Classification of driving characteristics using smartphone sensor data

C. Antoniou[1], V. Papathanasopoulou, V. Gikas, C. Danezis and H. Perakis
National Technical University of Athens, Greece

EXTENDED ABSTRACT

I. INTRODUCTION

Human factors and driving characteristics have become a key consideration and design factor for all kinds of transportation systems and infrastructure elements. Classification of driving behaviors allows a finer perception of real traffic, as it helps distinguish and interpret the way that drivers react to different traffic states and situations. Until recently, obtaining detailed traffic information on individual vehicles required expensive and hard-to-operate, specialized equipment that had to be installed on the vehicle of study (thus making it clear to the driver that s/he was under observation and thus potentially affecting his driving behavior). During the past several years, a new type of phones has been prevalent, so called smartphones. These devices incorporate several powerful sensors that collect much of the same information, as those specialized devices. Of course, the accuracy and performance of these devices are not necessarily in par with their more elaborate and expensive counterparts. However, preliminary results suggest that for practical driving behavior analysis, commodity smartphone sensors might be a plausible option (a comparison and discussion of limitations is available in Antoniou et al., under review).

GNSS information is very widely used in transportation applications for several decades. However, there are situations, in which GNSS information is not available, such as indoor facilities, including e.g. parking structures, tunnels and even urban canyons. In this research, we explore the distributions of data (accelerations) collected via smartphones and perform driving behavior classifications. A literature review is presented next, followed by an overview of the methodology and the experimental setup. A preliminary analysis of the data is presented next, followed by the discussion of the driving behavior clustering results. A discussion concludes the paper.

II. LITERATURE REVIEW

An increasing number of research studies focuses on detection and identification of driving behaviors patterns under various circumstances. Some of these studies involve in-vehicle data recorders (e.g. Toledo et al., 2008, Prato et al., 2009, Farah et al., 2013, 2014), which provide very detailed data regarding the vehicle and the driver. The authors focus on applications to young drivers, as well as the relation of driving patterns within the family. However, nowadays technology provides ubiquitous and affordable sensors (e.g. accelerometers and gyros in smartphones) that can provide reliable sensor platforms to monitor driving behavior, as shown e.g. by Antoniou et al. (2014). Johnson and Trivedi (2011) developed a system that recognizes a driver's profile using a smartphone. Aggressive drivers could be identified and therefore awareness and driving safety could be increased. Yuan et al. (2011) have also suggested the detection of a user's driving behavior using mobile sensors towards a self-adaptive driving direction service. Paefgen et al. (2012)

[1] Corresponding author: antoniou@central.ntua.gr

tried to capture driver variability under real traffic conditions using mobile sensing platforms while minimizing the impact of external factors.

Several different approaches have been presented to analyze and classify driving behavior especially in terms of traffic safety. Lotan and Toledo (2005) selected data using an In-Vehicle Data Recorder and classified driving maneuvers as safe, unsafe and dangerous. Observed time series data, including vehicle speed or acceleration, could be plotted vs. time and be subject to a classification system (Ma and Andreasson, 2008). Other researchers have attempted to differentiate cautious from aggressive drivers (Miles and Johnson, 2003; Knapper and Cropley, 2008). Jensen et al. (2011) proposed a driver classification system in terms of drivers' cautiousness based on data obtained from in-vehicle electronic devices, including coordinates, vehicle speed, two axis accelerations, engine speed, wheel speeds, brake pressure, etc. Other studies have focused on more specific driving behaviors. Dai et al. (2010) used accelerometer data from a mobile phone to detect drunk drivers' patterns. Another example is a Driver Monitor System appropriate for recognizing driving patterns of the elderly (Baldwin et al., 2004).

In this research a classification according to x-, y-, z-acceleration measurements is attempted so as different driving patterns to be identified across different drivers or across varying traffic conditions (even of one driver).

