2011 OTC Research Project Report

Integrating Traffic Operation with Emission Impact using Dual-loop Data

Submitted by

Heng Wei, Ph.D., P.E.
Associate Professor, School of Advanced Structures
Director, ART-Engines Transportation Research Laboratory
College of Engineering and Applied Science
University of Cincinnati, 792 Rhodes Hall
Cincinnati, OH 45221-0071
Tel: 513-556-3781; Fax: 513-556-2599
Email: heng.wei@uc.edu

Submitted to:

Ohio Transportation Consortium (OTC)
The University of Akron
Akron, OH 44325-6106

February 22, 2012

This Report is prepared by

Principal Investigator (PI):

Heng Wei, Ph.D., P.E.
Associate Professor, School of Advanced Structures
Director, ART-Engines Transportation Research Laboratory
College of Engineering and Applied Science
University of Cincinnati, 792 Rhodes Hall
Cincinnati, OH 45221-0071
Tel: 513-556-3781; Fax: 513-556-2599
Email: heng.wei@uc.edu

Project Researcher:

Hao Liu
Ph.D. Candidate, School of Advanced School
Research Assistant, ART-Engines Transportation Research Laboratory
College of Engineering & Applied Science
University of Cincinnati, 735 ERC
Cincinnati, Ohio 45221-0071
Tel: 513-828-4738
E-mail: liuh5@mail.uc.edu

Co-Principle Investigator:

Mingming Lu, Ph.D.
Associate Professor
School of Energy, Environmental, Biological and Medical Engineering
College of Engineering and Applied Science
University of Cincinnati, 797 Rhodes Hall
Cincinnati, OH 45221-0012
Tel: 513-556-0996
E-mail: mingming.lu@uc.edu

Collaborator:

Ping Yi, Ph.D., P.E.,
Professor, Civil Engineering
The University of Akron
Auburn Science & Engineering Center Rm 213
Akron, OH 44325-6106
Tel: 330-972-7294; Fax: 330-972-5449
Email: pyi@uakron.edu

February 22, 2012

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGEMENT

The authors would like to thank Ohio Transportation Consortium (OTC) for its strong support of the grant. Special thanks go to Mr. Qingyi Ai, Ph.D. candidate in the Advanced Research in Transportation Engineering and Systems (ART-Engines) laboratory at the University of Cincinnati (UC), for his great assistance in collecting video data, extracting vehicle trajectory datasets, and modeling efforts as needed for this project. The authors also express their thanks to Dr. Ben Coifman at the Ohio State University for providing the event dual-loop data of the study sites. Special gratitude goes to Ms. Angela Brodie, Program Assistant of OTC, for her administrative assistance during the implementation of the project.

EXECUTIVE SUMMARY

Transportation contributes great amount of green house gases and other pollutant emissions to the global environment. Localized analysis of on-road traffic source emissions is often required by the U.S. Environmental Protection Agency (EPA) for project-level conformity analysis of transportation projects in accordance with the Phase Implementation Plan (SIP). In public environmental health field, some healthy concerns have been addressed on their association with mobile source air pollution. However, studies on levels of exposures to mobile source emissions and their influence on children's health obviously lack local micro-scale data to reflect the characterization of variability in emissions under various representative real-world vehicle activities and traffic conditions. The dual-loop data could be utilized as a rich data source for micro-scale emission study. However, little research has been reported on it and one of reasons lies in the incapability of producing accurate traffic inputs for the prevailing emission tool MOVES (Motor Vehicle Emission Simulator) by using current dual-loop models.

In this study a methodological framework is developed for integrating the improved dual-loop models and the MOVES emission model to estimate the emission impact of traffic flow operation by utilizing the data source available at dual-loop monitoring stations in highways. Firstly, we apply the framework to eliminate dual-loop data errors by a dual-loop data screening algorithm. There are six filters in the dual-loop data screening algorithm, namely, 0-1 filter, minimum headway filter, pairing filter, sensitivity filter, on-time filter, and median speed filter. The screened dual-loop data is then processed by a traffic flow phase identification algorithm in order to prepare the dual-loop data for the MOVES inputs calculations. The traffic flow phase identification algorithm uses a hybrid model that incorporates the level-of-service based method and K-means clustering method to separate free flow records, free flow to synchronized flow transition records, synchronized flow records, and traffic jam records from the raw dataset. The processed dual-loop data, which carries traffic flow phase information, is used by the MOVES inputs calculation algorithm to generate traffic volume, vehicle composition and operating mode distribution for the MOVES emission analysis. Specifically, a vehicle specific power (VSP) based model is used along with vehicle acceleration data to determine the operating mode distribution in the algorithm.

The presented framework is validated using video based vehicle trajectory data. The traffic parameters generated by the framework have the same pattern as it is revealed by the video based ground truth data. A case study is presented to demonstrate the application of the presented framework. It is found that the impact of traffic flow operation on vehicle emission along a specific roadway section can be associated with three quantified traffic flow variables, i.e., operating mode distribution, traffic volume, and traffic fleet composition. On the other hand, the data analysis shows that traffic flow phase can be mathematically featured with those traffic parameters. The connection between traffic operation and vehicle emission impact has been established through the application of the three traffic parameters.

This study makes it feasible for a project level mobile source emission impact study to be performed by using microscopic real-world traffic data. In the future, more comprehensive ground truth data are needed to further validate the proposed methodology. It is also desirable to expand the methodology in order to take advantage of other similar traffic data sources such as radar data and video detection data sources.

Table of Contents

DISCLAIMER **3**
ACKNOWLEDGEMENT **4**
EXECUTIVE SUMMARY **5**
CHAPTER 1: INTRODUCTION **1**
1.1 BACKGROUND 1
1.2 GOAL AND OBJECTIVES 3
CHAPTER 2: LITERATURE REVIEW **4**
2.1 DUAL-LOOP DATA SCREENING 4
2.2 TRAFFIC FLOW PHASES IDENTIFICATION 6
2.3 TRAFFIC INPUTS FOR MOVES 7
CHAPTER 3: GENERATING MOVES INPUTS FROM DUAL-LOOP DATA **10**
3.1 STUDY SITE 10
3.2 DATA COLLECTION 11
3.3 DUAL-LOOP DATA SCREENING 12
3.3.1 Dual-loop Data Screening Filters *12*
3.3.2 Dual-loop Data Screening Results *17*
3.4 TRAFFIC FLOW PHASE IDENTIFICATION 18
3.5 MOVES INPUTS CALCULATION 23
3.6 MODEL VALIDATION 28
3.6.1 Validation Data *28*
3.6.2 Validation Results *29*
3.7 CASE STUDY 30
CHAPTER 4: SUMMARY AND CONCLUSIONS **36**
REFERENCES **37**

List of Figures

FIG. 1. SKETCH OF DUAL-LOOP DETECTOR STATION 2
FIG. 2. FRAMEWORK OF TRANSFORMING DUAL-LOOP DATA TO MOVES INPUTS 10
FIG. 3. LOOP STATION V1002 ON I-70/71 IN DOWNTOWN COLUMBUS, OH 11
FIG. 4. VIDEOTAPING AT THE SELECTED DUAL-LOOP STATION 11
FIG. 5. ALGORITHM FOR 0-1 FILTER 13
FIG. 6. ALGORITHM FOR MINIMUM HEADWAY FILTER 14
FIG. 7. ALGORITHM FOR PARING FILTER 15
FIG. 8. ALGORITHM FOR SENSITIVITY FILTER 16
FIG. 9. ALGORITHM FOR MEDIAN SPEED FILTER 17
FIG. 10. TRAFFIC FLOW IDENTIFICATION FOR INDIVIDUAL VEHICLE 19
FIG. 11. SCHEMATIC OF TRAFFIC FLOW PHASES IDENTIFICATION ALGORITHM 19
FIG. 12. STEP 1 OF TRAFFIC FLOW PHASE IDENTIFICATION ALGORITHM 20
FIG. 13. STEP 2 OF TRAFFIC FLOW PHASE IDENTIFICATION ALGORITHM 20
FIG. 14. TEMPORAL CHANGE OF TRAFFIC FLOW PHASES 23
FIG. 15. ALGORITHM FOR OPERATING MODE DISTRIBUTION CALCULATION 28
FIG. 16. OPERATING MODE DISTRIBUTION FROM VIDEO DATA 29
FIG. 17. OPERATING MODE DISTRIBUTION FROM DUAL-LOOP DATA 29
FIG. 18. COMPARISON OF OPERATING MODE DISTRIBUTION AND TRAFFIC FLOW PHASES 31
FIG. 19. RELATIONSHIP OF TRAFFIC VOLUME, VEHICLE COMPOSITION AND EMISSION 32
FIG. 20. OPMODE DISTRIBUTION AND PM FOR WESTBOUND LEFT LANE, JULY 14, 2009 34
FIG. 21. OPMODE DISTRIBUTION AND PM FOR WESTBOUND RIGHT LANE, JULY 14, 2009 35

List of Tables

TABLE 1. EXEMPLARY SAMPLE OF THE EVENT DUAL-LOOP DATA 2
TABLE 2. SPEED COMPARISON BETWEEN RAW DATA AND SCREENED DATA 17
TABLE 3. ON-TIME COMPARISON BETWEEN RAW DATA AND SCREENED DATA 18
TABLE 4. OPERATING MODE BINNING DEFINITION (EPA, 2009) 27
TABLE 5. OPERATING MODE DISTRIBUTION COMPARISON 30

CHAPTER 1: INTRODUCTION

1.1 Background

Transportation has been proven one of the most stubborn challenges in reducing carbon emissions such as toxic carbon monoxide (CO) and carbon dioxide (CO_2) (Millard-Ball, 2008; Duhme et al., 1998). Transportation greenhouse gases (GHGs) emissions are primarily in the form of CO_2 and it is the fastest growing source of GHG emissions in the world (Sperling 2006). Motor vehicles account for more than 40% of GHG emissions in California, and 28% in the entire United States (California EPA, 2006; Energy Information Administration, 2006). Most of the existing emission models such as MOVES (Motor Vehicle Emission Simulator) use the results from travel demand models as partial traffic input in estimating regional mobile source emissions. Some major output emission parameters involved in the model are determined primarily based upon laboratory engine dynamometer experiments. However, these methods offer little linkage between the emission estimates and the real-world duty cycles that vehicles experience on roads (Zhai et al. 2008; Younglove et al. 2005). Localized analysis of on-road traffic source emissions is often required by the U.S. Environmental Protection Agency (EPA) for project-level conformity analysis of transportation projects in accordance with the Phase Implementation Plan (SIP). In public environmental health field, some healthy concerns have been addressed on their association with mobile source air pollution. For instance, infants living near traffic or elementary students at schools near heavy-duty transportation infrastructures are particularly susceptible to ambient air pollutants (Ryan et al. 2008; Ryan et al. 2007; Brunekreef et al., 1997). However, studies on levels of exposures to mobile source emissions and their influence on children's health obviously lack local micro-scale data to reflect the characterization of variability in emissions under various representative real-world vehicle activities and traffic conditions (Zhai et al., 2008).

It has been widely recognized that the carbon emissions are associated with the traffic operation conditions (Zhai et al., 2008; Younglove et al., 2005; Jemenez-palacios, 1999; Nesamani and Subranmanian, 2006; Frey et al., 2007; Frey et al., 2006). In the latest EPA's vehicle emission simulator, MOVES2010a, traffic operation conditions are represented by several factors, namely traffic volume, vehicle composition and operating mode distribution. Operating mode distribution represents percentages of various vehicle emission-producing activities on a specific road segment. It is calculated using vehicle speed, vehicle acceleration and Vehicle Specific Power (VSP) across the entire vehicle fleet. In previous studies, Lin et al. (2011) and Xie et al. (2011) applied simulation experiments to calculate those parameters. The accuracy of the simulation in representing real-world situations under various operations (e.g., free flow and congested traffic) is often a challenging issue. The key problem hence lies in the availability of operating mode distribution, traffic volume and vehicle composition for MOVES analysis based on available real-world data.

Dual-loops (in-pavement sensors, as Fig. 1 shows) are widely utilized in collecting continuous traffic data (Coifman 1999; Coifman 2004; Nihan et al. 2006; Coifman and Kim 2008). The event dual-loop data is referred to as a kind of high-resolution data of the detected individual vehicles (Table 1). The primary components of the event data include the

timestamps of vehicles arriving and leaving the M and S loop. Event dual-loop data is usually applied to traffic analysis in order to obtain accurate travel features as individual vehicles traveling over the loop station (Nihan et al., 2006; Coifman et al., 2004). The event dual-loop data could be utilized as a rich data source for calculating traffic inputs for MOVES. However, little research has been reported on it. One main reason lies in the incapability of producing accurate vehicle classifications by current dual-loop models.

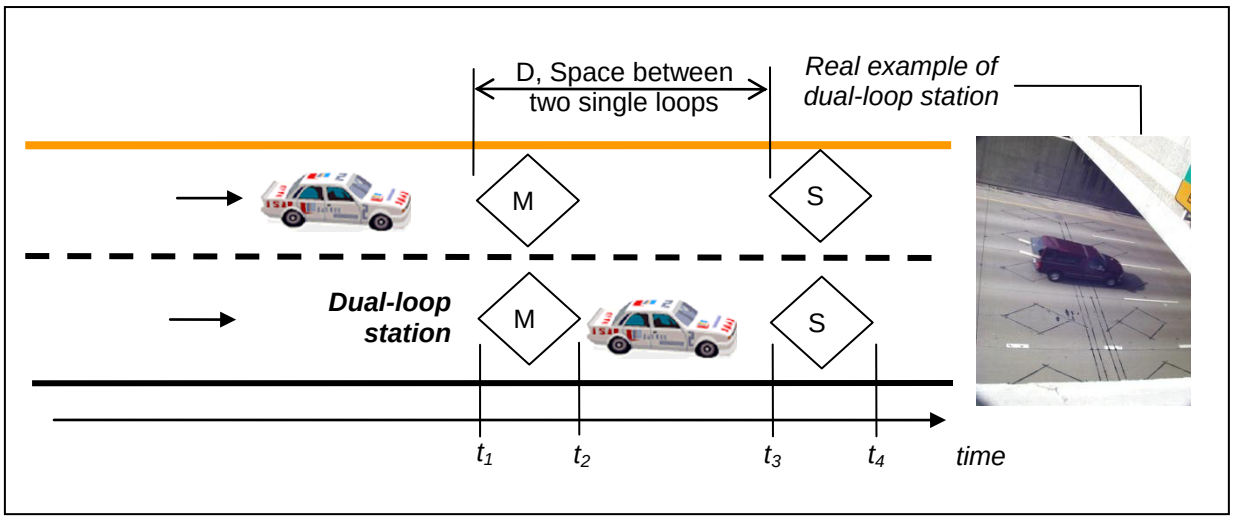

Fig. 1. Sketch of Dual-loop Detector Station

Table 1. Exemplary Sample of the Event Dual-loop Data

M loop (Upstream)		S loop (Downstream)	
Status	Timestamp	Status	Timestamp
1	3522267	1	3523667
0	3524341	0	3524489
1	3524504	1	3524652
0	3524675	0	3524795
1	3524817	1	3524919
0	3525598	0	3525914

The capability of measuring vehicle lengths makes dual-loop detectors a potential real-time data source for vehicle classifications (Nihan et al. 2006; Coifman 1999; Coifman 2004a; Coifman and Kim 2008). While most existing loop models for measuring speeds and vehicle classifications have been proved accurate against light traffic, they are not reliable in accuracy under other traffic conditions like synchronized or traffic jam congestions. Fortunately, the project conducted by the PI and collaborator, titled “Optimal Loop Placement and Models for Length-based Vehicle Classification and Stop-and-Go Traffic” (termed as the “OTC vehicle classification” project in this proposal), which is funded by OTC, has resulted in positive solutions to ensure the accuracy. That makes us to hold the technical promise for developing models of estimating micro-level emissions under various traffic operations by

using dual-loop data.

1.2 Goal and Objectives

The goal of this research project is to develop a methodological framework for integrating the improved dual-loop models and the MOVES emission model to estimate the emission impact of traffic flow operation by utilizing the data source available at dual-loop monitoring stations in highways.

Due to improper dual-loop sensitivity configuration malfunction of loops, and/or ambient environmental impact, dual-loop raw datasets often contain various errors such as data missing or data discontinuity. Those errors could impair or even make it impossible to extract reliable traffic parameters for MOVES analysis from the dual-loop data. Although missing data was usually coped with in engineering statistical approach in practice, it is not sufficient to satisfy the accuracy requirements of operating mode calculation. Therefore, a dual-loop data screening algorithm is necessary to filter out erroneous data records.

