Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

K-MEANS CLUSTERING METHOD TO CLASSIFY FREEWAY TRAFFIC FLOW PATTERNS

OTOYOL TRAFİK AKIM KOŞULLARINI SINIFLAMADA K-ORTALAMALAR KÜMELEME YÖNTEMİ

Mehmet Ali SİLGU[1], Hilmi Berk ÇELİKOĞLU[1]*

[1]Department of Civil Engineering, Faculty of Civil Engineering, Technical University of Istanbul, Istanbul, Turkey.
msilgu@itu.edu.tr, celikoglu@itu.edu.tr

Abstract

In this paper, performances of multivariate clustering methods in specifying flow pattern variations reconstructed by a macroscopic flow model are sought. In order to remove the noise in and the wide scatter of traffic data, raw flow measures are filtered prior to modeling process. Traffic flow is simulated by the cell transmission model adopting a two phase fundamental diagram. Flow dynamics specific to the selected freeway test stretch are used to determine prevailing traffic conditions. The classification of flow states over the fundamental diagram are sought utilizing the methods of partitional cluster analyses by considering the stretch density. The fundamental diagram of speed-density is plotted to specify the current corresponding flow state. Non-hierarchical or partitional clustering analysis returned promising results on state classification which in turn helps to capture sudden changes on test stretch flow states. The procedure followed by multivariate clustering methods is systematically dynamic that enables the partitions over the fundamental diagram match approximately with the flow patterns derived by the static partitioning method. The measure of determination coefficient calculated by using the K-means method is comparatively evaluated to statistically derive this conclusion.
Keywords: Traffic engineering, Traffic flow state.

Öz

Bu çalışmada; makroskobik bir trafik akım modeliyle oluşturulmuş akım koşulu farklılaşmalarını belirlemek amacıyla çok değişkenli kümeleme yöntemlerinin başarımları araştırılmıştır. Trafik verisindeki gürültüyü gidermek ve geniş saçılımı kabul edilebilir düzeye getirmek amacıyla, ham trafik değişkenleri modelleme öncesi filtrelenmiştir. Trafik akımı, iki fazlı bir temel eğriyi baz alarak hesap yapan hücre geçişi modeliyle benzetilmiştir. Seçilen otoyol kesimindeki akım dinamikleri, varolan akım koşullarını belirlemek amacıyla irdelenmiştir. Temel eğri üzerinde akım koşullarının sınıflanması, kesim yoğunluk değişkeni gözetilerek kümeleme yöntemleriyle aranmıştır. Hiyerarşik olmayan kümeleme yaklaşımları, örnek otoyol kesimi üzerindeki ani koşul değişimlerini tespit etmeye yarayan başarılı sınıflama sonuçları vermiştir. Çok değişkenli kümeleme yöntemlerince izlenen prosedür, sistematik olarak dinamiktir ve temel eğri üzerinde statik bölütleme yöntemiyle elde edilen kümelere oldukça yaklaşık kümeler oluşturabilmektedir. K-ortalamalar yöntemiyle elde edilen sonuçlar üzerinden hesaplanmış belirlenim katsayıları, elde edilen sonuçları istatistik yönden karşılaştırmalı olarak değerlendirmek amacıyla kullanılmıştır.

Anahtar kelimeler: Trafik mühendisliği, Trafik akım koşulu.

1 Introduction

In this paper, performances of multivariate clustering methods in specifying flow pattern variations reconstructed by a macroscopic flow model are sought. Pattern, or alternatively state, specification is generally described as estimating the traffic flow variables along a road stretch with an adequate spatial resolution at each time instant based on a limited amount of available measurements from detectors, where the pattern variables are the flows, space-mean speeds and densities [1]. Flow as a single variable is insufficient to exactly specify any pattern since a certain value of flow may correspond to two distinct density and speed values in two completely different flow conditions, i.e., congested and un-congested, which motivated us to obtain the pattern variable, density, specific to condition transitions of flow over a freeway stretch. The pattern variable reconstructed by a macroscopic flow model is further used to classify the fundamental diagram of speed vs. density.