III. METHODOLOGY

The objective of the proposed methodology is to classify driving behaviors using field measurements of acceleration, obtained from ubiquitous, opportunistic sensors. Clusters of observations with similar characteristics are identified. Different clusters may represent different driving behaviors and therefore either different drivers' behavior in certain traffic conditions or one driver's behavior under different traffic regimes. As data within a cluster are similar, a traffic model could be calibrated once for each cluster and then the calibrated parameters could be retrieved when the model need to be applied (applying e.g. the approaches developed by Antoniou et al., 2013). Thus, the performance of the models could be improved, as dedicated models would be applied to different conditions.

In this research available field observations are split into *k* clusters, according to the relationships which are identified among measured x, y and z-accelerations. As the number of clusters increases, the most specific each cluster is. On the one hand, this implies that more elaborate and detailed behavioral patterns might be identified; on the other hand, however, limited data and restricted behavioral patterns might impose a very large strain on classification/matching of the appropriate cluster.

In this research, the well-known k-means algorithm, which is one of the approaches found in the literature for clustering, was used. Other similar algorithms could also be used, and their performance could be compared to determine their relative advantages and disadvantages. As its name suggests, the k-means algorithm (MacQueen, 1967, Hartigan, 1975, Hartigan and Wong, 1979) minimizes the distance between each point and the center of its cluster for *k* given clusters. This is achieved by assigning each point to the nearest mean and re-estimating or moving the mean to the center of its cluster. It is regarded as a maximum likelihood clustering. The objective function to be minimized is:

$$\min_{\{\mu_1,\ldots,\mu_K\}} \sum_{h=1} \sum_{x \in X_h} \|x - \mu_h\|^2$$

where μ_i is the mean of cluster *i*.

A hypothesis h_1=<μ_1,…..,μ_K> with the means of the *k* different normal distributions is requested. A random hypothesis is assumed for the initialization of the procedure. Each instance could be written as <x_i, z_{i1}, $z_{i2,...,}$ z_{ik}> where x_i is the observed variable and z_{ij} is equal to 1 if it was obtained by the *j*th normal distribution or 0 otherwise. A maximum-likelihood hypothesis is sought after iterative re-estimations of the expected values of z_{ij}. Then, a new maximum likelihood hypothesis h_2 is calculated using the expected values in the previous step. Finally, the new hypothesis replaces the earlier one and iterations are going on until the algorithm converges to a value for the hypothesis.

IV. EXPERIMENTAL SETUP

Two small-scale experiments (henceforth called A and B) using a relatively recent smartphone (2 year old iPhone 5) were conducted within this research. They took place at the campus of National Technical University of Athens, Greece. The drivers were different between the two experiments but the traffic environment was the same (even though the trajectories were not exactly the same, i.e. they used the same roads, but not necessarily in the same order).

The driving course selected included driving a total distance of approximately 2.5 Km spanning a time period of 15 min in a mixed indoor/outdoor environment. The travelled path included maneuvers in a small indoor parking facility, where GNSS signal was lost, whereas in outdoor sections a good GNSS signal receipt was guaranteed. Data were collected at a normal city driving speed range with higher accelerations associated with straight-line sections.

V. PRELIMINARY ANALYSIS OF DATA

A preliminary insight of the available data could be obtained by examining some general statistics, including minimum and maximum values, as well as the first two moments. This information is presented for the x-, y-, and z-acceleration for the two data sets in Table. Note that z-axis acceleration includes the earth gravity g (~-9.81m/s^2).