Another challenge remaining in fulfilling the goal is associated with traffic flow phases identification problems with use of dual-loop data. Traffic phases identification is important for this research because the successful application of the length-based vehicle classification algorithms proposed by OTC vehicle classification project is dependent upon accurate identifications of traffic flow phases. The vehicle classification data is in turn the basis for estimating vehicle composition and operating mode distribution for MOVES emission analysis. Therefore, traffic flow phase identification problem becomes a mandatory target in addition to the vehicle classification modeling. In the wake of such an understanding, an effort in developing quantitative traffic flow phases identification algorithm is incorporated into the project tasks.

Another critical challenge as identified for this project is the lack of an available model for automatically generating traffic inputs for MOVES from dual-loop data source. The reason lies in the difficulty to produce vehicle acceleration from dual-loop data source due to the precision limitation of the data. Thus, an algorithm for transforming dual-loop data into the required traffic inputs to the vehicle emission model is developed through the project.

Based on the above understanding, the objects of this project are designated as follows.

1) To screen raw dual-loop data so that the dual-loop data are prepared for the next two algorithms;
2) To develop a traffic phases identification methodology and an associated computing algorithm which could be used to identify unique features of the traffic operation in varied traffic flow phases ;
3) To develop a computing algorithm to transform traffic inputs for MOVES from dual-loop datasets.

CHAPTER 2: LITERATURE REVIEW

2.1 Dual-loop Data Screening

The performance of dual-loop based traffic models is greatly affected by the dual-loop data errors. Many researchers and practitioners have tried to filter out dual-loop data errors using aggregated or event dual-loop data. Those methods are briefly introduced in this section.

Ishak (2003) proposed a methodology to filter out erroneous aggregated dual-loop data using fuzzy K-means clustering algorithm. The algorithm is developed based on the fact that a large dataset is able to show variation pattern of the real situation. In this study, a large dual-loop dataset was collected continuously in a year and used as the ground truth dataset for the presented algorithm. Then observed data was studied by first determining the number of clusters and cluster centroids based on the ground truth dataset. In the observed dataset, valid observations have high membership level to one cluster while erroneous observations do not have high membership to level any cluster. The valid observations are separated from erroneous observations by examining the membership level. The major drawback of this method is that huge amount of data are needed in the beginning to construct the ground truth dataset. It is not feasible for projects with limited data sources.

Vanajakshi and Rilett (2004) presented a dual-loop data filtering algorithm based on the principle of conservation of vehicles. The principle stipulates that the cumulative entering traffic volume for a section of road is always equal or larger than the exiting volume. The difference of the entering and exiting volumes should not be larger than the maximum vehicles that can be accommodated by the road section. He argued that classic dual-loop data error checking algorithms mainly focused on individual dual-loop station. Passing error checking at individual dual-loop station does not guarantee that the principle of conservation of vehicles is met at several consecutive dual-loop stations. Based on the idea, a Generalized Reduced Gradient (GRG) optimization algorithm was proposed to adjust flow output of a series of dual-loop stations. It was assumed that the vehicle conservation error was evenly distributed among all the study loops and all the study loops are adjusted equally. However, such ‘average’ adjustment method may cause the results of good loops over-adjusted and results from bad loops under-adjusted.

Al-Deek et al. (2004) presented an algorithm to flag out bad data samples from both 30 seconds and 5 minutes aggregated dual-loop data. A set of filtering criteria, such as maximum possible speed and flow and occupancy limitations, were used for screening the 30 seconds data. After the 30 seconds data was filtered, it was aggregated in 5 minutes interval and was input into an average vehicle length based algorithm for screening. After erroneous data examination, a pair-wise regression model was used to impute the bad data samples. The data from adjacent lanes and neighboring loop stations was used to develop a series of regression models. The median values of the outputs from the regression models were used to replace the bad data samples.

Above studies applied aggregated dual-loop data to address data errors. However, Cheevarunothai et al. (2007) argued that the ability of aggregated methods is limited because

those methods only use partial information from the dual-loop data. Information associated with individual vehicle is lost during the aggregation process. Therefore, dual-loop data screening and imputation using event loop data is developed by many studies.

Coifman and Dhoorjaty (2004b) proposed eight test algorithms to filter out erroneous loop records using event loop data. They reported that seven different errors that could be found from dual-loop data. Among them, premature rising edge and delayed falling edge could extend occupancy measurement. Delayed rising edge, premature falling edge, and flicker (turning off and back on in the middle of a vehicle) could shorten the occupancy measurement. A missed vehicle or a false detection in the absence of a vehicle could result missing or extra data records. The proposed test algorithms apply moving median speed, the boundary and variation of on time, headway limitation, vehicle length limitation and consecutive congestion samples as filtering criteria.

Cheevarunothai et al. (2006) presented that dual-loop sensitivity problems are main causes of inaccuracy of current WDOT vehicle classification algorithm. The presented dual-loop screening model filters out the sensitivity problem by first adjusting the sensitivity of two loops to the same level; and then adjusting both loops' sensitivity to the correct level. The correct level was obtained based on the observed ground truth vehicle length distribution for short vehicles (i.e., passenger cars). The length of short vehicle ranges from 9 feet to 25 feet and all short vehicles were divided into 17 classes (e.g., 9 feet vehicles belong to class 1, 10 feet vehicles belong to class 2, etc.). An indicator named sum of square errors was used to represent the difference of dual-loop output and the ground-truth (equation 1).

$$\text{sum of square errors} = \sum_{i=1}^{17}(\text{observed}_i - \text{expected}_i)^2 \quad (1)$$

In equation 1, 'observed_i' represents the number of vehicles in class i observed from dual-loop data; 'expected_i' represents the number of vehicles in class i from observed data. If the sum of square errors is larger than a threshold (400 in the study), the sensitivity level for M and S loops is not proper and thus needs to be adjusted.

Cheevarunothai (2007) identified four major problems that affect dual-loop data accuracy. Those problems were "(a) split of loop signals for multiunit trucks, (b) cross talk between adjacent loops, (c) constrained sensitivity adjustments of discrete levels at loop amplifiers, and (d) unsuitable thresholds of on-time differences between the two single loops", respectively. In order to sovle problem (a), it is assumed that time headway of two distinct vehicles should be larger than 0.63 seconds. Two records will be combined if their time headway is less than the threshold. In order to address the cross talk problem, the adjustment rule stipulated that a vehicle's entering time on M loop must be earlier than its entering time on S loop. Another adjustment rule for the cross talk problem stipulated that actuations of a loop cannot occur when loop at adjacent lane is actuated at approximately the same time unless match actuations can be found from the paired loop in the adjacent traffic lane. The error screening method for the sensitivity problem could be found elsewhere (Cheevarunothai et al., 2006). The threshold problem in problem (d) was caused by vehicle lane changing. The record of the vehicle was discarded accordingly.

2.2 Traffic Flow Phases Identification

Highway capacity manual (TRB, 2000) defines level-of-service (LOS) concept and classifies six levels of traffic flow phases on expressway, ranging from LOS A through LOS F. Traffic flow density is used as a sole variable to quantify traffic flow phases in terms of LOS. The congested traffic flow is represented by LOS F. However, Kerner (1996a) proposed that there are actually two traffic conditions under congested traffic, namely synchronized flow and traffic jam, respectively. Under synchronized traffic, average speeds of vehicles on different lanes are almost identical because the density is so large that vehicles within a fleet cannot find gaps to change lanes. The phenomenon of traffic jam is similar to that of synchronized flow in view of their common features of low speed and high density. The difference between them is that traffic jam has negative slope of time sequence (i.e., plotting flow-density points in time series) on flow-density plane and the time sequence of the synchronized flow is randomly changed. In addition, Kerner examined the characteristics of congested traffic and identified the features and causes of moving jam (Kerner, 1996b) and the stop-and-go traffic (Kerner, 1998). He argued that the complexity of congested traffic is caused by the temporal-spatial distribution of three traffic phases (i.e., free flow, synchronized flow and traffic jam, respectively) and the transitions among these conditions. He also (Kerner 1997; 1999) explored the features of phase transitions and proposed that the nucleation effect is responsible for the transition from free flow to synchronized flow and jam formation (free flow to traffic jam or synchronized flow to traffic jam).

Kerner categorized traffic flow conditions into free flow, synchronized flow and traffic jam. In addition, transitions from free flow to synchronized flow, transition from synchronized flow to traffic jam and transition from free flow to traffic jam are important parts of traffic flow conditions. Bank (1999) used the flow-occupancy ratio as an indicator to separate the synchronized flow and traffic jam. His results showed that the traffic phases similar to Kerner's definitions were verified. The traffic phase below flow-occupancy ratio of 55 was similar to traffic jams, with negative transference in flow-occupancy plane. However, the condition above flow-occupancy ratio of 55 had generally positive transference in flow-occupancy, which was different with Kerner's observation that the transference was random within the synchronized flow. Bank further argued that such slope difference could be caused by the variation in time gaps and the relationship among speed, flow, density and average time gap.

Although dispute exists, classifying traffic into three categories (i.e., free flow, synchronized flow and traffic jam) is widely accepted. A promising way to quantitatively identify the traffic phases is to use the clustering method. Azimi and Zhang (2010) noted that the "clustering is an unsupervised learning method that assigns observations into different groups or clusters. Its goal is to attribute multidimensional observations into different clusters such that points within a cluster are similar and differ from points in other clusters." There are three widely used clustering methods, namely, K-means, fuzzy C-means and CLARA (Azimi and Zhang, 2010). K-means method is the most commonly used method since it can produce reasonable results and also can be easily understood (Azimi and Zhang, 2010; Xia and Chen, 2007a; Xia and Chen, 2007b; Sun and Zhou, 2005). The results of the clustering methods are validated by comparing their results with the LOS classification results. It was found that the

K-means method can produce results that were mostly close to the LOS classification outcomes. In addition, the clustering method can further breakdown the congested traffic condition (e.g., synchronized flow and traffic jam), while the LOS method is incapable to distinguish the traffic conditions that are worse than LOS F (Azimi and Zhang, 2010).

Identification of clustering variables and cluster number is the major issue that should be emphasized in applying the clustering method. Usually basic traffic parameters, such as speed, density, occupancy, and volume, are used as the clustering variables (Azimi and Zhang, 2010; Xia and Chen, 2007a; Xia and Chen, 2007b; Sun and Zhou, 2005). In order to determine number of clusters, some researchers rely on engineering experience. For example, Azimi and Zhang (2010) used 6 clusters based on the LOS concepts. Xia and Chen (2007a) used five clusters: two for free flow condition, one for transition and the last two for congested condition. Sun and Zhou (2005) employed two and three clusters, when they developed the multi-regime speed-density model. Other researchers applied statistical method to determine the number of clusters (Xia and Chen, 2007b). However, the statistical method is incapable of linking the clusters to the real-world traffic conditions from their inherent features and rationale.

Another interesting finding on the clustering method is that the results of the clustering method are very close to the LOS classifications when traffic is in free flow condition (Azimi and Zhang, 2010). When traffic phase begins to change (i.e., from free flow to synchronized flow or when LOS is E), the results of clustering method exhibit different patterns comparing with the LOS classification (Xia and Chen, 2007a). Since the LOS method cannot take into account differences of traffic phases when traffic becomes congested, it is therefore more proper to use the clustering method in such a case. In general, traffic phases identification criteria based on traffic flow parameters work well under non-congestion traffic conditions. When traffic becomes congested, the interaction of traffic parameters become complicated so that it is hard to determine boundaries between different traffic phases. In this case, the clustering method is a better option since the method finds boundaries according to statistical characteristics of the study data.

2.3 Traffic Inputs for MOVES

Many researchers and practitioners have studied ways to generate traffic inputs for MOVES using various transportation models. Lin et al. (2011) proposed a framework that extracts traffic inputs for MOVES from a dynamic traffic assignment model, DynusT. Link drive schedule, operating mode distribution and link source type hour are generated from DynusT for MOVES analysis. However, no information is available in this study about how the validation of the proposed model is conducted. Xie et al. (2011) presented a framework integrating MOVES and traffic simulation tool Paramics for the purpose of studying vehicle energy consumption. In the proposed procedure, Paramics outputs including link volume, speed and vehicle composition were used for MOVES inputs. Three levels of inputs, namely average speed, link drive schedule and operating mode distribution, can be used in MOVES for defining emission-producing vehicle activities. Xie et al. used link average speed, which led relatively coarse emission and energy consumption results in comparison to the results when link drive schedule or operating mode distribution is used.

Scora et al. (2011) applied a computer vision-based system, called VECTOR, to extract traffic information such as vehicle classification, speed, and acceleration to calculate traffic inputs for MOVES. It provides a promising way to study mobile source emission impact if video sensors are widely deployed. Shi and Yu (2011) applied a second-by-second VSP model to study the vehicle emission. The model was built based on emission and vehicle operation data measured by a portable emission system. Because the portable emission system collected emission data for only one vehicle type, the differences in emission between various vehicle classes could not be determined by the model.

A dual-loop station records both the times (timestamps) when a vehicle enters and leaves either of the two inductive loop detectors which are placed 20 feet apart in Ohio. In a traditional dual loop model, the dual loop station can only produce one speed record when a vehicle passes it (Wei et al., 2010), while two speed records are needed to calculate the vehicle's acceleration. In order to obtain acceleration input, one possible way is to break a dual loop station into two single loop stations and apply two speeds estimated from each single loop for calculating acceleration. However, the single loop station itself is not an accurate enough to produce vehicle speed data because it cannot determine vehicle length, which is important in the accurate speed calculation (Coifman, 2001). As opposed to the single loop, the dual loop station has the ability to produce vehicle length information. The length data produced by dual loop station can be used for each single loop to help produce speed records with increased accuracy. Wei et al. (2010) proposed advanced models to estimate vehicle length under both free flow and congested traffic flow conditions. It has been proven that those models can derive accurate vehicle length results even when traffic flow is congested.

According to Wei et al. (2010), the vehicle length can be calculated using equation (2) and (3) when traffic is under free-flow condition.

$$speed = \frac{D}{t} \tag{2}$$

$$vehicle\ length = speed \times \frac{onT_1 + onT_2}{2} - loop\ length \tag{3}$$

Where:

D = distance between two single loops in the dual-loop station (ft);

$t = t_3 - t_1$;

$OnT_1 = t_2 - t_1$; and

$OnT_2 = t_4 - t_3$.

t_1, t_2, t_3, and t_4 = timestamps when a vehicle enters or leaves the upstream loop (M loop) or downstream loop (S loop).

The vehicle length can be calculated using equations (4) to (6) when it is synchronized flow.

$$L_v = v_0 \cdot onT_1 + \frac{1}{2}a(onT_2)^2 - L_s \tag{4}$$

$$v_0 = \frac{D}{t} - \frac{a \cdot t}{2} \quad (5)$$

$$a = \frac{D}{t}\left[\frac{2 \cdot (onT_1 - onT_2)}{(onT_2)^2 - (onT_1)^2 + (onT_1 + onT_2) \cdot t}\right] \quad (6)$$

Where:

L_v = length of the detected vehicle (ft);
L_s = length of each single loop within the dual-loop (ft);
v_0 = speed of the vehicle entering the upstream loop (M loop) (ft/s);
a = vehicle acceleration (ft/s2); and
D, t, OnT_1, and OnT_2 are the same as defined earlier in the previous equation.

The vehicle length can be calculated using equations (7) to (9) when it is traffic jam.

$$L_v = f_1 \cdot t_{dec} \cdot D \cdot \frac{1}{t} + \frac{1}{2} \cdot f_2 \cdot a \cdot t_{acc}^2 - L_s \quad (7)$$

$$t_{dec} + t_{acc} = OnT_1 - t_s \quad (8)$$

$$t_s = t_2 - t_3 - f_3 \cdot \frac{t_{acc}^2}{v_{min}} \quad (9)$$

Where:

Lv = length of vehicle (ft);
Ls = length of each single loop within the dual-loop (ft);
t_{dec} = time period from a vehicle entering the M loop to its stop (s);
t_{acc} = time period from a vehicle starting to move to leaving the M loop (s);
a = the average acceleration rate of vehicles when they start to move under stop-and-go traffic (ft/s2);
t_s = time period for a vehicle stopping on both loops (s);
v_{min} = the minimum speed which can maintain a vehicle running without stop (ft/s);
f_1, f_2, and f_3 = adjusting factors for different vehicle types (in this study, $f_1 = f_2 = f_3 = 1$);
D, t, t_2, t_3, OnT_1, and OnT_2 = as the same as defined previously.

CHAPTER 3: GENERATING MOVES INPUTS FROM DUAL-LOOP DATA

In this chapter, the algorithms for dual-loop screening, traffic flow phases identification and MOVES inputs transforming are developed. The traffic parameters resulted from dual-loop data are then compared against the ground-truth data for the purpose of validation. Video data is used as the ground-truth data in this research. The VEVID software is used to process the video data. It extracts ground truth vehicle type, speed and acceleration to calculate operating mode distribution which is then compared with the distribution estimated by the presented model. After validation, a case study at a section of I-71 in Columbus Ohio is presented to illustrate the procedure of the proposed methodology. Fig. 2 illustrates the framework of the proposed methodology.