The proposed flow pattern classification procedure is adopted simplifying the approach in [2]. The overall procedure involves two sub-processes succeeding the pre-process of noisy traffic data. The flow modeling incorporates a discrete approximation to a simple continuum model of macroscopic

approach that adopts a two-phase fundamental diagram under stationary and spatially homogeneous equilibrium conditions. Following the process of flow modeling, flow patterns over the fundamental diagram are segmented by considering the dynamics of actual traffic simulation. Since the ultimate aim of the present study is to classify the flow patterns and capture the transitions among these patterns in a dynamic fashion, the time-dependent densities of freeway test stretch are matched on the fundamental diagram that is partitioned using the clustering methods. In order to comparatively evaluate the performance of multivariate classification approach, an example by re-simulating the overall process with transferring the level of service measures [3] is provided.

Traffic flow pattern specification is a fundamental task for freeway traffic surveillance and control and has attracted considerable attention in the past three decades. Although a great number of studies dealing solely with traffic flow performance modeling do exist, many of them ignore a solid motivation for the solution of real-life problems. Therefore it is essential to note that there is a need to differentiate this so-called short-term traffic flow forecasting studies from the ones that traffic pattern is sought explicitly considering a traffic flow model besides an estimator.

Studies on flow state estimation, or equivalently pattern specification, show that both a flow modeling and an estimation component are to be processed in an integrate manner. In these studies, the flow state equation has been sought under different approaches. Some successful trials incorporating various approaches include; constructing linear state equation through flow measure based weighting scheme with interpolation [4]; utilizing linear measurement equations to employ flow measures under Eulerian framework [5]; employing filtering methods integrated to high-order traffic flow models [1],[6] using piecewise linear equations of bi-regime cell based traffic flow model [7]; and establishing linear measure equations to process moving observer data under Lagrangian framework [8].

A modified version of the macroscopic cell transmission model [9] to the estimation of density at unmonitored locations along a freeway segment is introduced in [7]. The model in [7] is simplified in [5] by reducing modes of flow states to two, which are free and congested. A comprehensive extended Kalman filter approach for the estimation and prediction of freeway flow states those are represented by the sectional density is resented in [1] and [6]. Based on a second-order traffic flow model, linearization around the current state is required to determine the transferring flows between sections [1],[6]. A variant of the cell transmission model based on a second-order flow model and adopted an alternative particle-filtering framework to avoid linearization operations is developed in [10].

In a number of relevant studies, modified versions of the wave model originated from the pioneering formulation of Lighthill and Whitham [11] and Richards [12] are used to model the traffic flow. An interpolation method by employing a kernel function to set up the state equation for forward and backward waves and integrated these two equations into a linear state equation through a speed measurement-based weighting scheme is proposed in [4]. In order to reduce the discrepancy between the Lagrangian measurements and the estimated states, a correction term to the Lighthill-Whitham-Richards (LWR) partial differential equation [11],[12] is incorporated in [8].

Considering the works summarized above, a dynamic modeling approach composed of estimation and classification modules is recently introduced in [2] in which neural network methods are used as approximators to map lane-specific sectional density from vehicle count and speed measures obtained simultaneously at multiple sensors, and the simultaneous functioning of the neural approximators in the estimation module with the traffic flow model that simulates actual traffic dynamics is utilized. The approach considering lane-specific measures in [2] is further modified by making use of section-specific measures in [13] in which a multi-mode discrete cell transmission model as the flow model adopts a two-phase fundamental diagram as in [2] but makes use of the multi-modes of flow conditions and transitions within them with the explicit considerations of wave-fronts that exit throughout the decomposed freeway stretch. The approach in both studies [2],[13] seeks the dynamic segmentation of the fundamental diagram in each time interval by considering the dynamics of traffic simulation. Simulated flow patterns are input to be dynamically classified in order to capture the transitions among patterns. The present study introduces the use of multivariate methods in flow state classification process as an alternative to the segmentation method in [2] and [13].

2 Flow Pattern Specification Process

The proposed process for flow pattern classification is composed of three sub-processes succeeding the pre-process of raw traffic data. The density derivation and the flow modeling sub-processes are simultaneous. The third sub-process is the classification of mapped density as pattern variable with the corresponding flow-rate and average speed measures considering a dynamic segmentation on flow patterns over the fundamental diagram of speed-flow. The overall flow pattern specification process is shown in Figure 1.
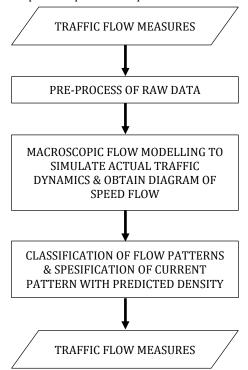

Figure 1: Overall flow pattern classification process.

