# BUS CRASH PATTERNS IN THE UNITED STATES: A CLUSTERING APPROACH BASED ON SELF-ORGANIZING MAPS

*Carlo Giacomo PRATO, Department of Transport, Technical University of Denmark*

*Sigal KAPLAN, Department of Transport, Technical University of Denmark*

## ABSTRACT

Accident taxonomy is widely used by researchers and practitioners worldwide as a tool for understanding accident risks and designing effective policy measures to lessen these risks. Interestingly, despite the usefulness of accident taxonomy for recognising accident risks and the growing interest in improving bus safety operations, information regarding the taxonomy of bus accidents is limited. The current study analyses the risk-factors underlying bus accidents in the United States by uncovering prevailing typologies and analysing their severity. Data from the General Estimates System (GES) crash database were clustered by means of a two-stage clustering method consisting of self-organizing maps (SOM) followed by neural gas, Bayesian classification and unified distance matrix edge analysis. A multi-layer perceptron (MLP) neural network was employed to confirm the correctness and usefulness of the SOM-based clustering process. Five clusters were identified: (i) multi-vehicle collisions at intersections: vehicle encroaching or travelling; (ii) multi-vehicle collisions with school bus at an intersection: distracted drivers; (iii) multi-vehicle collisions in road sections: infrastructure and traffic; (iv) single-vehicle bus accidents off-road: bus travelling and bus driver distraction at low speeds; (v) single-vehicle collisions with non-motorists: pedestrian and cyclists. The analysis pointed out conflicts among buses and other road users and indicated possible cluster-driven directions towards the enhancement of bus safety.

*Keywords: bus crashes, crash clustering, self-organizing maps*

## INTRODUCTION

In recent years there has been a growing interest across world regions in improving bus safety as a result from the number of injuries and fatalities as well as the high media exposure of mass casualty bus crashes (e.g., Blower and Green, 2010; Pearce et al., 2000; Björnstig et al., 2005; Barua and Tay, 2010). In countries throughout the developing world, bus safety is a major concern due to the high share of bus crashes and the main role of buses in providing adequate and affordable accessibility to the vast majority of the population and in promoting

rural and urban development (Pearce et al., 2000). In several countries including Bangladesh, Sri Lanka, India, Nepal, Zimbabwe, and Tanzania, bus crashes represent between 14 and 24% of the reported accidents and are responsible for up to 30-40% of the number of injuries. In Bangladesh, buses constitute about 17% of all vehicles involved in crashes, and 33% of all vehicles involved in fatal crashes (Barua and Tay, 2010). In Sri Lanka, bus crashes comprise 18.5% of the reported accidents, of which roughly 30% result in injuries that require hospitalization (Jayatilleke et al., 2009). In Karachi, Pakistan, buses are responsible for 27% of the injuries and 43% of the fatalities in road traffic crashes, although they constitute only 1.8% of the registered vehicles in the city (Mizra et al., 1999). In Nepal, bus crashes comprise 14% of the reported crashes and are responsible for 39% of the annual road fatalities. In the state of Maharashtra, India, bus crashes comprise 14% of the reported crashes and are responsible for 12% of the fatalities, although buses constitute only 1% of the vehicle fleet. In Zimbabwe, bus crashes comprise 14% of the reported crashes and are accountable for 11% of the injuries. In Tanzania, conventional buses and shared-taxis account for 24% percent of the vehicles involved in crashes and are responsible for 39% of fatalities and injuries (Pearce et al., 2000). In Chile, during the year of 2001 there were over 7,392 bus crashes in the Santiago urban area involving about 5,587 injuries and 112 fatalities (Estache and Gómez-Lobo, 2005).
In Europe and the United States, although buses are considered as a safe mode of transport (Barua and Tay, 2010) due to the relatively small share of bus crashes in comparison with car crashes, the number of injuries and fatalities is far from being negligible. In Europe, about 20,000 coaches are involved in crashes every year for total numbers of 30,000 injuries and over 150 fatalities (ECBOS, 2004) that have been steady over recent years (Björnstig et al., 2005). In the United States, between the years 1999 and 2005 about 63,000 buses were involved in traffic crashes each year, including 14,000 with a nonfatal injury and 325 with a fatal injury (Blower and Green, 2010). Moreover, while bus crashes comprise a relatively small share of the total crashes (0.6%) in the United States, the number of bus crashes per million passenger miles (3.04) is comparable to the number of car crashes per million driven miles (3.2). Decision-makers interest in improving bus safety in Europe is reflected through recent bus safety projects funded by the European Commission, namely the European Coach and Bus Occupant Safety project (ECBOS) and Road Safety in School Transport (RSST), while in the United States the interest is manifested the recognition of bus safety as a high priority issue by the National Transportation safety Board (NTSB, 2011) and the introduction of the new Motor-coach Enhanced Safety Act of 2011 in the House of Representatives and in the Senate. The act aims at establishing new safety standards and regulations for interstate buses, conducting research on bus safety, and creating a new training curriculum for interstate bus operators.
Despite growing concerns with respect to bus safety, studies on bus crashes are scarce, and many basic questions remain unaddressed (Af Wåhlberg, 2004a; Barua and Tay, 2010). Interestingly, while crash taxonomy is widely used by researchers and practitioners worldwide, information regarding the taxonomy of crashes involving buses is scarce (Af Wåhlberg, 2002). Moreover, in the rare cases that it is considered, the taxonomy is defined a-priori by the researcher. For example, Rahman et al (2011) divided bus accidents in Alberta, Canada into four different types based on the number of vehicles and the road type. The four types considered are single-vehicle collisions on highways and non-highway roads and two-vehicle collisions on highways and non-highway roads. The only study focusing on bus