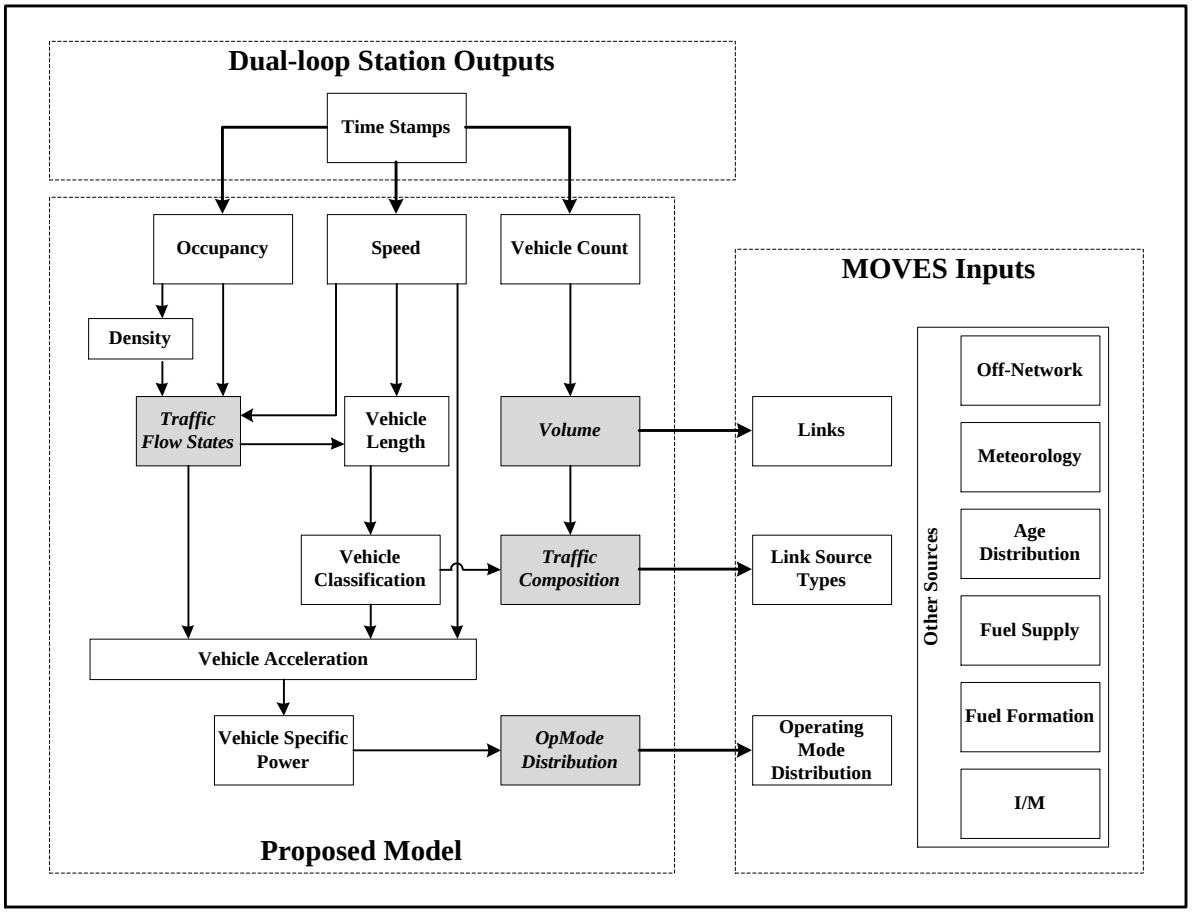

Fig. 2. Framework of transforming dual-loop data to MOVES inputs

3.1 Study Site

The data for was collected at dual-loop station, numbered V1002, in Columbus, Ohio. The dual-loop station is in good working condition and the traffic flows over the loops has been videotaped at the study site. As shown by Fig. 3 and 4, V1002 station is located in I-70/71 at West Mound Street within downtown Columbus. This station has 6 dual-loop detectors in both directions. The Franklin County Juvenile Parking Garage is near this station. A video camera is placed on the top floor of it. ODOT Traffic Management Center (TMC) is running

those stations and provides the loop data.

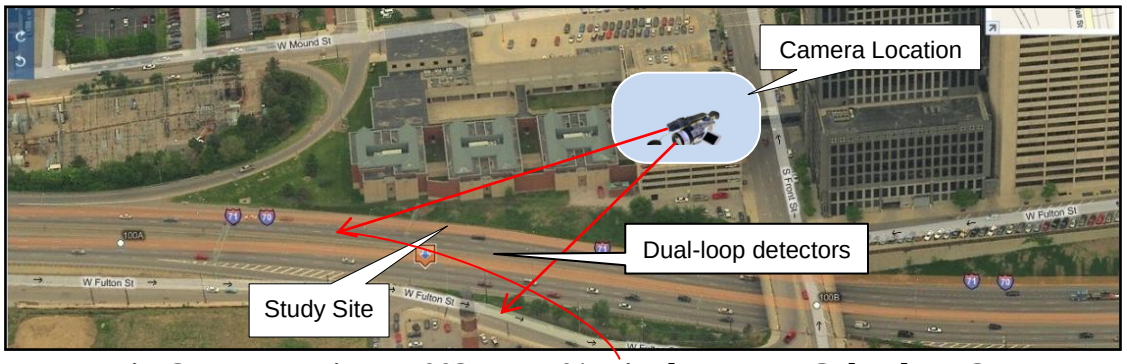

Fig. 3. Loop station V1002 on I-70/71 in downtown Columbus, OH

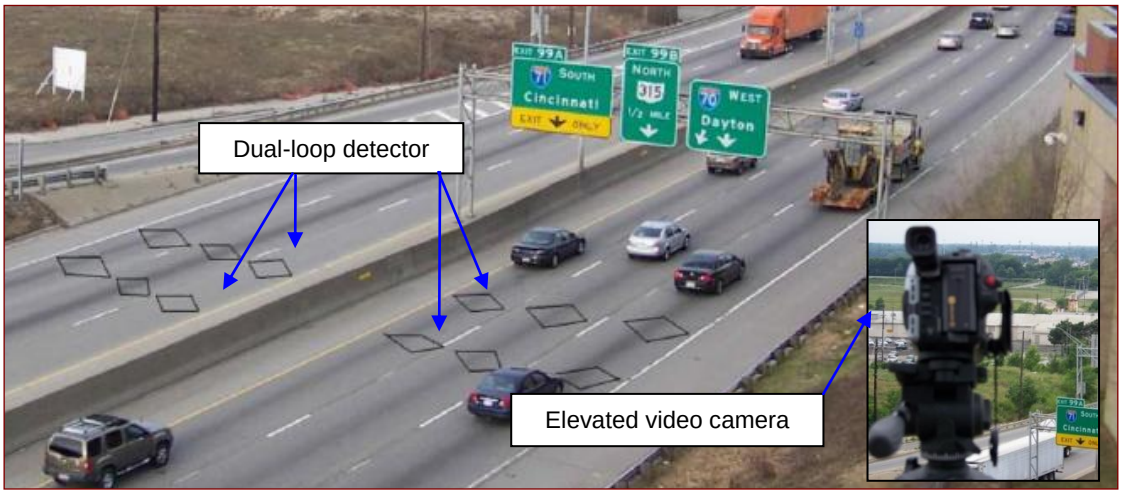

Fig. 4. Videotaping at the Selected Dual-loop Station

3.2 Data Collection

Dual-loop data was collected from July 14th to July 16th, 2009. The video data for the selected dual-loop station was collected using digital camcorders for the purpose of validation. Video data was collected on July 15th, 2009 from 15:25 to 16:25. There were 1699 vehicles during the study hour.

A video data extraction has been performed for this project. 178 vehicles, including 126 passenger cars and 52 heavy-duty-vehicles, were extracted out of 1695 vehicles from the one hour video data. The sampling rate is 10.5%. The one hour video data was divided into 52 clips. Each clip had 2000 video frames. In order to avoid sampling bias, two sampling methods were used. The first method was used to extract 4 vehicles records, including speed, acceleration and vehicle type, from 1 video clip. The first heavy-duty-vehicle appearing in the clip was extracted and the rest three vehicles were passenger cars that appeared at 500th, 1000th and 1500th frame. Another method was applied to extracted 2 vehicles from 1 video clip. Similarly, the information for the first heavy-duty-vehicle appearing in the clip and the passenger car appearing at 1000th frame was extracted. Such sampling method ensures the randomness of the extracted data.

3.3 Dual-loop Data Screening

The purpose of the dual-loop data screening is to ensure that the dual-loop data is able to be applied for the traffic flow phases identification and MOVES inputs transforming algorithm. There are six filters for the presented dual-loop data screening algorithm. They are 0-1 filter, minimum headway filter, pairing filter, sensitivity filter, on-time filter and median speed filter, respectively. The filters are used following the listed order because the later filter must be applied based on results from the previous filters. The mechanism of the filters is that the dual-loop data should be arranged in a sequence of arriving and leaving timestamps. Arriving time must be earlier than leaving time. The records for upstream and downstream loops are paired based on the sequence of their timestamps. In addition, the on-time and headway must be within reasonable limits.

3.3.1 Dual-loop Data Screening Filters

0-1 Filter**:**

The 0-1 filter is designed to eliminate dual-loop communication errors. Due to communication errors the raw dual-loop data may exhibit leaving timestamp (timestamp 0) appears ahead of arriving timestamp (timestamp 1). It may also cause that several 1s or 0s appear continuously. In order to eliminate the communication error, the 0-1 filter applies the algorithm in Fig. 5.

The minimum headway filter is used to eliminate dual-loop records whose estimated time headway values are less than a threshold. According to Cheevarunothai et al.'s study (2007), two vehicles' time headway should be larger than 0.63 second. If an estimated headway is less than the threshold, the two corresponding records are considered to represent one vehicle (e.g., a truck with a trailer). Then those records are combined. Fig. 6 shows the minimum headway filter algorithm.

Paring Filter:

The paring filter eliminates cross talk or lane changing errors. Cheevarunothai et al. (2007) presented that the cross talk error is erroneous activation of a loop detector when vehicles pass the loop's adjacent lane or its upstream (or downstream) loop. It causes M (or S) loop has vehicle record while the other loop does not have the corresponding record. The lane changing error occurs when vehicles change lanes as they travel through the loop station area. The error causes similar errors as cross talk error does. The paring filter eliminates cross talk and lane changing error and its algorithm is illustrated by Fig. 7. It pairs the upstream and downstream records using the following criteria:

- M loop arriving time is earlier than S loop arriving time;
- M loop leaving time is earlier than S loop living time;
- The paired records are the closest records. For example, an M loop record (called current M record) and an S loop record (called current S record) are considered. If the occurrence time of next M loop record is closer to the current S record than the occurrence time of the current M record, the next M record and the current S record are paired.

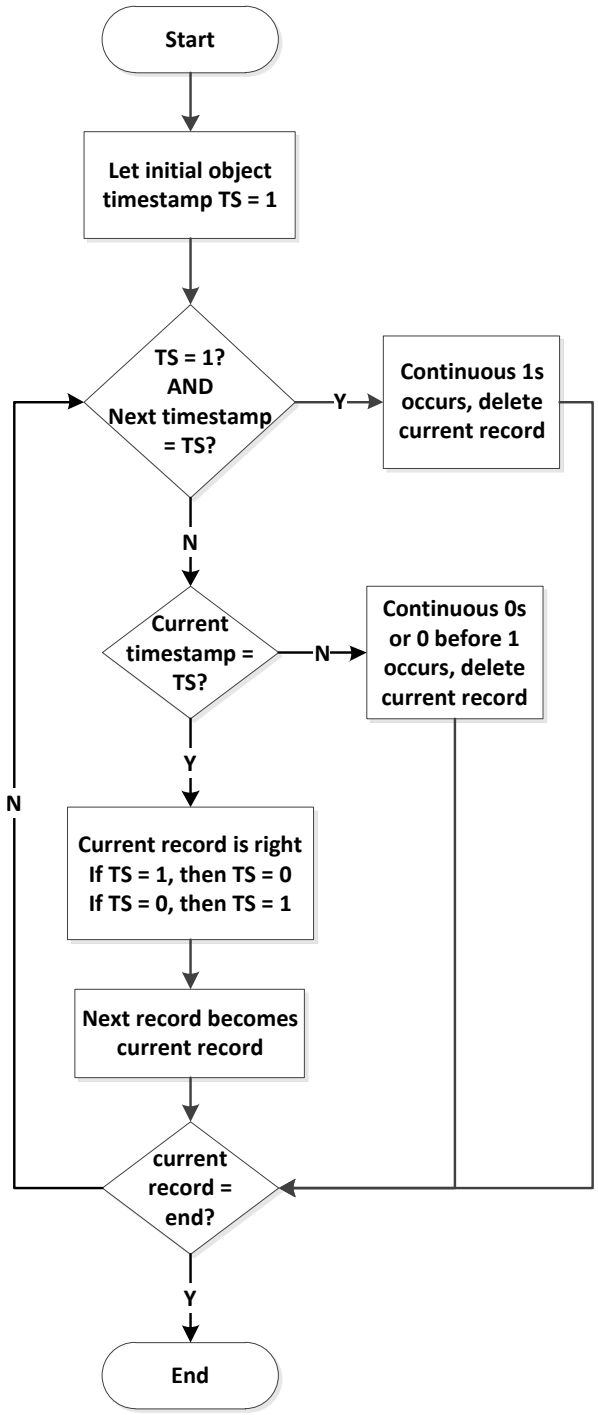

Fig. 5. Algorithm for 0-1 filter

Referring to Fig. 7, the M records and S records are all paired after paring filter is applied. The numbers of the upstream and downstream records are equal.

For a series of records, the algorithm is designated to look for timestamp 1 first and then locate timestamp 0. The search process continues until all records are addressed. An indicator TS is used to identify whether current target timestamp is 1 or 0. The continuous 0s and disordered 0-1 sequence can be filtered out by the indicator TS. The continuous 1s are identified using the timestamp of the next record. Note that the algorithm is applied to M loop data and S loop data separately. It is possible that the output data table has different record

numbers between the M and S loop's columns.

Minimum Headway Filter:

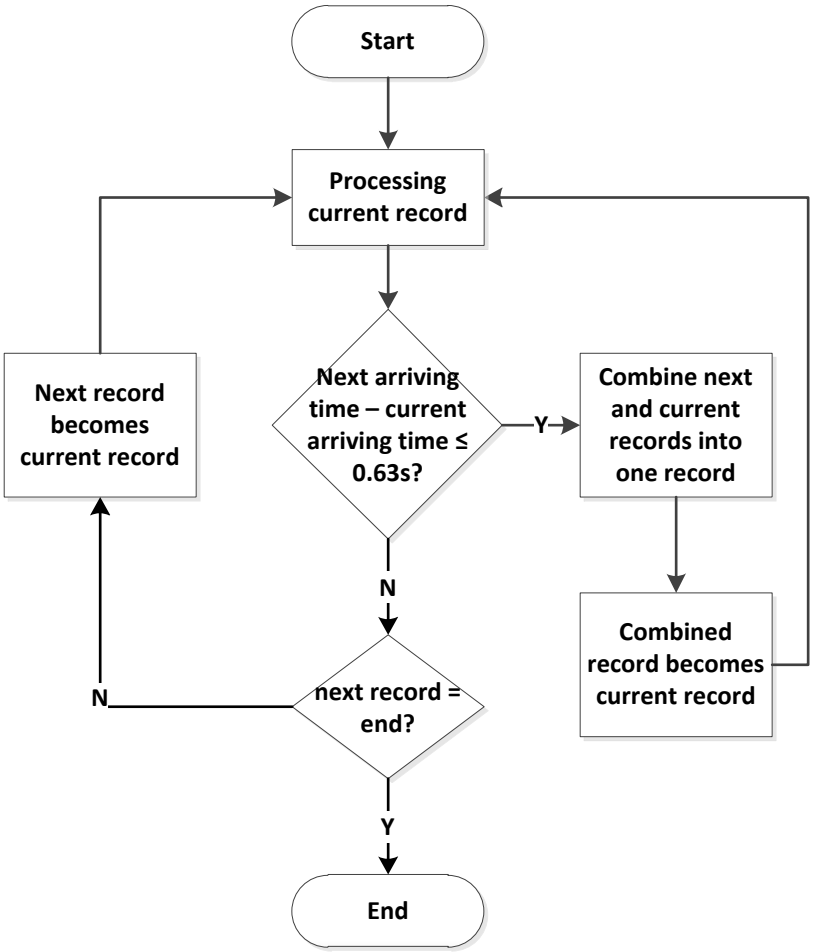

Fig. 6. Algorithm for minimum headway filter

On-time Filter:

The on-time filter eliminates records whose M loop on-time or S loop on-time exceeds boundaries. Coifman and Dhoorjaty (2004b) presented that the on-time boundaries can be calculated using the maximum and minimum possible speeds and the maximum and minimum possible vehicle lengths. It was calculated that the lower boundary of on-time is 0.16 second and the upper boundary of on-time is 1.3 seconds. The filter processes each record. If a record's M loop on-time or S loop on-time exceeds the boundary, the record is deleted.