The macroscopic flow modeling sub-processes run with the input filtered traffic flow measures those are collected by sensors mounted on a freeway stretch. The macroscopic flow model reconstructs section dynamics to obtain the diagram of speed-flow so that it can be partitioned in whether a dynamic or static manner. The traffic flow model is used to simulate and describe the dynamic behavior of flow propagation on the freeway stretch from data set of filtered measures. The resultant measures of density variable with the corresponding flow-rates and average speeds are used in the clustering analyses by multivariate methods. Finally, the current flow pattern is specified by the match-up of current value of the pattern indicator with the appropriate speed-flow diagram segment. It is worthy to note that the number of flow patterns to be classified is decided considering the level of service concept that is defined in the Highway Capacity Manual 2010 [3].

2.1 Information on Study Area and Data Set

In this study the remote traffic microwave sensor (RTMS) data is processed to derive the macroscopic features of traffic flow. To detect vehicles the RTMS uses microwave signals by transmitting frequency modulated continuous wave form in which the transmitted frequency constantly changes with respect to time in a fixed fan-shaped beam [14].

Field data is obtained in January 2009 from three successive microwave sensor units located with the spacing of 250 m and 750.50 m in side-fired position on a 4-lane freeway stretch. The test stretch is on the approach of a Bosphorus strait crossing, the Fatih Sultan Mehmet Bridge, on the European side of Istanbul. Lane-specific measurements on traffic variables, i.e., long vehicle counts, n1(t), non-long vehicle counts, n2(t), and speed, U(t), are obtained for the reference time period 't'. Measurements are aggregated in 2 minutes time interval. For each time interval, flow-rate is determined by the vehicle counts of 20 minutes, and the speed by the harmonic mean of the individual vehicles passing in this time period. Such a procedure is followed since the available measurements on speed are averages of spot speeds and the arithmetic mean of spot speeds leads to systematic overestimations in congested conditions.

Although the technology on data collection systems has been experiencing tremendous progressions, the problem of measurement noise on empirical data still exists, as in our case. It is shown that if the spacing of two adjacent detectors is no longer than about 1 km, it is sufficient to apply a linear smoothing/interpolating filter [4]. It is also noted in [15] that the wide scattering is partly a side effect of the measurement process. Therefore, in order to reconstruct observed dynamics at any point in a space-time plane from sensors positioned at discrete locations, the thin plate spline (TPS) interpolation method [16] is applied to post-process raw data prior to macroscopic model computations.

The TPS filtering helped to tolerate the unrealistic empirical measures and measurement noise which led to fluctuations by affecting the underlying patterns.

Considering the first four order central moments on the distribution of filtered measurements it is seen that variation on actual speeds increase from lane no. 1 to other lanes as lane no. 1 is more frequently used by heavy vehicles those tend to change lane relatively lesser than other vehicle classes. Please note that it is inappropriate to support the same behavior in terms of vehicle counts since the maximum of the vehicle counts is observed when the traffic is flowing at an optimum speed. Therefore, it is the main reason that we consider density as a function of both the flow and the speed, as traffic flow theory dictates.

2.2   Flow Modelling Sub-Process

Macroscopic flow models, those are analogous enough to make the hydrodynamic theory useful in describing traffic dynamics, trace the collective vehicle dynamics in terms of aggregate variables such as density, flow-rate, and speed and are useful in reproducing freeway flow dynamics [17]. One of the frequently argued problems of vehicular flow modeling has been the need to adopt fundamental diagram in flow modeling process. Considering the findings on the wide scattering of empirical freeway data [18],[19],[15], we have chosen to utilize a macroscopic flow model with a fundamental diagram and remove the noise in real data by filtering.