accident taxonomy was conducted by Af Wåhlberg (2002, 2004) for low-speed bus crashes in Sweden. While this pioneer study provides insights regarding certain types of bus crashes and relationships between crash type, bus movement, location and adverse road surface, it is limited in three important ways. Firstly, the analysis is limited to low-speed crashes in urban areas. Secondly, the number of variables that are used for constructing the taxonomy is rather limited, since the main purpose of the taxonomy is its use in studies on bus drivers' accident liability (Af Wåhlberg, 2002). Last, it is unclear whether the developed taxonomy is suitable for other world regions and countries.

The current study contributes to the knowledge about bus crashes by presenting a data driven taxonomy of bus crashes, namely the classification of bus crashes according to their recurrent features and correlation patterns that emerge from the data. The unraveling of crash patterns that emerge from the data is conducted by means of cluster analysis through self-organizing feature maps (SOM) and the usefulness of the formed patterns is confirmed by the implementation of a multi-layer perceptron (MLP) supervised learning algorithm. Consequently, the analysis enables to provide integrative and multi-faceted map containing information regarding the correlation among geographical, demographic, infrastructural and environmental dimensions of bus crashes. Following, the severity distributions of the various clusters are analysed and compared in order to identify priorities in the application of preventive measures. The advantage of employing pattern recognition compared to other methods such as frequency analysis, multiple correspondence analysis, and discrete choice modelling, consists in the ability of unraveling crash patterns directly from the data, without the need of analysing independently crash factors, restricting significantly the number of examined variables, or introducing restrictive a-priori assumptions about relevant crash factors to be considered (Prato et al., 2012).

Bus crash data for the analysis were retrieved from the National Automotive Sampling System General Estimates System (GES) crash database for the years 2005-2009. The GES contains a representative probability sample that is annually drawn from police-reported crashes in 60 geographic areas across the United States. The data contain details about crash location, infrastructure characteristics, environmental conditions, driver attributes and driving behaviour, vehicles and persons involved in each crash. The selected database allows mapping bus crash patterns at a national level rather than at a local level, accounting for bus crashes involving both urban and intercity services, and considering a wide variety of crash risk-factors.

The remainder of the chapter is organized as follows. The next section presents the bus crash data. The third section describes the methodology applied to unravel bus crash patterns. The fourth section presents the results of the cluster analysis and derived taxonomy. Last, the fifth section discusses the major findings of this study and recommends possible preventive measures.

# DATA

Bus crash data from the National Automotive Sampling System General Estimates System (GES) crash database, maintained and published by the National Highway Traffic Safety Administration, served as the data source in the present study.