Variable	Data series	min	max	mean	var
xacc (m/s^2)	A	-6.90	4.92	0.47	1.79
	B	-3.91	5.11	-0.29	0.73
yacc (m/s^2)	A	-6.78	6.65	0.14	1.81
	B	-4.02	6.83	0.32	1.19
zacc (m/s^2)	A	-14.81	-5.06	-9.83	0.44
	B	-13.12	-7.41	-9.87	0.14

Table 1: Acceleration profile and experimental data

Figure 1 presents the complete data sets. The top row of subfigures illustrates the density distributions for x-, y-, and z-axis accelerations, while the bottom row shows the empirical cumulative distributions for the same data. As one might expect, the x-axis acceleration pattern seems to be the most differentiated between the two drivers, indicating that their differences (different driver and different vehicle) are more easily demonstrated in the longitudinal axis of movement. Furthermore, the second driver (red dashed line) seems to be more stable to his speed avoiding high accelerations or decelerations. On the other hand, the first driver (black line) may be more adaptable to traffic circumstances or be more abrupt to accelerate/ decelerate. Lateral accelerations (y-axis) seem to be similar for the two drivers. Finally, the z-axis acceleration distributions suggest that the first driver exhibited higher accelerations across the vertical axis, possibly suggesting that he moved more abruptly in road elements with grade (such as ramps).

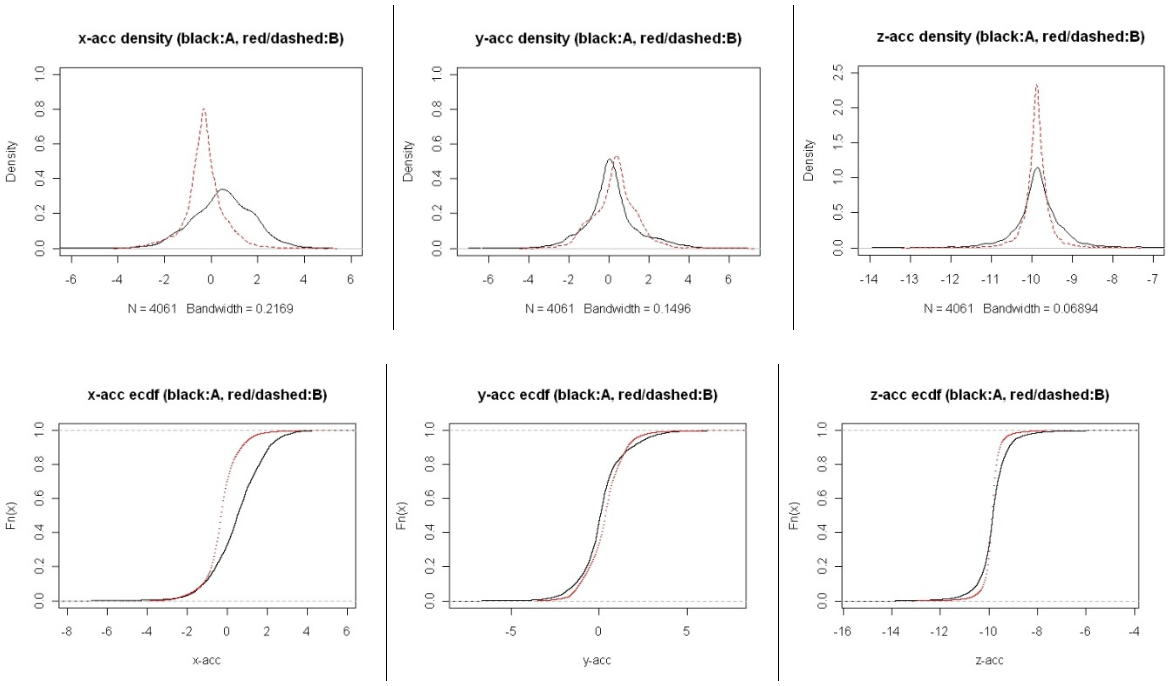

Figure 1: Densities (top) and empirical cumulative distributions (bottom) of x-, y-, and z-axis accelerations for the two data series (black: A, red/dashed: B)

V. DRIVING BEHAVIOR CLUSTERING

For a more in-depth analysis of the available data, a clustering procedure is demonstrated, using the k-means algorithm. Relationships among the data are identified and different driving patterns, according to the acceleration values used for clustering, are recognized. The number of clusters is an exogenous input to this process (unlike e.g. the model-based clustering approach of Fraley and Raftery, 2002, 2003, used in the transportation context e.g. by Papathanasopoulou and Antoniou, 2014). In this research, we consider two different numbers of clusters, 3 and 5, leading to fairly rich, but still easily interpretable models.