For modeling, we follow the fluid dynamics approach to theory of continuous vehicular traffic flow, defined upon the variables of flow-rate, q, density, ρ, and speed, u, and referred to as the LWR theory [11],[12]. This theory assumes that flow is strictly a function of density, $q = Q(\rho)$, and consequently speed is strictly a function of density, $u = U(\rho)$. The LWR model can be described by a single partial differential equation in conservation form as given by Equation (1) or alternatively as given by Equation (2),

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho \cdot U(\rho))}{\partial x} = 0 \qquad (1)$$

$$\frac{\partial \rho}{\partial t} + \left(C(\rho) \cdot \frac{\partial \rho}{\partial x} \right) = 0 \qquad (2)$$

where; $C(\rho) = (\partial Q(\rho))/\partial \rho$. The LWR theory expresses that slight fluctuations in flow are propagated upstream along kinematic waves, where the speed is given by $c = C(\rho)$, such as the slope of flow-density curve. Given the appropriate boundary conditions, solution to this model can be obtained by determining the function $\rho(x, t)$, where x and t represent space and time respectively. Different variations of the macroscopic model given by Equation (2) can be characterized by the speed-density relationship $u = U(\rho)$ and, consequently, by the adopted fundamental diagram.

In order to obtain a convergent approximation to the continuous LWR model, we utilize the discrete cell transmission approach of Daganzo [9] that adopts a two-phase simplified fundamental diagram as shown in Figure 2.
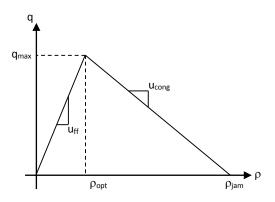

Figure 2: Simplified two phase fundamental diagram of flow-density.

In the graphical representation the fundamental diagram (see Figure 2), u_{ff} is the free-flow speed, u_{cong} is the backward wave propagation speed in congestion, ρ_{jam} is the jamming density, ρ_{opt} is the optimum density and q_{max} is the capacity.

The cell transmission model (CTM) is discrete both in space and time. It divides the freeway into sections called 'cells'. The traffic flow entering a cell bounded by points s and s+1, is considered to be constant between two successive times t and t+Δt and is determined by Equation (3) [9],

$$q^{s, s+1}(t) = \min \left\{ \left(u_{ff}^{s-1,s} \cdot \rho^{s-1,s}(t) \right), \left(u_{cong}^{s,s+1} \cdot \left(\rho_{jam}^{s,s+1} - \rho^{s,s+1}(t) \right) \right), \left(q_{max}^{s,s+1} \right) \right\} \qquad (3)$$

where; $\rho^{s,s+1}(t)$ is the average density of cell s, s+1 between times t and t+Δt, $\rho_{jam}^{s,s+1}$ is the jamming density of cell s, s+1, $u_{ff}^{s-1,s}$ is the free flow speed in cell s-1, s, $u_{cong}^{s,s+1}$ is the congestion wave speed in cell s, s+1, and $q_{max}^{s,s+1}$ is the capacity of cell s, s+1. A schematic representation of our freeway test stretch divided into two cells those correspond to sections is provided in Figure 3. Note that the formulation of flow propagation via cell interfaces is useful for real-time applications, as flow variables

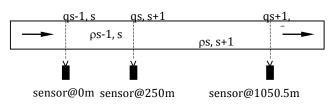are measured by point sensors and sensors may be mounted to match exact points.

Figure 3: Schematic representation of freeway test stretch divided into sections.

Considering the prevailing phase speeds and densities, the flow on each section is determined by Equation (9) for each time interval Δt.

The flow modeling component uses the procedure explained above for the real-time simulation of actual traffic dynamics and the consequent reconstruction of section performances. The real-time reconstruction of flow variables are sequentially used to update and partition the fundamental diagrams of speed-flow and flow-density at each computation time interval, as explained in the following.

2.3 Flow Pattern Classification

The third sub-process involves the partitioning of fundamental diagrams considering pre-defined rules and matching the predicted density to the appropriate partition. It is important to note that in the present study, partitioning a diagram in real-time by updating the critical values of flow in each time step corresponds to the dynamic -may also be referred to as time-varying- classification and specification of traffic patterns. In order to figure out exclusively the performance of pattern specification, an additional example in which neural mappings are matched using the level of service (LOS) measures [3].