Sensitivity Filter:

The sensitivity filter corrects possible sensitivity errors in dual-loop data. There are two kinds of sensitivity errors. One is caused by the inconsistency sensitivity levels between the M loop and S loop. Another is caused by the improper sensitivity levels for both loops. In this project, VEVID data are used as the ground-truth data to study the sensitivity errors. It is found that the S loop has proper sensitivity level while the M loop does not. Therefore, the M loop's records are adjusted according to S loop's data. The sensitivity filter uses all free-flow

records from the S loop data to find out the adjustment factor. The sensitivity filter algorithm is shown in Fig. 8.

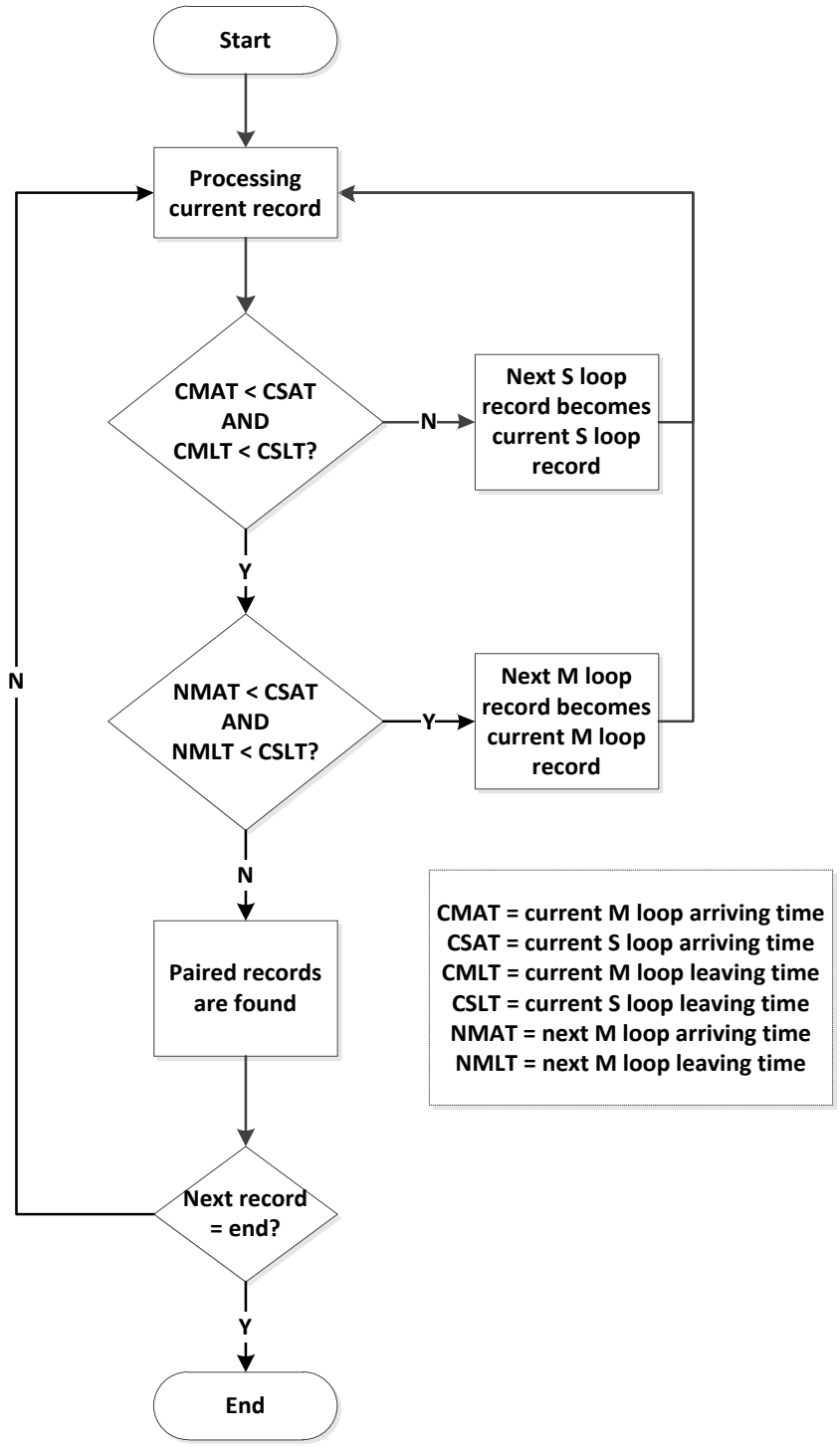

Fig. 7. Algorithm for paring filter

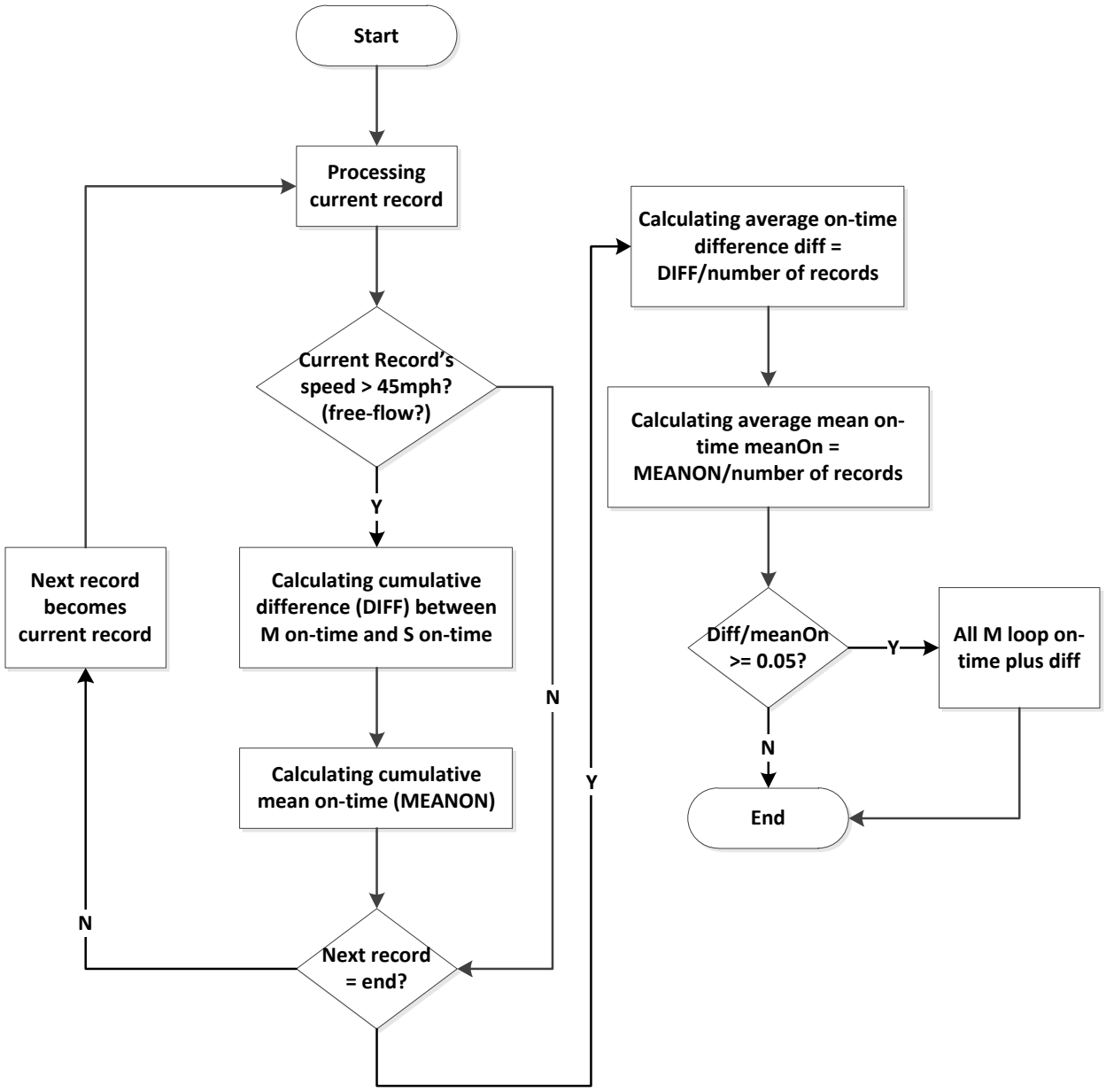

Fig. 8. Algorithm for sensitivity filter

Median Speed Filter:

The median speed filter eliminates records that result speeds 15 mph larger than the estimated median speed. The median speed is calculated using 11 vehicles' records centered at the subject record. We chose 11 vehicles because their median speed can represent the localized speed situation and they are not affected by the spatial changes of traffic condition. The moving median technique is used for calculating the median speed. The moving median concept is shown in Fig. 9.

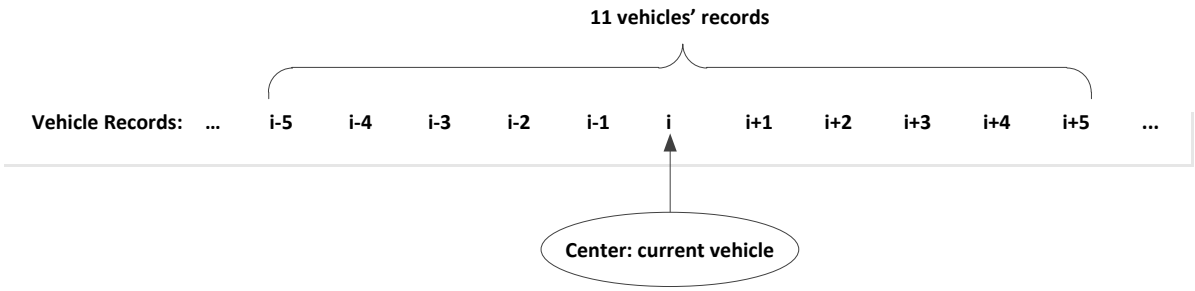

Fig. 9. Algorithm for median speed filter

3.3.2 Dual-loop Data Screening Results

The dual-loop data from midnight to 8:00 am on July 14, 2009 was used for testing the filters. No traffic congestion was observed from the study time period. In order to test the dual-loop data screening algorithm, several errors were added to the raw data on purpose. The locations of those errors were randomly selected. The added errors covered all error types mentioned in the previous section. The screening results show that the proposed error screening algorithm can eliminate the added errors completely.

The dual-loop datasets before and after filtering were compared. There were 5137 records in the raw data. After data screening, there were 4882 records remaining. The difference was 4.9%. The result indicates that the proposed data screening method can retain most part of the raw data while filter out bad data records.

For each dual-loop record, there are three time parameters can be calculated. The first one is the time difference between M loop arriving time and S loop arriving time; the second is the time difference between M loop leaving time and S loop leaving time; the third one is the average time of the above two values. Accordingly, three speed parameters can be calculated by dividing the distance between M and S loop by those three time parameters. The comparison of the three speeds is listed in Table 2.

Table 2. Speed comparison between raw data and screened data

(mph)	Raw Max	Screened Max	Raw Min	Screened Min	Raw Mean	Screened Mean
Speed #1	163.64	81.78	40.4	37.6	65.83	58.53
Speed #2	77.92	81.85	0.3	16.45	52.96	59.37
Speed #3	89.09	84.42	28.09	28.98	59.4	58.95

Since the selected data represent free-flow traffic condition. The three speed parameters estimated from dual-loop data are expected to be similar. However, the results from raw data show large variation. On the other hand, the filtered data shows good consistency. The data screening algorithm also eliminates unreasonable outliners. The maximum speed estimated from raw data is 163.64 mph, which is not possible for normal vehicles on the expressway. The minimum estimated speed from raw data is 0.3 mph, which

cannot usually be observed during the free-flow traffic. On the contrast, the filtered data shows more reasonable results, indicating the presented filtering algorithms are able to eliminate extreme values.

In terms of the on-time difference, Table 3 shows the comparison results. It is expected that the M loop on-time and S loop on-time are similar under the free-flow condition. However, the raw data shows large difference. After filtering and eliminating unreasonable records, the filtered data shows very consistent on-time data.

Table 3. On-time comparison between raw data and screened data

(second)	M on-time	S on-time	Difference (M - S)
Raw data	0.3450	0.457283333	-24.56%
Screened data	0.4045	0.402916667	0.39%
Difference (Filtered - Raw)	17.25%	-11.89%	

3.4 Traffic Flow Phase Identification

Traffic flow phase identification algorithm provides critical input for the length-based vehicle classification models (equation 4 through 9). The vehicle length obtained from the models is important inputs when vehicle speed, acceleration and traffic composition are calculated from dual-loop data. The traffic flow phase identification algorithm can also determine traffic phase for each data record and make it possible to study traffic flow phase as a contributing variable when analyzing emission impact of various traffic flow operations.

Individual vehicle length is required from event dual-loop data for speed, acceleration and traffic composition calculation. However, traffic flow phases can only be identified for a time period using aggregated data. To fill in the gap, it is assumed that all vehicles in a data aggregation interval have the same traffic flow phase. Therefore, traffic flow phase for individual vehicle can be estimated according to the traffic flow phase of the specific aggregation interval where the vehicle stays. The concept is shown in Fig. 10.

A hybrid method is used to develop the traffic flow phase identification model. According to Kerner's studies (1997; 1998; 1999), traffic flow phases are classified as free flow, synchronized flow, traffic jam (stop-and-go), and free flow to synchronized flow (F→S) transition. The level of service (LOS) method is used in the first step to preliminarily classify free flow and non-free (congested) flow phases. The results of the step are candidate free flow records and raw non-free flow records. Speed threshold and historical record methods are then applied to the candidate free flow records to identify real free flow records and separate them from traffic jam records and synchronized flow records. Alternatively, the K-means clustering method is used to distinguish F→S transition, synchronized flow and traffic jam records from the raw non-free flow records. The schematic diagram of the proposed heuristic model is illustrated by Fig. 11 to 13.