First, the results of clustering into 3 clusters are presented in Figure 2 for each data series. It is noted that the two clustering applications are performed independently and therefore there is no correspondence of the colors between the two data series, i.e. the “green” cluster in the results for the first data set does not correspond to the “green” cluster (or any other) in the second data set.

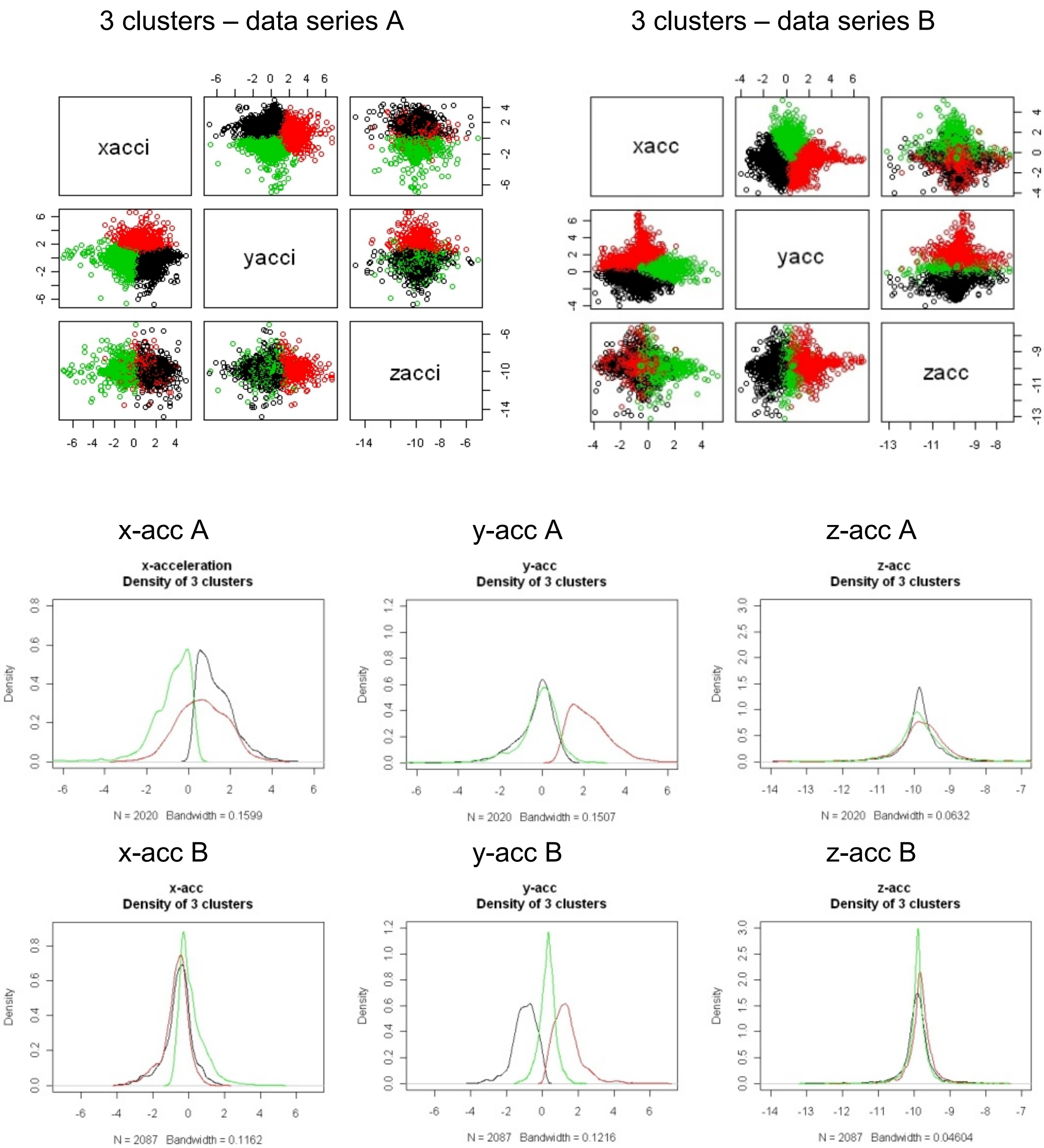

Figure 2: Clustering with 3 clusters for A and B data series

While the main parameter of interest in transportation modeling (especially when related to traffic simulation, where usually currently only the longitudinal movement is simulated) would be the longitudinal acceleration, looking at the acceleration across different axes helps provide further insight. For example, looking at the x-axis view of the three clusters for data series A, one can distinguish one cluster for acceleration (black), one for deceleration (green) and one that spans the entire acceleration range. When one looks at the y-axis accelerations, however, one notices that the red cluster is separated, while the other two clusters mostly coincide. z-axis acceleration does not show radical differences between the three clusters. On the other hand, the three clusters of data series B are differentiated mainly in accelerations of the y-axis, while the variability between the three clusters across the x-axis data is significantly lower. There is no significant differentiation in z-accelerations for both data series.

The richer analysis with the five clusters is considered next (Figure 3). Again, it is noted that the clustering has been performed independently for each data set and, therefore, there is no correspondence between the colors in the two clusters sets. Obviously, the larger number of clusters implies that they are more differentiated between each other and specialized than the previous clustering. The clustering is clearer and the driving behaviors are finer. In addition, clusters between the two data series tend to be similar.