2.3.1 Classification Using HC LOS Measures

The LOS measures for multi-lane freeways in the Highway Capacity Manual (HCM) [3] are used to rule partitions. We transfer the LOS measures A through F [3] for flow pattern (FP) representation in terms of vehicle per kilometer-lane, as given by Equation (4).

$$FP = \begin{cases} A, \text{ if } & \rho \le 7 \\ B, \text{ if } & 7 < \rho \le 11 \\ C, \text{ if } & 11 < \rho \le 17 \\ D, \text{ if } & 17 < \rho \le 22 \\ E, \text{ if } & 22 < \rho \le 28 \\ F, \text{ otherwise} \end{cases} \qquad (4)$$

In HCM [3], the boundary between LOS E and F represents capacity. The breakdown of flow, signified by the LOS F, is determined when demand flows are greater than the capacity. In cases where capacity is exceeded, the LOS method in HCM [3] does not produce density estimates and, consequently, specific values within the LOS F.

2.3.2 Classification Using Multivariate Methods of Partitional Cluster Analysis

The boundaries of partitions given by Equation (4) remain constant throughout the simulation horizon, where a time-varying partitioning is performed by the multivariate methods using the procedure in [2].

The multivariate classification approach is analogous to the dynamic classification procedure defined in [2] Critical measures to obtain clusters are dynamically updated throughout the simulation time interval that the whole data set is introduced. This approach differs from the LOS based classification in terms of bounding density measures all of which are obtained considering the mean and deviation statistics of data set, specifically the K-means clustering method which is performed using different distance measures, i.e., Square-Euclidian Distance (SED) and City-Block Distance (CBD). Considering such a possible on-site implementation of the proposed approach supported by information dissemination hardware, the variation on road users' perception should stay in certain limits for appropriate behavior adaptations. This requires the dissemination of a reasonable number of information messages those are signified by captured shifts on patterns. Moreover, in order to obtain a consistent comparative evaluation with the HCM LOS case, we set the user-defined class number to six. In each time step, the segmentation on the fundamental diagram is updated considering the critical distance measures that are calculated using data set of density accumulated 'till the reference computation time interval. Partitioning is applied considering the current density range and calculated multivariate measures. The flow pattern is specified by considering the match of the current density prediction to the appropriate class on the updated diagram. Therefore, the transitions and jumps on patterns between successive times are captured. Illustrations provided in the paper show the overall fundamental diagram partitions that are obtained at the end of the simulation.

In our research since we aimed to comparatively analyze the boundaries of clusters' with the corresponding HCM LOS boundaries the optimum number of clusters are initially assigned on purpose, assuming each LOS as a cluster.

2.3.2.1 K-Means Clustering

The Cluster analysis is used for classifying the objects which have similar characteristics. In cluster analysis, contrary to discriminant analysis, in the beginning there are no sets and we define the border of sets and member of sets. Cluster analysis is used to identify groups or clusters of homogenous individuals. Results of clustering process, there must be high internal homogeneity within cluster and high external heterogeneity between clusters [20].

There are two main clustering algorithms; one is hierarchical and the other is non-hierarchical algorithm. In this study, non-hierarchical algorithm is used due to its simplicity and consequent computation load in comparison to hierarchical algorithm.

Non-hierarchical clustering is applied initially selecting cluster seeds as cluster centers. In comparison to hierarchical methods, it requires the adequate number of clusters at the beginning of the clustering process. Following, non-hierarchical clustering needs the centroids. If there is no information about seed's point, the method takes the initial observation as a cluster seed. To conclude, the non-hierarchical clustering procedure uses the steps given in Figure 4 [21]. Procedure given by Figure 4 [21] can also be used for hierarchical clustering except for partitioning

methods and decide on the distance concept for calculating centroid steps.

The non-hierarchical clustering uses several algorithms. Sequential, parallel and optimization algorithms are well-known algorithms for non-hierarchical clustering algorithms [22]. For sequential threshold, all observations within a specified distance of the seed are clustered and then extra seeds are added. In parallel to threshold, several seeds are selected and assigned objects within the threshold distance to the closest seed. For optimization, it allows observations to be moved to a cluster that has become closer [22].

All the above mentioned clustering algorithms are known as the K-means algorithm. K-means term is suggested by different researchers for describing an algorithm which assigns each item to the cluster having the nearest centroid. In K-means clustering, there is a center point that represents a cluster [23]. Lots of partitional clustering methods do exist. In this paper, K-means algorithm is used for classification.