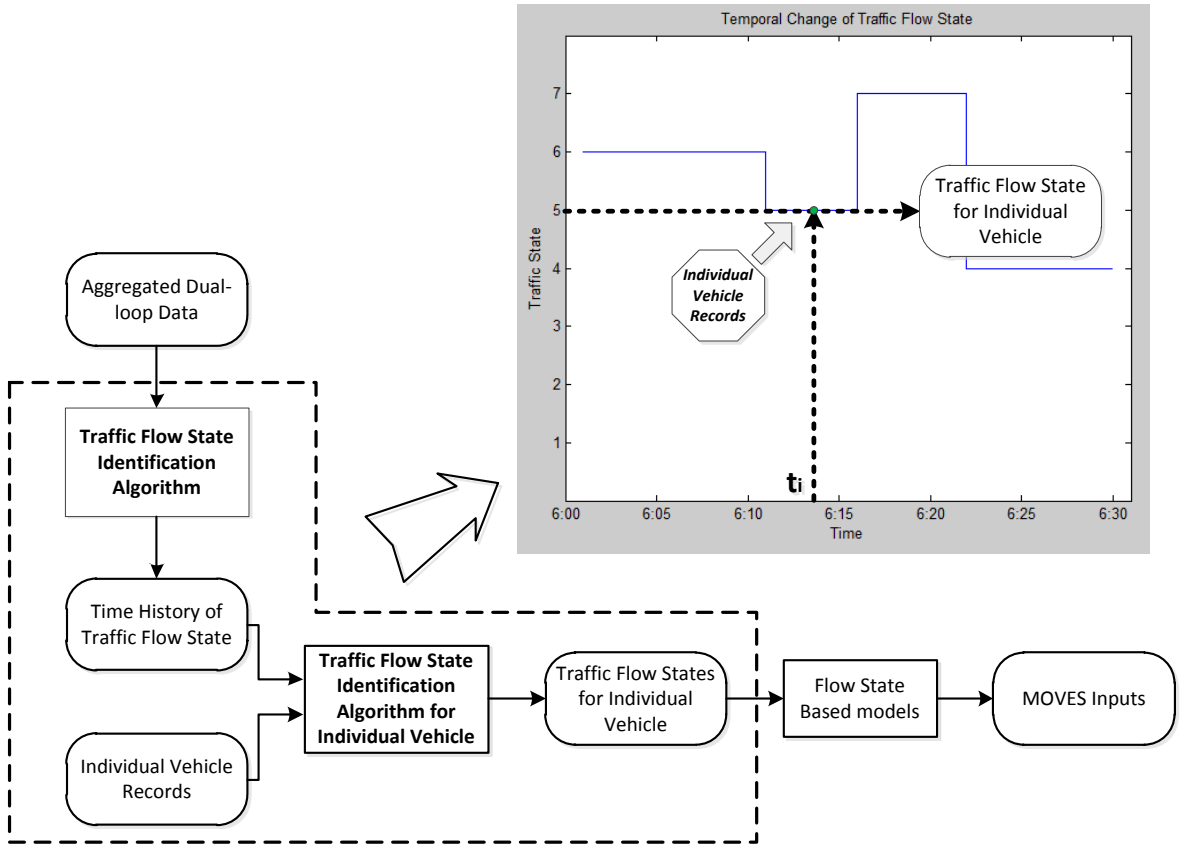

Fig. 10. Traffic flow identification for individual vehicle

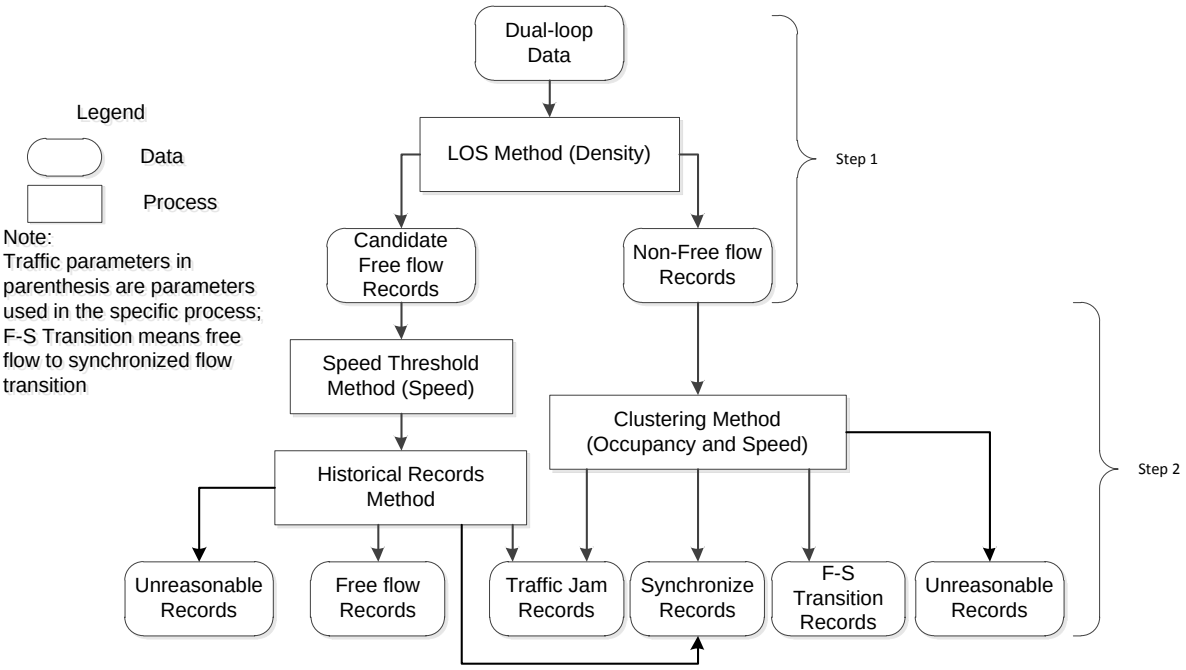

Fig. 11. Schematic of traffic flow phases identification algorithm

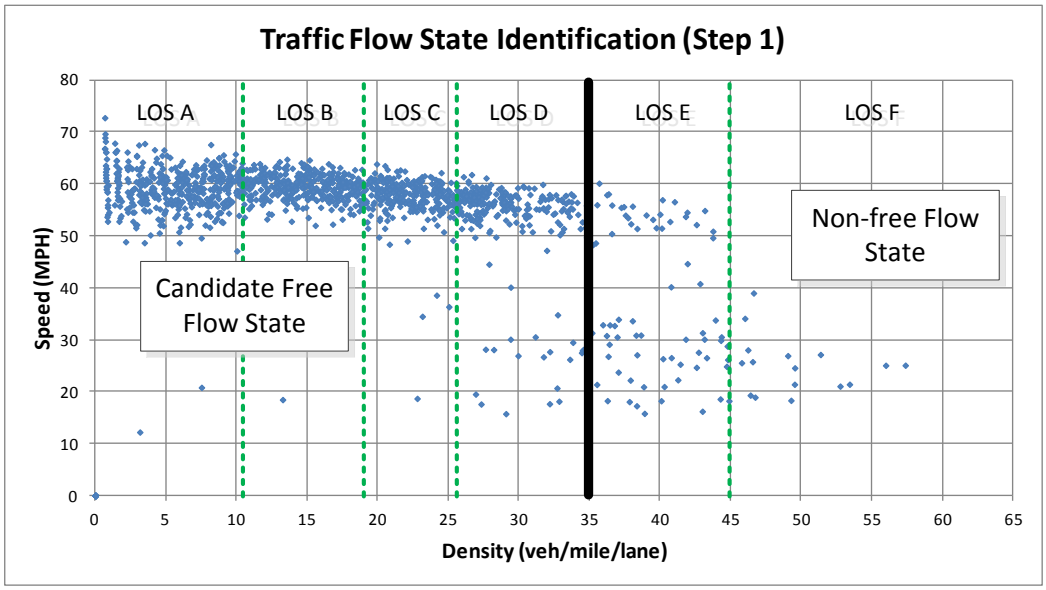

Fig. 12. Step 1 of traffic flow phase identification algorithm

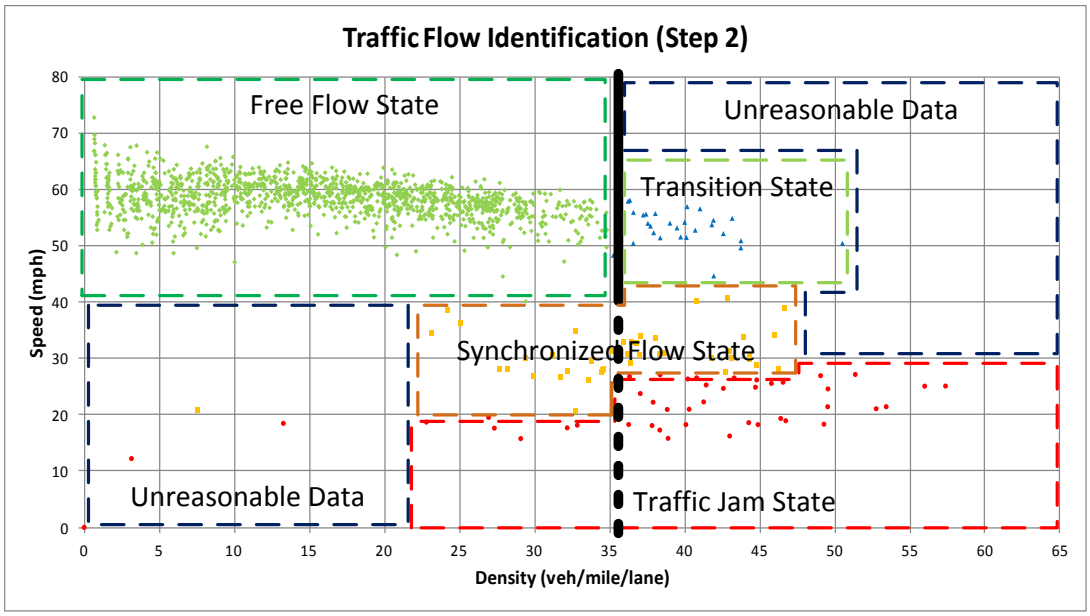

Fig. 13. Step 2 of traffic flow phase identification algorithm

LOS method is used for preliminarily distinguish free flow and non-free flow records. Records with LOS D or better are classified as candidate free flow records. Some non-free flow records may be mistaken as free flow records by LOS method under stop-and-go traffic condition. It occurs when one vehicle stops close to the downstream edge of the S loop, and its following vehicle stops close to the upstream edge of the M loop. The dual-loop station is inside the gap of the two consecutive stopping vehicles. If this situation lasts more than 30 seconds, the measured occupancy (and density) is low during this time interval. Because density is the only criterion used by LOS method, it misidentifies the traffic as free flow traffic.

In order to address this problem, another phase representative variable – speed – is used

to extract congestion data from the candidate free flow data. If a record belongs to the candidate free flow dataset while its speed is lower than 20 mph and larger than 0 mph, the traffic phase for the data record is identified as traffic jam. If a record's speed is lower than 40 mph and larger than 20 mph, the traffic phase for the data record is identified as synchronized flow.

Specifically in the candidate free flow dataset, if a record's speed is 0 mph, there are two possibilities. The first happens when traffic is light and no vehicle passes the dual-loop station during the observation interval. The other happens when traffic is so congested that vehicles stop close to both downstream and upstream edges of the dual-loop station. For the two cases when measured speed is extremely small or close to 0 mph, history records method is used to identify the real traffic phase. If the traffic phase for the previous interval is traffic jam (or synchronized flow), the traffic phase for the current record is traffic jam (or synchronized flow) as well. In rare cases, the data records indicate slow speed as well as small density. Those cases could be caused by special events (e.g., emergency stopping) or erroneous dual-loop outputs. Those records are beyond the study scope of this research and are eliminated from the study dataset. Based on above discussion, the traffic flow phase $F(t_i)$ at time interval t_i is determined by equation (10) as the density k is less than 35 vehicles/mile/lane.

$When\ k\ \leq 35$:

$$F(t_i) = \begin{cases} FF, & u \geq 40 \\ SF, & (k > 20\ and\ 20 < u \leq 40) or\ (u = 0\ and\ F(t_{i-1}) = SF) \\ TJ, & (k > 20\ and\ u \leq 20) or\ (u = 0\ and\ F(t_{i-1}) = TJ) \\ O, & k \leq 20\ and\ u \leq 40 \end{cases} \quad (10)$$

Where:

$k = density, vehicle/mile/lane;$
$u = speed, mph;$
$i = time\ interval\ i;$
$FF = Free\ flow\ state;$
$SF = Synchronized\ flow\ state;$
$TJ = Traffic\ jam\ state;$
$O = Unreasonable\ outliner.$

The raw non-free flow data records are further classified into F→S transition, synchronized flow and traffic jam records, respectively. When traffic becomes congested (LOS becomes E or F), LOS method is unable to distinguish the three traffic phases because those phases could occur within a single LOS level. In this case, the clustering method is applied because it has the ability to identify general traffic patterns from datasets where boundaries for different phases may change from case to case (Xia and Chen, 2007a & 2007b).

The clustering method quantifies traffic phases by first choosing proper clustering variables and number of clusters. Clustering variables are variables representing characteristics of traffic conditions. The clustering method determines class boundaries based on features of the clustering variables. Occupancy and speed are chosen as the phase representative variables in this study and thus they are used as clustering variables. Number of clusters is an empirical value indicating how many distinct classes there are in the study dataset. According to the

previous discussion, three traffic phases, namely F→S transition, synchronized flow and traffic jam, could exist during congested traffic. Therefore, the number of clusters is 3.

For the three traffic phases, data points in the same phase should be very similar. To measure the similarity, K-means clustering method is applied. K stands for number of clusters and is equal to 3. When applying K-means clustering method, initial cluster centroids are randomly selected for all clusters in the first step. Each point (occupancy and speed pairs) in the study dataset is assigned as a member of a certain cluster. The membership of a point is determined by finding the shortest distance between the point and candidate cluster centroids. Using Sun and Zhou's algorithm (2005), the membership of a point to a certain cluster can be calculated by equation (11).

$$u_{ij} = \begin{cases} 1 & if\ d_{ij}(x_i, \mu_j) < d_{ij}(x_i, \mu_{j'}) \\ 0 & otherwise \end{cases} \quad (11)$$

Where:

$x_i = ith\ point\ of\ the\ sutdy\ data\ set$;
$\mu_j = the\ centroid\ of\ the\ jth\ cluster$;

$\sum_{j=1}^{K} u_{ij} = 1 \quad \forall i = 1, \dots, n.$

After each point is assigned, the K-means algorithm seeks to minimize the objective function 12.

$$J = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^{m} d_{ij}(x_i, \mu_j) \quad (12)$$

Where:

$m = coefficient \geq 1, to\ adjust\ the\ blending\ of\ different\ clusters$;
$u_{ij} = membership\ of\ x_i\ in\ cluster\ i$;
$x_i = ith\ data\ point$;
$\mu_j = the\ centroid\ of\ the\ jth\ cluster$;

$d(x_i, \mu_j) = dissimilarity\ metric\ between\ data\ and\ cluster\ centroid.$

In this study, Euclidean distance is used to calculate dissimilarity, which is depicted by equation 13.

$$d_{ik}(x_i, x_k) = \left(\sum_{i=1}^{d} |x_i - x_k|^2\right)^{1/2} \quad (13)$$

When optimization is achieved, $\partial J / \partial \mu_j = 0$ and $\partial J / \partial u_{ij} = 0$. This leads to equations 14 and 15.

$$\mu_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \quad \forall j = 1, 2, 3 \quad (14)$$

$$u_{ij} = \left[\sum_{k=1}^{3} \frac{d_{ij}}{d_{ik}}^{1/(m-1)}\right]^{-1} \quad (15)$$

Every variable is defined in the previous equations. Since it is not easy to obtain analytical solutions for equations 14 and 15, an iterative procedure is needed to calculate approximate solutions. Such procedure keeps updating cluster centroid μ_j and membership variable u_{ij} until no further improvement of J could be achieved.

After the implementation of the presented traffic flow identification algorithm, the traffic flow phases for 24 hours' data on westbound left lane on July 14th, 2009 are shown in the Fig. 14. The identification results are compared with the traffic flow change over time.

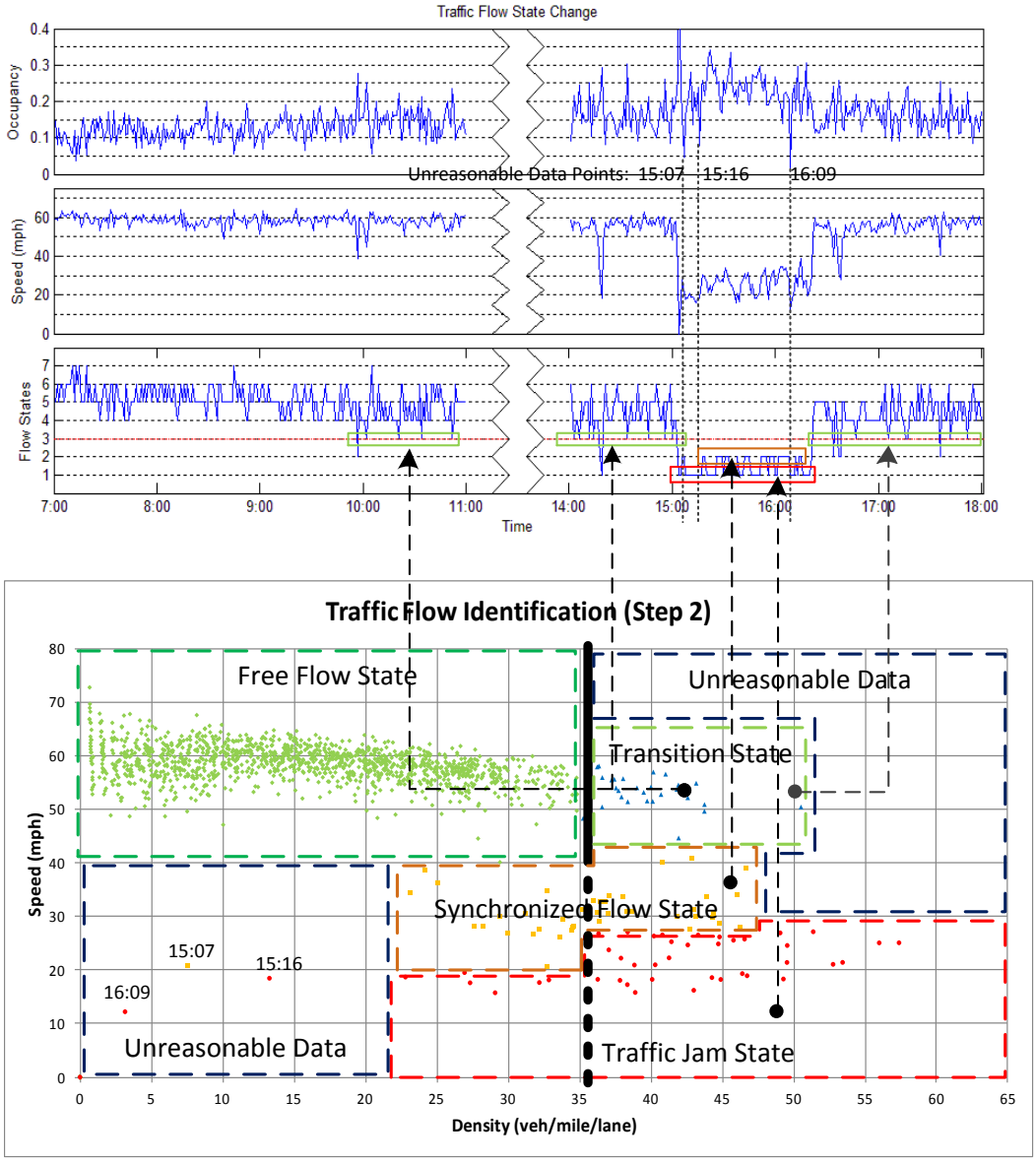

Fig. 14. Temporal change of traffic flow phases

3.5 MOVES Inputs Calculation

As Fig. 2 shows, traffic volume, vehicle composition and operating mode distribution are traffic inputs for MOVES. Traffic volume can be easily extracted from the dual-loop data by

counting the number of data records. The estimation of vehicle composition and operating mode distribution is introduced in this section.