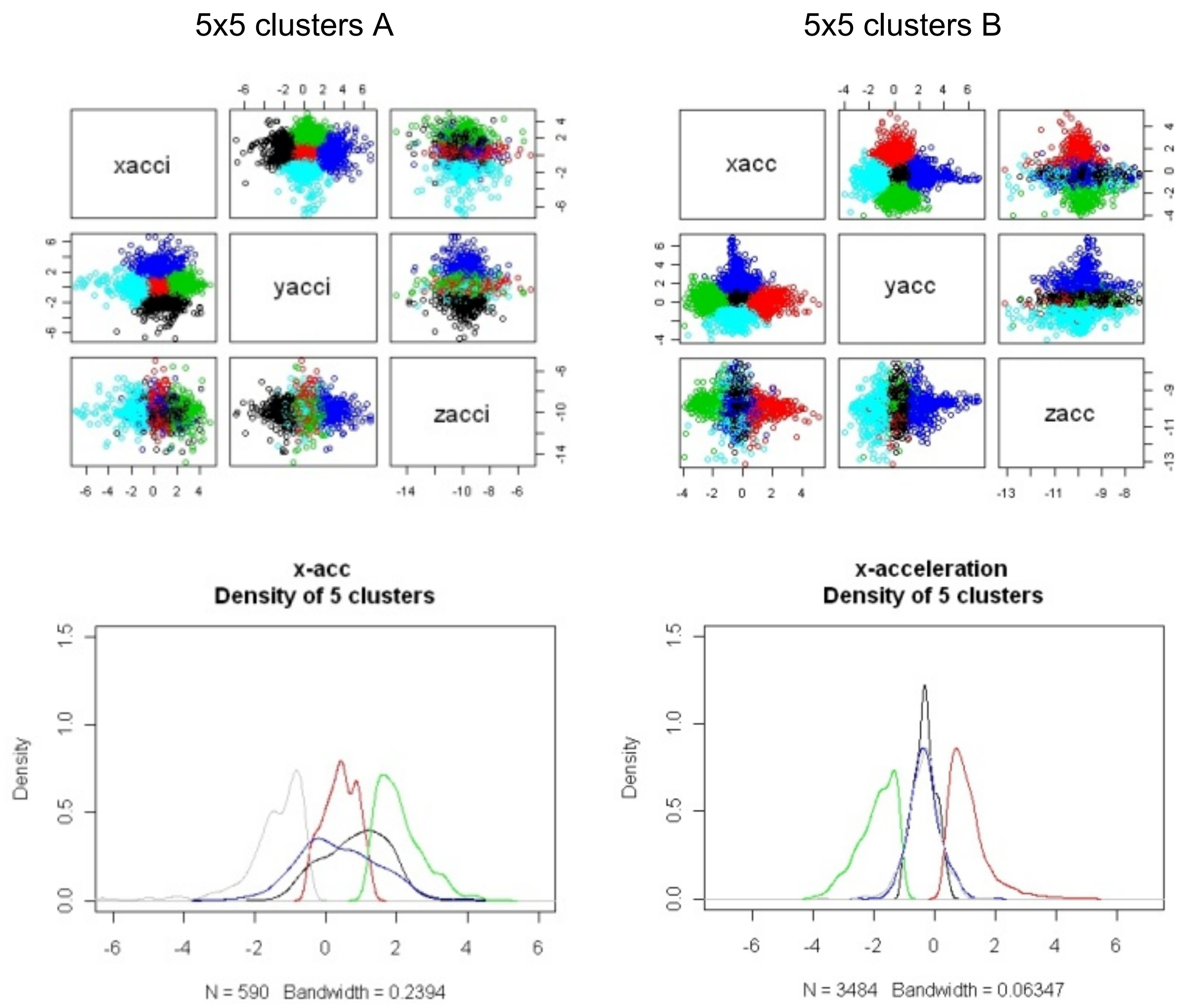

Figure 3: Clustering with 5 clusters (left: series A, right: series B)

V. DISCUSSION AND CONCLUSION

This research has suggested a simple methodology for identifying and classifying different driving behaviors. Having a clear insight of a driving pattern in a normal traffic state it is easier to recognize and study the impacts on driving behavior by certain factors. Several applications are considered for such a procedure. From a behavioral point of view, separating different driving patterns can help the researchers link them with the surrounding roadway environment, traffic conditions, or even operational conditions. For example, one might study the impact of information overload from advanced driver assistance systems within the vehicle, or distraction from mobile phone or other conversations, on driving performance. Moreover, the identification of different traffic behavior patterns could support the calibration of detailed traffic simulation models for each of these regimes, or allow the in-depth analysis of psychological factors of driving behavior.

Ongoing and future research includes the incorporation of more sources of data; modern smartphones incorporate also other sensors, such as gyros, which can be useful in providing information about driving conditions, especially in cases, where GNSS information is not available, such as indoor parking structures, tunnels and urban canyons.

ACKNOWLEDGEMENT

This research has been supported by the Action: ARISTEIA-II (Action's Beneficiary: General Secretariat for Research and Technology), co-financed by the European Union (European Social Fund – ESF) and Greek national funds.

REFERENCES

Antoniou, C., H.N. Koutsopoulos and G. Yannis (2013). Dynamic Data-Driven Local Traffic State Estimation and Prediction. Transportation Research Part C: Emerging Technologies, Vol. 34, pp. 89-107.

Antoniou, A., V. Gikas, V. Papathanasopoulou and T. Mpimis (under review). Towards distribution-based calibration for traffic simulation. 17th International IEEE Conference on Intelligent Transportation Systems, Qingdao, China, October 8-11, 2014.

Baldwin, K. C., Duncan, D. D., and West, S. K. (2004). The driver monitor system: A means of assessing driver performance. Johns Hopkins APL Technical Digest, 25(3), pp. 269-277.