The literature in which clustering methods are applied to transportation problems in Turkey is quite limited. A recent study investigating the performance of K-means and fuzzy clustering methods on traffic accident data of Denizli city for the years of 2004, 2005 and 2006 is resented in [24]. In this study, the initial numbers of clusters have been selected as 15. However, as a result of minimum iterations, the optimum numbers of clusters have been found 7. Moreover, fuzzy clustering methods have been found more useful than k-means.
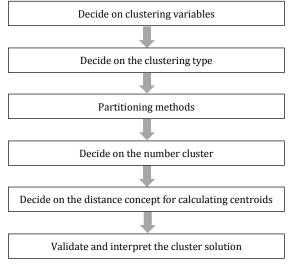

Figure 4: Procedure in Clustering [21].

2.3.2.2 K-means Clustering by Using Square Euclidian Distance

The "distance" concept has to be defined while using K-means clustering. Distance represents a member of data sets range between cluster's centroids. In this study, two types of distance are used for finding the centroids' of clusters. First is the Square Euclidian distance (SED), which needs to be explained by Euclidian distance. It is simply the geometric distance in a multidimensional space and computed as given by Equation (5) [25];

$$\text{distance}(x, y) = \left[\sum_{i \in n} (x_i - y_i)^2 \right]^{1/2} \tag{5}$$

Where x is denoted the coordinate of a point's value on the x axis in Cartesian system, y is denoted of a point's value on the y axis in Cartesian system, and n is the number of observations.

The SED is computed as given by Equation (6) (24);

$$\text{distance}(x, y) = \sum_{i \in n} (x_i - y_i)^2 \tag{6}$$

The distance between clusters is distance between centroids.

2.3.2.3 K-means Clustering by Using the City-Block Distance

Manhattan or City-Block distance is the distance between two points measured along axes at a right angle. In a plane with p_1 at(x_1,y_1) and p_2 at (x_2,y_2), the City-Block distance (CBD) can be obtained as given by Equation (7) [26];

$$|x_1 - x_2| + |y_1 - y_2| \tag{7}$$

Although some of relevant researches are aimed to find similarities, the CBD and the SED are computationally different [27].

3 Problem Formulation and Solution Analysis

The basic version of K-means clustering can be arranged in four steps [23];

1) Initialize a K-partition randomly or based on some prior knowledge. Calculate the cluster prototype matrix;

$M=[m_1,.........,m_K)$

Where M is denoted as a prototype matrix and m is the sample mean for the each cluster.

2) Assign each object in the data set to the nearest cluster C_l, i.e,

$x_j \in C_l, if \ \|x_j - m_l\| < \|x_j - m_i\|$ for j=1, ... , N, $i{\neq}l$, and i=1,.........,K;

Where x_j is a set of objects and C_l is denoted as clusters which are organized to objects.

3) Recalculate the cluster prototype matrix based on the current partition,

$$m_i = \frac{1}{N_1} \cdot \sum_{x_j \in C_i} (x_j)$$

4) Repeat steps 2 and 3 until there is no change for each cluster.

K-means assumes that the number of clusters K is already known by users, which, unfortunately, usually is not true in practice [23]. However, there is no method for specifying the best numbers of K clusters. Therefore, identifying K in advance becomes a very important topic in cluster validity [28].

In this study, there is no necessity to find the optimum number of clusters since a user defined pattern number adopted from the HCM LOS concept [3] exists.

Following the step of cluster number determination, the method to compute both the distances between centroids and the distance of each observation to the cluster centroid has to be selected.

The density values specific to centroids and borders obtained to cluster flow patterns in this study using SED and CBD methods are presented in Table 1.

Table 1: Calculation method of k values.

Value of $k_{centroid}$	Value of k_{border}
K	2*K=P
L	(L-P)+L=R
M	(M-R)+M=T
N	(N-T)+N=W
O	(O-W)+O=Z

The total number of density data input to clustering analyses is 8000. The number of clusters specific to SED and CBD measures based computations are resented in Table 2.

Table 2: Number of clusters specific to SED and CBD measure based calculations.

Flow State	Number of Clusters for SED	Number of Clusters for CBD
1	1620	992
2	2734	1543
3	2180	1737
4	600	1107
5	866	2621

The overall fundamental diagrams of flow vs. density specific to each lane are presented in Figure 5, Figure 6, Figure 7 and Figure 8.