Calculating vehicle composition requires vehicle length. The vehicle length is calculated using equation (2) and (3) under free flow phase. Vehicle length is calculated using equations (4) to (6) under synchronized traffic phase. Equations (7) to (9) are used to estimate vehicle length when it becomes stop-and-go traffic (Wei et al., 2010).

The observed vehicles are classified into two types. Vehicles shorter than 28 feet are considered as passenger cars. Vehicles no less than 28 feet are considered as heavy-duty vehicles. As a result, vehicle composition can be obtained by calculating fraction of passenger cars and heavy-duty vehicles for a specific time interval. Equations (16) and (17) are used to calculate vehicle composition.

$$VC_{ij} = \frac{C_{ij}}{C_i} \tag{16}$$

Where:

$C = vehicle\ count;$
$i = index\ of\ study\ time\ interval;$
$j = 1\ for\ passengercar, 2\ for\ HDV.$

$$C_{ij} = \sum_{n=1}^{N} v_{nj} \tag{17}$$

Where:

$n = index\ of\ observed\ vehicles;$
$v_{n1} = \begin{cases} 1, & length < 28\ feet \\ 0, & length \geq 28\ feet; \end{cases}$
$v_{n2} = 1 - v_{n1}.$

Operating mode distribution is determined by vehicle specific power (VSP) and vehicle speed. VSP is an advanced concept reflecting how vehicle operation impacts emissions as a function of engine load per unit mass of the vehicle. Jiménez-Palacios (Jiménez-Palacios, 1999) was the first to apply this concept. He provides equations to calculate VSP for passenger cars (PC) and heavy-duty vehicles (HDV), as shown in equation (18) and (19).

$$VSP(PC) = v(1.1a + 9.81grade(\%) + 0.132) + 3.02 \times 10^{-4}(v + v_w)^2 v \tag{18}$$

$$VSP(HDV) = v(a + 9.81\sin(grade) + 0.092) + 0.00021v^3 \tag{19}$$

Where:

$v = speed, m/s;$
$a = acceleration, m/s^2;$
$v_w = headwind\ into\ the\ vehicle, m/s.$

Operating mode distribution estimation starts by grouping vehicles into 10-vehicle fleets. Each fleet is one calculation unit. The data for all 10 vehicles in each fleet is aggregated for calculating fleet speed and acceleration. VSP for individual units is calculated according to fleet speed and acceleration using equation (18) or (19). Then operating mode

for each unit is determined using fleet VSP and fleet speed. Finally, operating mode distribution for a period of time (usually 1 hour) is estimated using operating mode data obtained from each unit observed in the time period.

Vehicle fleet is used instead of individual vehicle for operating mode distribution estimation because there is large variance in event dual-loop data. The variance makes the raw event dual-loop data not precise enough for estimating individual vehicle's acceleration. The 10-vehicle data aggregation method is used to reduce the variance and obtain more precise acceleration from the raw data. Those vehicles' records are aggregated to calculated average speed records over both M loop and S loop. Based on the two average speed records, the average acceleration is estimated for each vehicle fleet.

The aggregation method is applied separately for different vehicle classes and different traffic flow phases because simply aggregating all vehicles without classification may generate biased aggregation results by averaging data across different categories. For example, the maneuverability is different among different types of vehicles. Different types of vehicles have different acceleration patterns when they travel through the same road section. More importantly, the acceleration patterns over a dual-loop station change significantly in different traffic phases. Based on the understanding, the aggregation method is applied according to the following observations over the study dual-loop station.

- Acceleration is similar for the same type of vehicles under free-flow condition;
- All vehicles have identical behavior across different lanes in a fleet under synchronized flow condition. These vehicles have similar acceleration behavior;
- There are eight scenarios (Wei et al., 2010) when vehicles pass a dual-loop station under stop-and-go traffic phase. The acceleration is similar for all vehicles under the same scenario.

Vehicle type is considered when classifying vehicles for the purpose of data aggregation though all vehicles in a fleet have similar acceleration behavior under synchronized flow traffic phase. This consideration improves the precision of the results.

Based on the above criteria, the dual-loop data is aggregated based on vehicle types and traffic flow phases. For example, the data records for passenger car under free-flow traffic phase are categorized in one dataset. Every 10 consecutive data records in the dataset are used for fleet speed and acceleration estimation. The 10-vehicle aggregation interval can provide satisfying precision for acceleration estimation as well as keep those vehicles in the same traffic phase. Similarly, data records for other vehicle types and traffic flow phases are categorized and processed. Acceleration for vehicles in the same dataset can be calculated by equation (20).

$$a_i = \frac{\overline{u_{i2}} - \overline{u_{i1}}}{\bar{t}_i} \tag{20}$$

Where:

$i = index\ for\ vehicle\ fleet;$

$\overline{u_{i2}} = \frac{1}{10}\sum_{j=1}^{10}\frac{l_j}{t_{j2}},\ feet/second;$

$\overline{u_{i1}} = \frac{1}{10}\sum_{j=1}^{10}\frac{l_j}{t_{j1}},\ feet/second;$

$\overline{t_i} = \frac{1}{10}\sum_{j=1}^{10}(t_{Sj} - t_{Mj}),\ second;$

$l_j = length\ of\ the\ jth\ vehicle, feet;$
$t_{j1} = upstream\ loop\ on\ time\ for\ jth\ vehicle, second;$
$t_{j2} = downtream\ loop\ on\ time\ for\ jth\ vehicle, second;$
$t_{Sj} = instant\ when\ jth\ vehicle\ arriving\ downstream\ loop, second;$
$t_{Mj} = instant\ when\ jth\ vehicle\ arriving\ upstream\ loop, second;$

To further improve the precision of the estimated acceleration, it is assumed that the preliminary estimated acceleration provides information on the distribution of the real acceleration for the study vehicle fleet. Then the preliminary estimated acceleration data is standardized to reduce results' variance and improve precision. Such standardized data is proven to be a good approximation of the real acceleration situation and is more suitable for VSP calculation. The preliminary vehicle fleet acceleration is standardized using equation (21).

$$a_i' = \frac{a_i - \bar{a}}{\sigma} \quad (21)$$

Where:

$a_i' = standardized\ acceleration\ for\ ith\ vehicle\ fleet, feet/second^2;$
$\bar{a} = average\ acceleration, feet/second^2;$
$\sigma = standard\ deviation\ of\ original\ vehicle\ fleet\ acceleration, feet/second^2.$

With vehicle fleet speed, acceleration and type, VSP for individual vehicle fleet can be calculated using equation (18) and (19). Operating mode can be determined based on Table 4. When operating mode for each vehicle fleet is determined, the operating mode distribution can be determined accordingly using equation (22).

$$OpMode = \begin{bmatrix} op_i \\ \vdots \\ op_n \end{bmatrix} / N \quad (22)$$

Where:

$op_i = number\ of\ vehicle\ fleets\ within\ the\ ith\ operating\ mode;$
$N = total\ number\ of\ vehicle\ fleets.$

Table 4. Operating mode binning definition (EPA, 2010)

VSP	Instantaneous Speed		
	0-25 mph	15-50	>50
<0 KW/ton	Bin 11	Bin 21	N/A
0-3	Bin 12	Bin 22	N/A
3-6	Bin 13	Bin 23	N/A
<6	N/A	N/A	Bin 33
6-9	Bin 14	Bin 24	N/A
6-12	N/A	N/A	Bin 35
9-12	Bin 15	Bin 25	N/A
≥12	Bin 16	N/A	N/A
12-18	N/A	Bin 27	Bin 37
18-24	N/A	Bin 28	Bin 38
24-30	N/A	Bin 29	Bin 39
≥30	N/A	Bin 30	Bin 40

Note: Braking = bin 0; idle = bin 1.

The preceding steps can provide operating mode distribution for specific datasets (e.g., passenger car under free-flow phase, HDV under synchronized phase, etc.). However, combining results from different datasets is necessary because operating mode distribution must be provided for each source type (e.g., passenger cars or heavy-duty vehicles) in a specific period of time for the MOVES analysis. Since time information is included in the results, equation (23) can be used to obtain operating mode distribution for certain vehicle types during specific time period.

$$OpMode = \left(\sum_{j=1}^{m} OpMode_j \cdot n_j\right)/N \quad (23)$$

Where:

$OpMode_j = operating\ mode\ distribution\ for\ jth\ data\ set$;

$n_j = number\ of\ vehicles\ in\ jth\ data\ set$;

$N = \sum_{j=1}^{m} n_j$;

$m = number\ of\ data\ sets$.

In summary, Fig. 15 shows the flow chart of the presented MOVES inputs calculation algorithm.

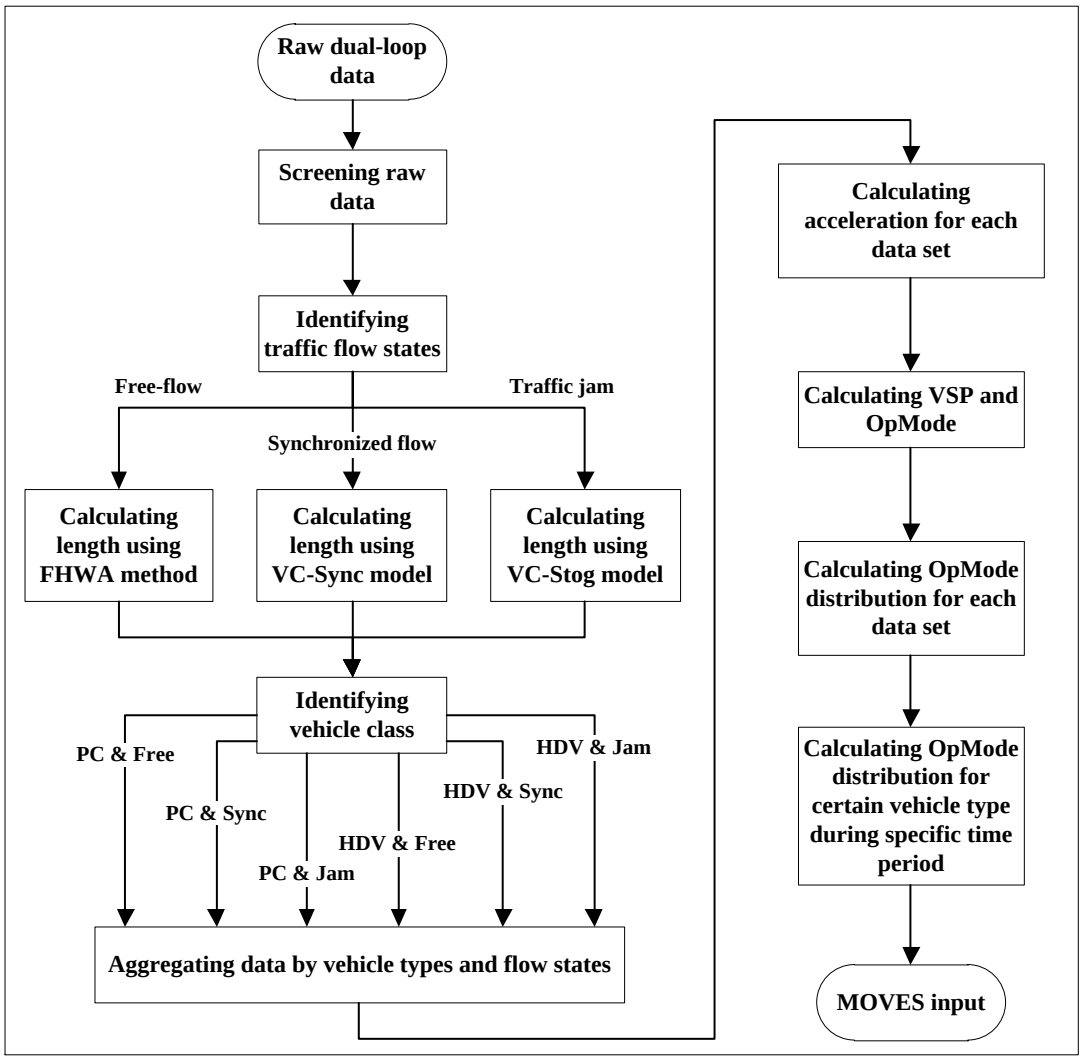

Fig. 15. Algorithm for operating mode distribution calculation

3.6 Model Validation

3.6.1 Validation Data

One hour video data of the traffic flowing over the dual-loop sensors was used as the ground truth to validate the algorithm. Since traffic volume and vehicle composition obtained from dual-loop data are comprehensively studied by previous researches, the validation work emphasizes on verifying the operating mode distribution estimated from dual-loop data.

VEVID software was used to extract the video-based ground truth operating mode distribution data (Wei et al, 2005). Since the location of video camera did not change during the video data collection, it ensures that the coordinate system for every frame of the video data is fixed. The coordinate system is determined by multiple reference points on the pavement of the study road section. The longitudes and latitudes of those reference points are collected by a GPS device embedded in a probe vehicle. Once the coordinate system is determined, the location of vehicles in a frame can be determined by manual tracking. Note the time between two consecutive frames is 1/30 second. If a vehicle is tracked in several frames, its speed and acceleration can be calculated based on the tracked locations and time

lapse. With the acceleration data and speed data provided by the video data processing, the ground truth operating mode distribution is calculated using equations (18), (19), (22) and Table 4. The operating mode distributions from dual-loop data and video data are then compared in order to validate the proposed dual-loop model.

3.6.2 Validation Results

Fig. 16 shows the operating mode distribution obtained from the video data. The video data indicates that the operating modes during the study hour are distributed in two regions. One region has operating mode ranging from bin 21 to 27; the other region has operating mode ranging from bin 33 to 39. This operating mode distribution indicates that three traffic flow phases, namely free-flow traffic, F→S transition and synchronized flow, occurred in the study hour. The outcome from video data analysis is consistent with the field observation.

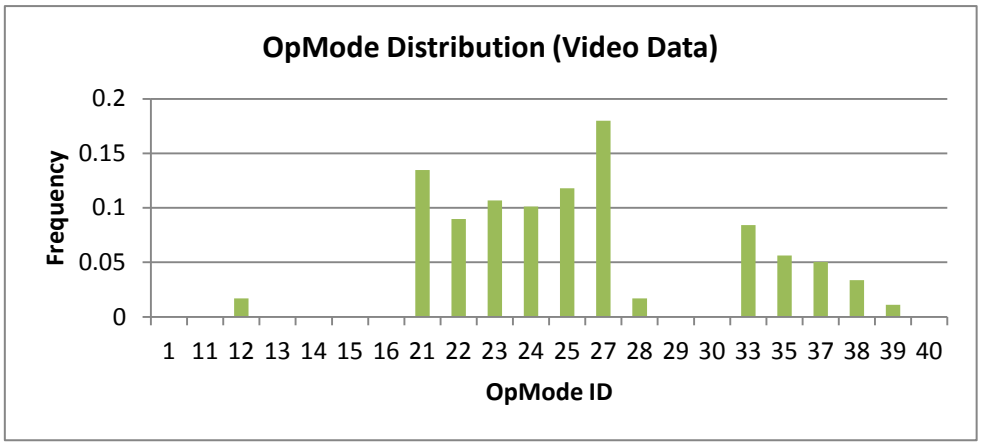

Fig. 16. Operating mode distribution from video data

Fig. 17 illustrates the operating mode distribution obtained from the dual-loop data. It exhibits the same pattern as Fig. 16 does. Table 5 compares the dual-loop results and video results. It can be noted that most differences are less than 5%. The largest two differences are 10% and 12%, which are still acceptable in this study. The differences of individual operating modes between the two datasets do not deny their similarity of the general patterns. The comparison shows that generating traffic inputs for MOVES from dual-loop data is feasible.

