control, Engineering & Information Technology CEIT2015. No. 1. IEEE, pp. 25–27.
- Zeroual, A., Messai, N., Kechida, S., Hamdi, F., 2017. A Piecewise Switched Linear Approach for Traffic Flow Modeling. *International Journal of Automation and Computing*.
- Zhang, H., sep 1998. A theory of nonequilibrium traffic flow. *Transportation Research Part B: Methodological* 32 (7), 485–498.
- Zhou, S., Sun, B., Shi, J., 2006. An SPC monitoring system for cycle-based waveform signals using Haar transform. *IEEE Transactions on Automation Science and Engineering* 3 (1), 60–72.