To represent deviations of K-means cluster partitioning from LOS-based partitioning the coefficient of determination is calculated specific to SED and CBD measures and presented in Table 3.

Table 3: R^2 values for each lane by calculated using the SED and CBD methods.

Method and Number of Lane	R^2
SED1	0.9385
SED2	0.8702
SED3	0.9733
SED4	0.9909
CBD1	0.8873
CBD2	0.8842
CBD3	0.9940
CBD4	0.9891

To visually express the comparative evaluation, the partitions obtained using SED, CBD and HCM LOS methods are resented in Figure 9 and Figure 10.
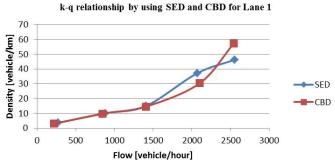

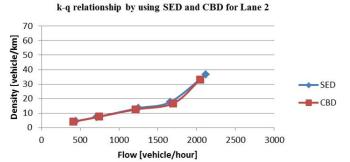Figure 5: Fundamental diagram of density vs. flow for lane1.

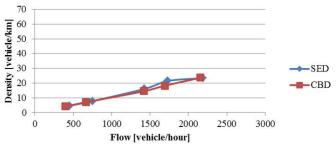Figure 6: Fundamental diagram of density vs. flow for lane2.

Figure 7: Fundamental diagram of density vs. flow for lane3.
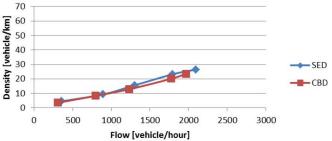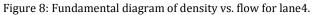

Figure 8: Fundamental diagram of density vs. flow for lane4.
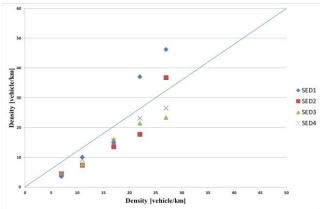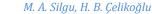

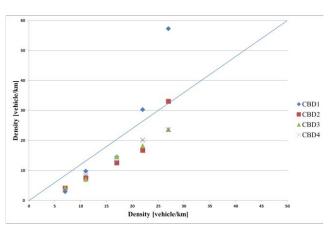Figure 9: Comparison between SED clustering and HCM LOS method.

Figure 10: Comparison between CBD clustering and HCM LOS method.Conclusions

This research showed that the application of K-means clustering method yielded to results consistent with static classification computations adopting the HCM LOS concept. It is seen that although the procedure followed by multivariate clustering methods is systematically dynamic the partitions over the fundamental diagram approximately match with the flow patterns derived by the static partitioning method. The values of determination coefficient calculated by using the K-means method are evaluated to statistically derive this conclusion.

Further directions in our research are to; i- alternate the critical distance measure to obtain cluster centroids, such as the Chebyshev distance, and etc., ii- apply multivariate reduction methods prior to clustering, such as the Principal Component Analysis and Discriminant Analysis Principal Component Analysis, and iii- to evaluate of performance of multivariate clustering methods in comparison to the dynamic classification methods in [2].

4 References

[1] Wang Y, Papageorgiou M. "Real-Time Freeway Traffic Pattern Estimation Based on Extended Kalman Filter: A General Approach". *Transportation Research Part B: Methodological*, 39(2), 141-167, 2005.

[2] Celikoglu HB. "An Approach to Dynamic Classification of Traffic Flow Patterns". *Computer-Aided Civil and Infrastructure Engineering*, 28(4), 273-288, 2013.

[3] Highway Capacity Manual 2010, Transportation Research Board of the National Academies, 2010.

[4] Treiber M, Helbing D. "Reconstructing The Spatio-Temporal Traffic Dynamics from Stationary Detector Data". *Cooper@tive Transport@tion Dyn@mics*, 1(3), 3.1-3.24, 2002.

[5] Sun X, Muñoz L, Horowitz R. "Highway Traffic Pattern Estimation Using Improved Mixture Kalman Filters for Effective Ramp Metering Control". *Proceedings of the 42nd IEEE Conference on Decision and Control*, Maui, Hawaii, USA, 9-12 December 2003.