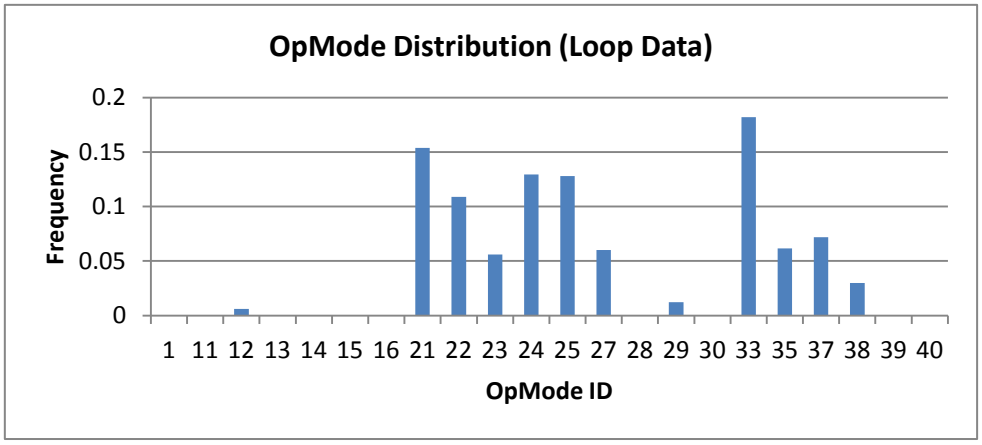

Fig. 17. Operating mode distribution from dual-loop data

Table 5. Operating mode distribution comparison

Operating Mode	1	11	12	13	14	15	16	21	22	23	24
Dual-loop	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.15	0.11	0.06	0.13
VEVID	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.13	0.09	0.11	0.10
Difference	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.02	0.02	-0.05	0.03

Operating Mode	25	27	28	29	30	33	35	37	38	39	40
Dual-loop	0.13	0.06	0.00	0.01	0.00	0.18	0.06	0.07	0.03	0.00	0.00
VEVID	0.12	0.18	0.02	0.00	0.00	0.08	0.06	0.05	0.03	0.01	0.00
Difference	0.01	-0.12	-0.02	0.01	0.00	0.10	0.01	0.02	0.00	-0.01	0.00

3.7 Case Study

The procedure of using dual-loop data for mobile source emission impact analysis has been applied and the results of the analysis are shown in this section. All available dual-loop data is used for the case study. The analysis focuses on the peak traffic hours. Data from 7:00 to 11:00 was chosen for AM peak study. Data from 14:00 to 18:00 was chosen for PM peak study. Traffic flow phases, traffic volume, vehicle composition and operating mode distribution are calculated from dual-loop data at first. The vehicle emissions are calculated using MOVES based on those traffic parameters. Particulate matter 2.5 (PM2.5) and particulate matter 10 (PM10) are selected as the target pollutants. Besides traffic inputs, MOVES also requires several non-traffic inputs. Among them, meteorological data is obtained from National Weather Service website. The average temperature and humidity for July in Columbus area are used. Fuel supply, fuel formation and age distribution data is obtained from local Metropolitan Planning Organization. Although the dual-loop station is a point measurement station, it is assumed that the 1-mile freeway section centered at the study dual-loop station has the homogeneous traffic situation. Therefore, the link length in MOVES is set as 1 mile.

The relationship between operating mode distribution and traffic flow phases is examined first. Fig. 18 shows AM and PM operating mode distribution vs. traffic flow phases for westbound left lane on July 14. The operating mode distribution is distributed across bin 33 to 40 during AM period. If a vehicle cruises on a freeway (acceleration is almost 0) and its operating mode is between bin 33 and bin 40, the vehicle's speed must range from 50 mph to 60 mph. It suggests that the vehicle is under free flow phase. The estimated operating mode distribution is compatible with the traffic flow phase identification result.

During 16:00 to 17:00, the dominant traffic flow phase is synchronized flow. In this case, the operating mode distribution is shifted left toward bin 21 to 25. The vehicle speed for those operating modes is between 25 mph and 50 mph and the vehicle acceleration varies from 0 to 1.5 feet/second^2. It indicates many vehicles operate at low speeds with frequent brake or acceleration during 16:00 to 17:00. The interpolation of operating mode distribution

matches the observation of traffic flow in this hour. During 15:00 to 16:00, the dominant traffic flow phase is traffic jam. The operating mode distribution is shifted further left toward bin 11 to 14. Vehicles operate at even lower speed and experience more frequent acceleration compared with vehicles observed during 16:00 to 17:00. The driving pattern of those operating mode bins is also consistent with observations under traffic jam phase.

By comparing the temporal change of operating mode distribution and traffic flow phases, we find that the lighter the traffic is, the higher the high level operating mode bins' percentages will be.

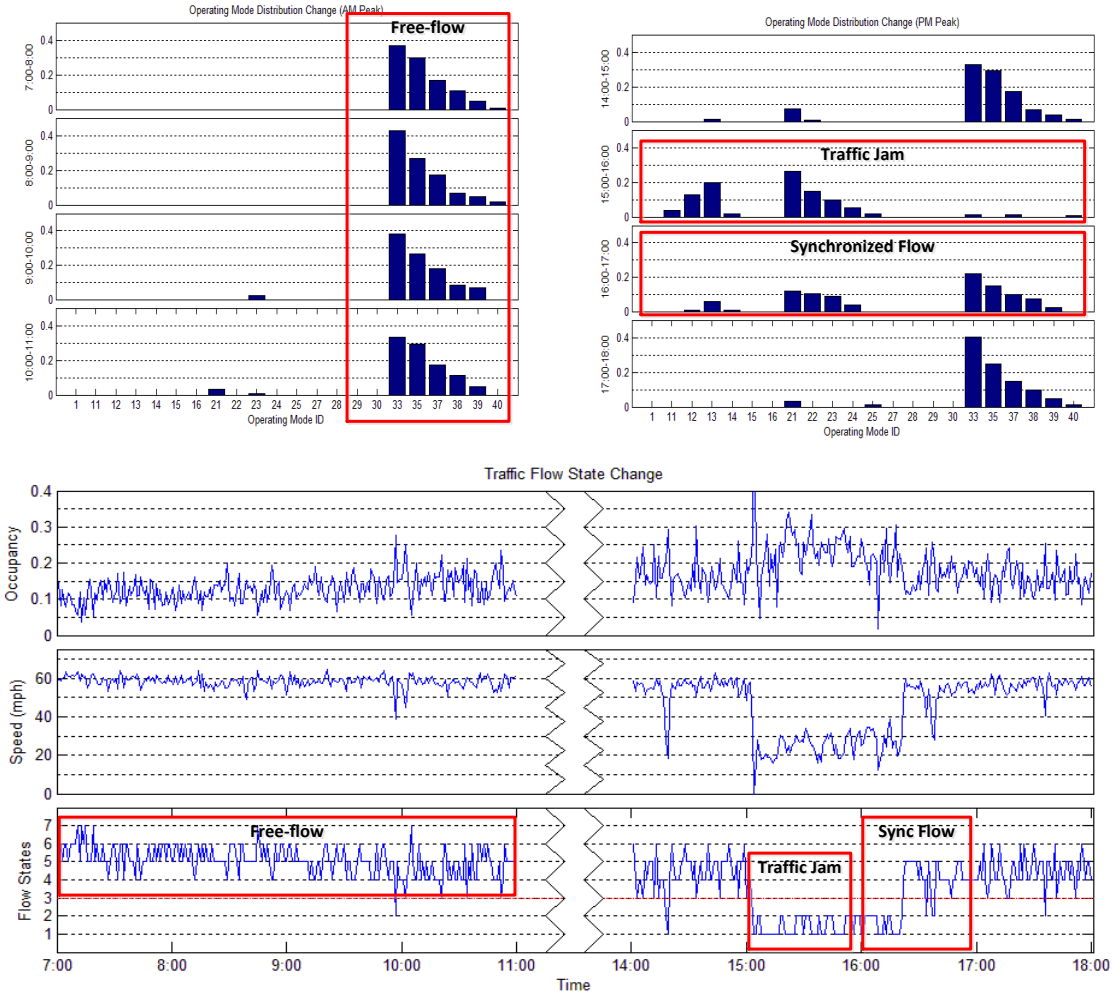

Fig. 18. Comparison of operating mode distribution and traffic flow phases

The emission results are examined as well. It is found that traffic volume, traffic composition and operating mode distribution are contributing factors to mobile source particulate matter (PM2.5 and PM10) emissions. The magnitude of particulate matter is determined primarily by the number of trucks and the traffic volume. Fig. 19 shows the relationship of traffic volume, vehicle composition and particulate matter emission. The traffic volumes for 14:00-15:00 and 16:00-17:00 are almost the same. The truck percentage for 16:00-17:00 is 10% less than that of 14:00-15:00. As a result, the particulate matter emission is 40% smaller during 16:00-17:00 than that of 14:00-15:00. For data collected

during 6:00-7:00 and 7:00-8:00, the traffic composition is the same while traffic volume is smaller for during 6:00-7:00. The emission results show that the particulate matter emission is also smaller during 6:00-7:00.

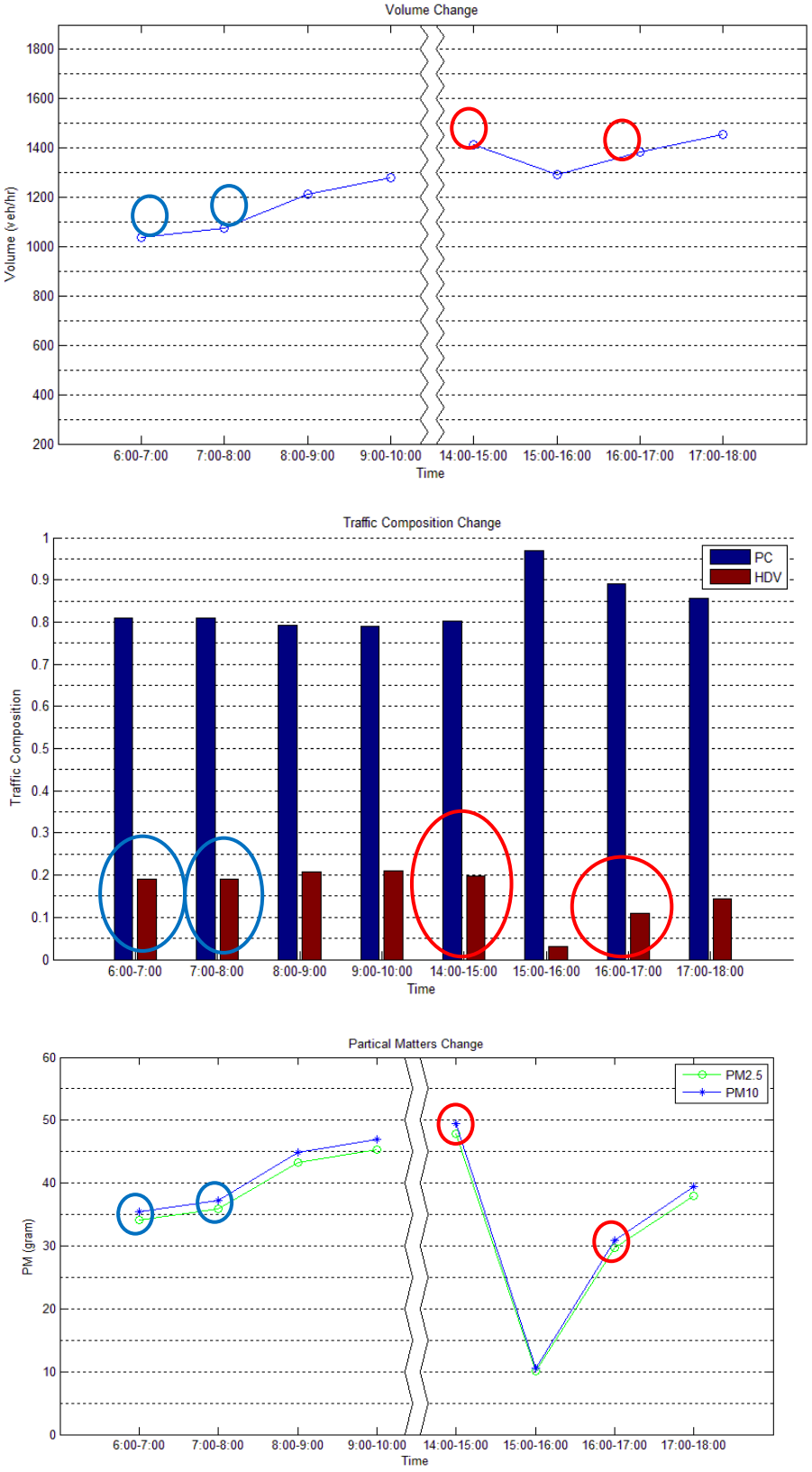

Fig. 19. Relationship of traffic volume, vehicle composition and emission

When traffic volume and vehicle composition are steady, particulate matter emissions change as a function of the operating mode distribution, as Fig. 20 and 21 show. Fig. 20 shows the operating mode distribution vs. particulate matter for westbound left lane on July 14, 2009. Fig. 21 shows the operating mode distribution vs. particle for westbound right lane on July 14, 2009. Data from left lane during 16:00-17:00 and data from right lane during 14:00-15:00 are compared. The volume is approximately 1300 vehicle/hour and the truck percentage is 10% for both datasets. In this case, the particulate matter changes follow operating mode distribution changes. The left lane experienced a long time period of synchronized flow in the study hour. Consequently, the fractions for operating mode bin 21 to 25 are large. The right lane observed the beginning of the synchronized flow during the study hour. The operating mode distribution mainly concentrates on bin 33 to 40. It can be noted that the right lane particulate matter is less than the left lane value, indicating that high level operating mode bins cause relatively low particulate matter emission.

The comparison of particulate matter emissions and traffic volume, vehicle composition and operating mode distribution shows that those parameters dominate mobile source particulate matter emissions given that other environmental factors, such as temperature and wind speed stay stable. In general, the traffic volume and hourly vehicle composition remain similar patterns in expressways and major highways during typical weekdays or weekends. Operating mode distribution then becomes the major traffic factor that influences mobile source emission. If hourly data for operating mode distribution is available for a typical weekday and a weekend for a particular site, it is possible to estimate mobile source emissions at the site for longer time period, such as a mouth. If seasonal environmental factors are considered, the annual emission can be estimated as well.