Dai, J., Teng, J., Bai, X., Shen, Z., & Xuan, D. (2010, March). Mobile phone based drunk driving detection. In Proceedings of the 4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 1-8. IEEE.

Farah H., Musicant O., Shimshoni Y., Toledo T., Grimberg E., Omer H. and Lotan T. (2014), Can providing feedback on driving behavior and training on parental vigilant care affect male teen drivers and their parents?, Accident Analysis and Prevention 69, pp. 62-70.

Farah H., Musicant O., Shimshoni Y., Toledo T., Grimberg E., Omer H. and Lotan T. (2013), The first year of driving: Can an in-vehicle data recorder and parental involvement make it safer?, Transportation Research Record 2327, pp. 26-33.

Fraley, C. and A. E. Raftery (2002). Model-Based Clustering. Discriminant Analysis and Density Estimation. Journal of the American Statistical Association, Vol. 97, No. 458, pp. 611-631.

Fraley, C. and A. E. Raftery (2003). Enhanced software for model-based clustering, den- sity estimation, and discriminant analysis: MCLUST. Journal of Classification, 20:263– 286, 2003.

Hartigan, J. A. (1975). Clustering Algorithms. John Wiley and Sons, New York.

Hartigan, J. A., and M. A. Wong (1979). A K-Means Clustering Algorithm. *Applied Statistics,* Vol. 28, pp. 100–108.

Jensen, M., J. Wagner, and K. Alexander (2011). "Analysis of in-vehicle driver behaviour data for improved safety." *International Journal of Vehicle Safety* 5(3), pp. 197-212.

Johnson, D. A., and M. M. Trivedi (2011). "Driving style recognition using a smartphone as a sensor platform." *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 5-7 Oct. 2011, Washington, D.C.

Knapper, C. and Cropley, A. (2008). 'Interpersonal factors in driving.', *Applied Psychology*, Vol. 29, No. 4, January, pp.415–438.

Lotan, T. and Toledo, T. (2005). 'Evaluating the safety implications and benefits of an in-vehicle data recorder to young drivers', *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment*, Training and Vehicle Design, Rockport, ME, June.

Ma, X. and Andreasson, I. (2008). 'Statistical analysis of driver behavior data in different regimes of the car-following stage.', *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2018/2007, pp. 87–96.

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proc., 5th Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1 (L. M. Le Cam and J. Neuman, eds.), University of California Press, Berkeley, pp. 281–297.

Miles, D. and Johnson, G. (2003). "Aggressive driving behaviors: are there psychological and attitudinal predictors?.', *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 6, No. 2, June, pp. 147–161.

Paefgen, J., Kehr, F., Zhai, Y., & Michahelles, F. (2012). Driving behavior analysis with smartphones: insights from a controlled field study. In *Proceedings of the 11th International Conference on mobile and ubiquitous multimedia* (p. 36). ACM.

Papathanasopoulou, V. and C. Antoniou (2014). Simulation optimization of car-following models using flexible models. Opt-i: 1st International Conference on Engineering and Applied Sciences Optimization, Kos, Greece, 4-6 June 2014.

Prato C.G., Toledo T. And Lotan T. (2009). Intra-familial transmission of driving behavior: evidence from in-vehicle data recorders, Transportation Research Record 2138, pp. 54-65.

Toledo T., Musicant O. and Lotan T. (2008), In-vehicle data recorders for monitoring and feedback on drivers' behavior, Transportation Research Part C 16(3), pp. 320-331.

Toledo T., Musicant O. and Lotan T. (2008), In-vehicle data recorders for monitoring and feedback on drivers' behavior, Transportation Research Part C 16(3), pp. 320-331.

Yuan, J., Zheng, Y., Xie, X., & Sun, G. (2011). Driving with knowledge from the physical world. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 316-324). ACM.