[6] Wang Y, Papageorgiou M, Messmer A. "Real-Time Freeway Traffic Pattern Estimation Based on Extended Kalman Filter". *Transportation Science*, 41(2), 167-181, 2007.

[7] Muñoz L, Sun X, Horowitz R, Alvarez L. "Traffic Density Estimation with the Cell Transmission Model". *Proceedings of the 2003 American Control Conference*, Denver, Colorado, USA, 4-6 June 2003.

[8] Herrera JC, Bayen AM. "Incorporation of Lagrangian Measurements in Freeway Traffic Pattern Estimation". *Transportation Research Part B: Methodological*, 44(4), 460-481, 2010.

[9] Daganzo CF. "The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory". *Transportation Research Part B: Methodological*, 28(4), 269-287, 1994.

[10] Mihaylova L, Boel R, Hegyi A. "Freeway Traffic Estimation within Particle Filtering Framework". *Automatica*, 43(2), 290-300, 2007.

[11] Lighthill MJ, Whitham GB. "On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads". *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 229(1178), 317-345, 1955.

[12] Richards PI. "Shock Waves on the Highway". *Operations Research*, 4(1), 42-51, 1956.

[13] Celikoglu HB. "Dynamic Classification of Traffic Flow Patterns Simulated by A Multi-Mode Discrete Cell Transmission Model". *IEEE Transactions on Intelligent Transportation Systems*, 15(6), 2539-2550, 2014.

[14] Klein LA, Mills MK, Gibson DRP. "Traffic Detector Handbook, 3rd edition, Vol. I". US Department of Transportation (DOT), Federal Highway Administration (FHWA), FHWA-HRT-06-108, 2006.

[15] Treiber M, Kesting A, Helbing D. "Three-Phase Traffic Theory and Two-Phase Models with a Fundamental Diagram In The Light of Empirical Stylized Facts". *Transportation Research Part B: Methodological*, 44(8-9), 983-1000, 2010.

[16] Wahba G. *Spline Models for Observational Data*. Philadelphia, USA, Society for Industrial and Applied Mathematics (SIAM), 1990.

[17] Celikoglu HB. "A Dynamic Network Loading Model for Traffic Dynamics Modeling". *IEEE Transactions on Intelligent Transportation Systems*, 8(4), 575-583, 2007.

[18] Kerner B, Rehborn H. "Experimental Properties of Complexity in Traffic Flow". *Physical Review E*, 53(5), R4275-R4278, 1996.

[19] Kerner BS. *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Heidelberg, Germany, Springer, 2004.

[20] Hair JF, Black WCJ, Babin BJ, Anderson RE. *Multivariate Data Analysis*. 7th ed. Upper Saddle River, NJ, USA, Prentice Hall, 2009.

[21] Mooi E, Sarstedt M. *A Concise Guide to Market Research*. Berlin, Germany, Springer-Verlag, 2011.

[22] Green PE. *Analyzing Multivariate Data*. Oak Brook, Illinois, USA, Dryden Press, 1978.

[23] Xu R, Wunsch DC. *Clustering*. New Jersey, USA, Wiley-IEEE Press, 2008.

[24] Murat YŞ, Şekerler A. "Trafik Kaza Verilerinin Kümelenme Analizi Yöntemi ile Modellenmesi". *İMO Teknik Dergi*, 4759-4777, 2009.

[25] Statsoft Inc. "How to Group ObjectsInto Similar Categories, Cluster Analysis". http://www.statsoft.com/Textbook/Cluster-Analysis#d (21.08.2013).

[26] National Institute of Standarts and Technology. "Manhattan Distance". http://xlinux.nist.gov/dads//HTML/manhattanDistance.html (12.09.2014).

[27] Hasnat A, Halder S, Bhattacharjee D, Nasipuri M, Basu DK. "Comparative Study of Distance Metrics for Finding Skin Color Similarity of Two Color Facial Images". *National Conference on Advancement of Computing in Engineering Research (ACER 13)*, Krishnagar, West Bengal, India, 2013.

[28] Dubes R. *Cluster Analysis Related Issue*. Editors: Chen C, Pau L, Wang P. Handbook of Pattern Recognition and Computer Vision, 3-32, River Edge, NJ, USA, World Science Publishing Company, 1993.