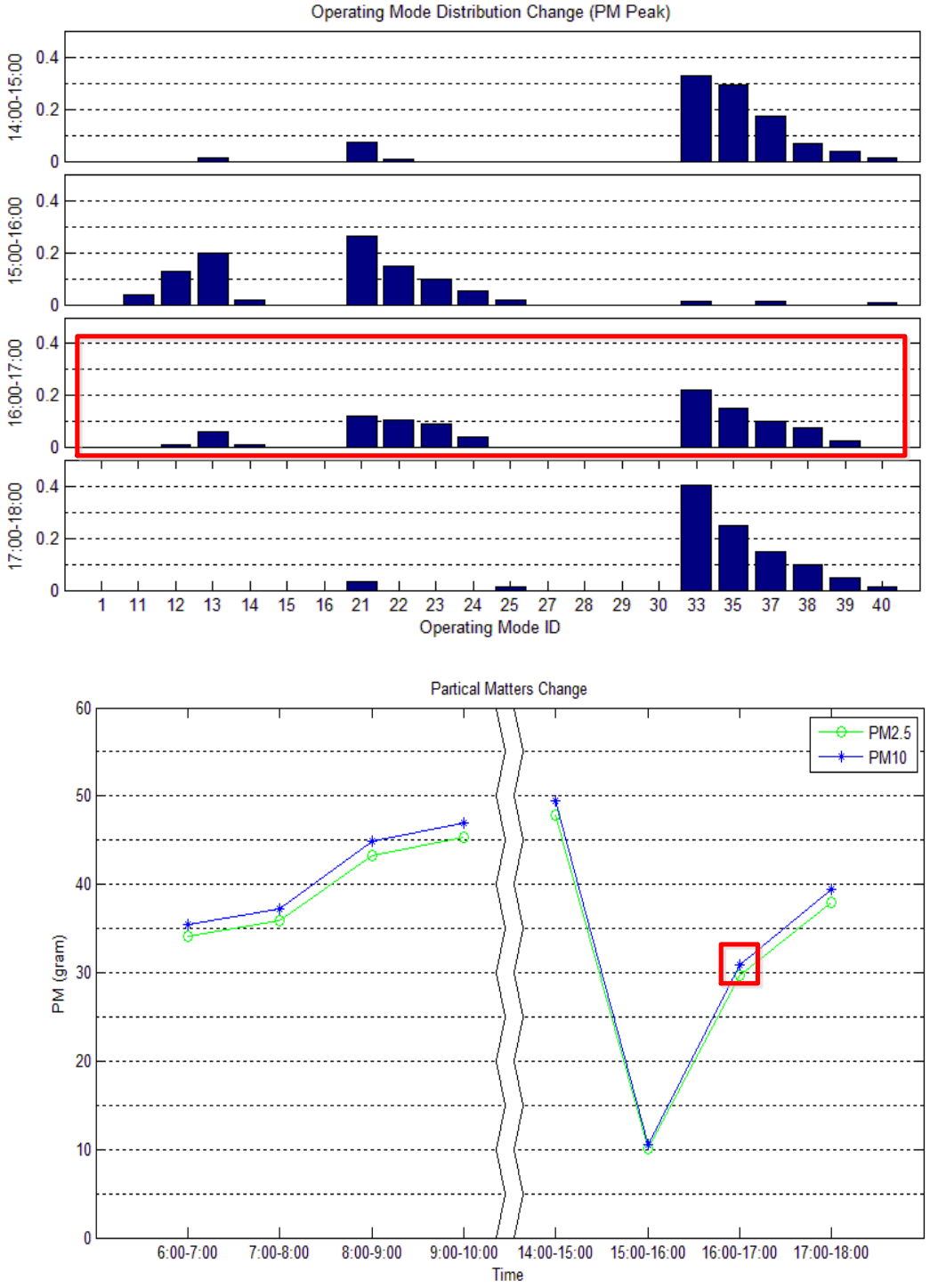

Fig. 20. OpMode distribution and PM for westbound left lane, July 14, 2009

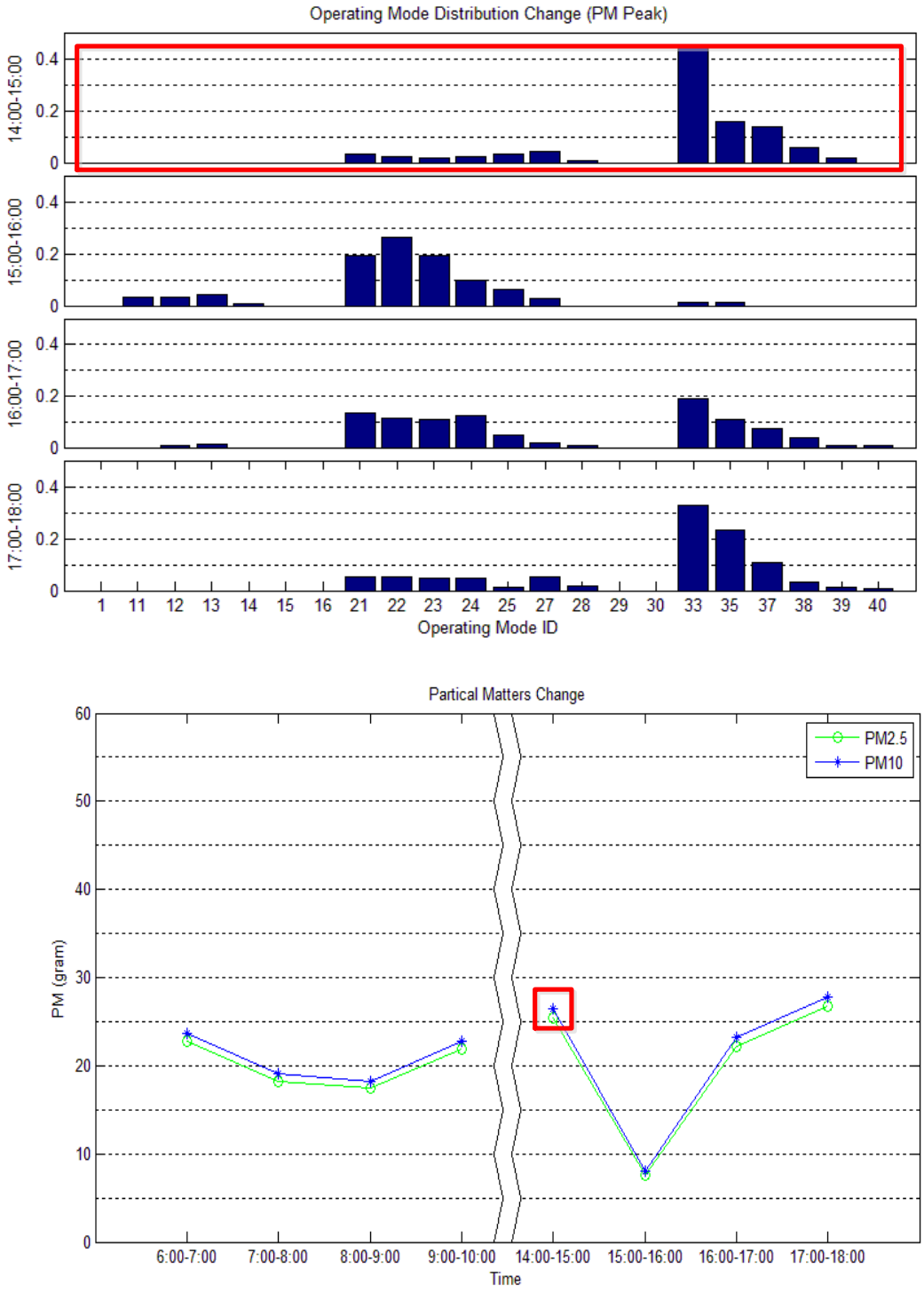

Fig. 21. OpMode distribution and PM for westbound right lane, July 14, 2009

CHAPTER 4: SUMMARY AND CONCLUSIONS

This research provides a methodology for generating traffic inputs for MOVES from the dual-loop data source. The traffic inputs include operating mode distribution, traffic volume and vehicle composition. The procedure for extracting those traffic parameters from the dual-loop data source is demonstrated in this project by using three days of dual-loop data. The presented methodology is validated by video ground-truth data.

In the project, a dual-loop data screening algorithm, a traffic phase identification algorithm and an MOVES inputs calculation algorithm are presented. The dual-loop data screening algorithm is used for filtering out erroneous dual-loop data before it is used to estimate traffic parameters such as vehicle classes, speed, acceleration, VSP and operating mode. There are six filters in the dual-loop data screening algorithm, namely, 0-1 filter, minimum headway filter, pairing filter, sensitivity filter, on-time filter and median speed filter. Most common errors reported by previous studies can be eliminated through these filters. The traffic phases identification algorithm determine free flow, F→S transition, synchronized flow and traffic jam traffic phases by a hybrid method. The method uses level-of-service concept to determine traffic flow phases when it is not congested. When traffic is congested, the K-means clustering method is used to determine synchronized flow and traffic jam phases. The flow phase identification algorithm can help calculate vehicle length since vehicle length calculation algorithm uses different equations under different traffic flow phases. In addition, the flow identification algorithm makes it possible to use flow phases as independent variables when analyzing emission impact of different flow phases. The MOVES inputs calculation algorithm calculates operating mode distribution, traffic volume and vehicle composition based on data processed by dual-loop screening algorithm and traffic flow identification algorithm. In the algorithm, the vehicle acceleration estimation algorithm is of most importance. The preliminary acceleration data was obtained using classified dual-loop data. An aggregation and standardized method is used to improve the precision of the preliminary acceleration data. Then, VSP was calculated using the VSP-based equations and operating mode distribution was determined by EPA specified criteria.

The impact of traffic flow operation on vehicle emission along a specific roadway section can be associated with three quantified traffic flow variables, i.e., operating mode distribution, traffic volume, and traffic fleet composition. On the other hand, the data analysis shows that traffic flow phase is a function of those traffic parameters. The bridge between traffic operation and vehicle emission impact analysis has been established through the application of the three traffic parameters.

This study makes it feasible for a project level mobile source emission impact study to be performed using microscopic real-world traffic data. In the future, more comprehensive ground truth data are needed to further validate the proposed methodology. It is also desirable to expand the methodology in order to take advantage of other similar traffic data sources such as radar data and video detection data sources.

REFERENCES

Al-Deek, H.M., Venkata C., and Ravi, Chandra, S.R. 2004. "New Algorithms for Filtering and Imputation of Real-Time and Archived Dual-Loop Detector Data in I-4 Data Warehouse." Transportation Research Record: Journal of the Transportation Research Board, 1867, Transportation Research Board of the National Academies, Washington, D.C., 116–126.

Azimi M., and Zhang, Y. 2010. "Categorizing Freeway Flow Conditions by Using Clustering Methods." Transportation Research Record: Journal of the Transportation Research Board, 2173, Transportation Research Board of the National Academies, Washington, D.C., 105–114.

Banks, J.H. 1999. "Investigation of Some Characteristics of Congested Flow." Transportation Research Record: Journal of the Transportation Research Board, 1678, Transportation Research Board of the National Academies, Washington, D.C., 128-134.

Brunekreef, B., Janssen, N.A., de Hartog, J., Harssema, H., Knape, M., and van Vliet, P. 1997. "Air Pollution from Truck Traffic and Lung Function in Children Living near Motorways," Epidemiology 8, pp: 298–303.

California Environmental Protection Agency. 2006. Climate Action Team Report to Governor Schwarzenegger and the Legislature.

Cheevarunothai, P., Wang, Y., and Nihan, N.L. 2006. "Identification and Correction of Dual-Loop Sensitivity Problems." Transportation Research Record: Journal of the Transportation Research Board, 1945, Transportation Research Board of the National Academies, Washington, D.C., 73–81.

Cheevarunothai, P., Wang, Y., and Nihan, N.L. 2007. "Using Dual-Loop Event Data to Enhance Truck Data Accuracy." Transportation Research Record: Journal of the Transportation Research Board, 1993, Transportation Research Board of the National Academies, Washington, D.C., 131–137.

Coifman, B. 1999. "Using Dual-loop Speed Traps to Identify Detector Errors," Compendium of Papers CD-ROM, 78th Transportation Research Board Annual Meeting, Washington, DC, January 10-13, 1999.

Coifman, B. (2001). "Improved velocity estimation using single loop detectors." Transportation Research part A 35 (2001) 863-880.

Coifman, B., and Dhoorjaty, S. 2004a. "Event Data-Based Traffic Detector Validation Tests." Journal of Transportation Engineering, Vol. 130, No. 3, 313 -321.

Coifman, B. 2004b. Research Reports: An Assessment of Loop Detector and RTMS Performance. Report No.: UCB-ITS-PRR-2004-30. ISSN 1055-1425

Coifman, B., and Kim, S. B. 2008. "Speed Estimation and Length Based Vehicle Classification from Freeway Single Loop Detectors." Compendium of Papers CD-ROM for the 87th Transportation Research Board Annual Meeting, Washington, DC, Jan.

13-17, 2008.

Duhme, H., Weiland, S.K., and Keil, U. 1998. “Epidemiological Analyses of the Relationship between Environmental Pollution and Asthma,” Toxicology Letters 102–103, pp: 307–316.

Energy Information Administration. 2006. Emissions of Greenhouse Gases in the United Phases 2005, Washington, DC.

EPA. (2010b). “Motor Vehicle Emission Simulator (MOVES) User Guide for MOVES2010a.”

Frey, H.C., Rouphail, N.M., Zhai, H., Fraias, T.L., Goncalves, G.A. 2007, “Comparing real-world fuel consumption for diesel- and hydrogen- fueled transit buses and implication for emissions”. Transportation Research Part D: Transport and Environment, Volume 12, Issue 4, June 2007, Pages 281-291

Frey, H. C., Rouphail, N. M., Zhai, H. 2006. “Speed- and facility-specific emission estimates for on-road light-duty vehicles based on real-world speed profiles.” Transportation Research Record No. 1987, pp. 128–137.

Ishak, S. 2003. “Fuzzy-Clustering Approach to Quantify Uncertainties of Freeway Detector Observations.” Transportation Research Record: Journal of the Transportation Research Board, 1856, Transportation Research Board of the National Academies, Washington, D.C., 6-15.

Jemenez-palacios, J. 1999. “Understanding and Quantifying Motor Vehicle Emissions with Vehicle Specific Power and TILDAS Remote Sensing”. PhD Dissertation. Massachusetts Institute of Technology.

Kerner, B.S., and Rehborn, H. 1996a. “Experimental properties of complexity in traffic flow.” Physical Review E, Vol. E 53, 4275-4278.

Kerner, B.S., and Rehborn, H. 1996b. “Experimental features and characteristics of traffic jams.” Physical Review E, Vol. E 53, 1297-1300.

Kerner, B.S., and Rehborn, H. 1997. “Experimental Properties of Phase Transitions in Traffic Flow .” Physical Review I, Vol. I 79, 4030-4033.

Kerner, B.S. 1998. “Experimental Features of Self-Organization in Traffic Flow.” Physical Review Letters, 81, Vol. 81, 3797-3800.

Kerner, B.S. 1999. “Congested Traffic Flow Observations and Theory.” Transportation Research Record: Journal of the Transportation Research Board, 1678, Transportation Research Board of the National Academies, Washington, D.C., 160-167.

Lin, J., Chiu, Y., Vallamsundar, S., and Bai, S. 2011. “Integration of MOVES and Dynamic Traffic Assignment Models for Fine-Grained Transportation and Air Quality Analyses.” 2011 IEEE Forum on Integrated and Sustainable Transportation Systems.

Millard-Ball, A. 2008. “The Municipal Mobility Manager: A New Transportation Funding Stream from Carbon Trading?” 2008 TRB Annual Meeting

Nesamani, K.S., and Subranmanian, K.P. 2006. "Impact of Real-World Driving Characteristics on Vehicular Emissions." JSME International Journal, Series B, Vol. 49, No. 1, 2006.

Nihan, N. L., Wang, Y.H., and Cheevarunothai, P. 2006. "Improving Dual-Loop Truck (and Speed) Data: Quick Detection of Malfunctioning Loops and Calculation of Required Adjustments." University of Washington, Seattle, Washington.

Ryan, P.H., LeMasters, G.K., Levin, L., Burkle, J., Biswas, P., Hu, S., Grinshpun, S., and Reponen, T. 2008. "A Land-use Regression Models for Estimating Microenvironmental Diesel Exposure Given Multiple Addresses from Birth though Childhood," Science of Total Environment 404, pp: 139-147.

Ryan, P.H., LeMasters, G.K., Biswas, P., Levin, L., Hu, S.,Lindsey, M., Bernstein D.I., Lockey, J., Villareal, M., Hershey, G.K.K., and Grinshpun, S. 2007. "A Comparison of Proximity and Land Use Regression Traffic Exposure Models and Wheezing in Infants," Environmental Health Perspectives, Vol. 115 No. 2, pp: 139-147.

Scora, G., Morris, B., and Tran, C. 2011. "Real-Time Roadway Emissions Estimation using Visual Traffic Measurements." 2011 IEEE Forum on Integrated and Sustainable Transportation Systems.

Shi, Q., and Yu, L. 2011. "Evaluation of Mobile Source Greenhouse Gas Emissions for Assessment of Traffic Management Strategies." Report for Southwest Region University Transportation Center, Report 161142-1.

Sperling, D. 2006. "Asilomar Declaration on Climate Policy." Access 29: 11-21.

Sun, L., and Zhuo, J. 2005. "Development of Multiregime Speed–Density Relationships by Cluster Analysis." Transportation Research Record: Journal of the Transportation Research Board, 1934, Transportation Research Board of the National Academies, Washington, D.C., 64–71.

TRB. (2000). "Special Report 209: Highway Capacity Manual 4th edition." National Research Council, Washington DC.

Vanajakshi, L., and Rilett, L. R. 2004., "Loop Detector Data Diagnostics Based on Conservation-of-Vehicles Principle." Transportation Research Record: Journal of the Transportation Research Board, 1870, Transportation Research Board of the National Academies, Washington, D.C., 162–169.

Wei, H., Meyer, E., Lee, J. and Feng, C.E. (2005). "Video-Capture-Based Approach to Extract Multiple Vehicular Trajectory Data for Traffic Modeling," *ASCE Journal of Transportation Engineering* Volume 131, No. 7 (2005), pp. 496-505.

Wei, H., Ai, Q., Eustace, D., and Yi, P. 2010. "Optimal Loop Placement and Models for Length-based Vehicle Classification and Stop-and-Go Traffic." 2009 Ohio Transportation Consortium (OTC) Research Project Report, University of Cincinnati.

Xia, J., and Chen, M. 2007a. "Defining Traffic Flow Phases Using Intelligent Transportation Systems-Generated Data." Journal of Intelligent Transportation Systems, 11(1), 15–24.

Xia, J., and Chen, M. 2007b. "A Nested Clustering Technique for Freeway Operating Condition Classification." Computer-Aided Civil and Infrastructure Engineering 22, 430–437.

Xie, Y., Chowdhury, M., Bhavsar, P., and Zhou, Y. 2011. "An Integrated Tool for Modeling the Impact of Alternative Fueled Vehicles on Traffic Emissions: A Case Study of Greenville, South Carolina." Compendium of Papers CD-ROM for the 90th Transportation Research Board Annual Meeting, Washington, DC, Jan. 2011.

Younglove, T., Scora, G., and Barth, M. 2005. "Designing on-road vehicle test programs for the development of effective vehicle emission models." Transportation Research Record No. 1941, pp. 51-59.

Zhai, H., Frey, H.C., and Rouphail N. M. 2008. "A Vehicle-Specific Power Approach to Speed- and Facility-Specific Emissions Estimates for Diesel Transit Buses." Environ. Sci. Technol., 2008, 42 (21), pp 7985–7991.