The GES contains a 1% representative probability sample of road crashes that is annually drawn from roughly 6 million annual police-reported crashes involving severe property damage, injury or loss of life in 60 geographic areas across the United States. The sampling procedure includes stratification by geographic region, primary sampling unit type, police jurisdiction, and accident groups. The data are obtained by GES data collectors that make weekly or monthly visits to approximately 400 police agencies within the sampled geographical areas. The data collectors send copies of the Police Crash Reports (PARs) for selected crashes to a contractor for coding, and then trained personnel interpret and code data directly from the PARs into an electronic file. Data are initially checked for validity and consistency during the coding phase. After the data files are created, they are further checked for quality in order to make reliable and trustworthy electronic data available to governments, researchers, motor vehicle manufacturers, insurance companies, and others (NHTSA, 2010). Since the GES data are obtained from a probability sample, a weight variable is provided in the GES data files in order to produce national estimates of crash characteristics (NHTSA, 2010).

The GES database consists of three main files: accident, vehicle/driver, and person. The accident file reports details concerning each accident, including crash type, date, time of day, pre-crash critical event and manner of collision, number of involved vehicles and road users, level of severity, infrastructure, environmental conditions, and details about specific circumstances and accident typologies. The vehicle and driver file describes each vehicle and driver involved in the accident, with each record listing generic vehicle information, vehicle situational factors, vehicle damage, drivers' demographic characteristics, driver permanent and temporal impairment, and driver's crash avoidance maneuvers. The person file provides information about each person involved in the crash, including demographics, permanent and temporal impairment, and injury severity. The relevant variables for the analysis of bus crashes and their categories are summarized in table 1.

Given the focus of the present study, only crashes involving buses are considered. The buses include transit buses, intercity buses and school buses, although the GES database differentiates only between school buses and other buses. A five-year period is chosen since it and traffic conditions at the national level. Both single-vehicle and multiple-vehicle crashes are considered in the analysis. Overall, the data sample contains information about 2,564 buses representing 325,000 buses involved in crashes between the years 2005-2009.

Table I – Categorical variables for bus accidents

| Variable | Categories |
|---|---|
| Accident characteristics | |
| Maximum injury severity | No injury-possible injury-non incapacitating-incapacitating-fatal |
| Vehicles involved | One-two-three or more |
| Non-motorists involved | Yes-no |
| Manner of collision | None-rear-end-head-on-rear-to-rear-angle-side-wipe same and opposite direction |
| Position | On-roadway-off-roadway |
| Critical event that made the crash imminent | Loss of control-vehicle travelling-another vehicle encroaching-pedestrian-cyclist-animal-object-unknown |
| Infrastructure and traffic control highway | |

| | |
|---|---|
| Interstate highway | Yes-no |
| Number of lanes | Single lane or unknown - multiple lanes |
| Location | Road section-intersection |
| Road alignment | Straight-curve |
| Road profile | Level-grade |
| Road surface | Dry-adverse |
| Traffic control | No traffic control-regulatory sign-traffic signal |
| Speed limit | 15 or less-20-25-30-35-40-45-50-55-60-65 or above |
| Day and time of day | |
| Daylight conditions | Daylight-dawn and dusk-dark with illumination-dark without illumination |
| Time of day | Day-night |
| Day of the week | Weekday-weekend |
| Bus characteristics | |
| Bus service type | School bus-other bus |
| Bus age | Continuous variable |
| Bus movement | Parking-going-straight-stopping-accelerating-decelerating-turning right-turning left-overtaking another vehicle-reversing-negotiating a curve-other |
| Bus driver | |
| Gender | Male-female |
| Age | Continuous variable |
| Distracted driving | Yes-no |
| Speeding | Yes-no |
| Impaired/alcohol/drugs | Yes-no |
| Bus collision partner: vehicles involved | |
| Involved car | Yes-no |
| Involved pick-up trucks | Yes-no |
| Involved SUV | Yes-no |
| Involved vans | Yes-no |
| Involved light trucks | Yes-no |
| Involved heavy trucks | Yes-no |
| Involved motorcycle | Yes-no |
| Involved other | Yes-no |
| Bus collision partner: driver involved | |
| Age | Yes-no |
| Visual obstruction | Yes-no |
| Distracted driving | Yes-no |
| Speeding | Yes-no |
| Impaired driving | Yes-no |
| Alcohol involvement | Yes-no |
| Drugs involvement | Yes-no |

# METHODOLOGY

The current study attempts to individuate patterns of bus crashes and to investigate their recurrent characteristics with respect to the accident, the vehicles and the persons involved.
In order to identify bus crash patterns, cluster analysis is conducted by means of a two-stage approach embedded in the software Synapse (Peltarion, 2012) that consists of SOM followed by neural gas, Bayesian classification and unified distance matrix (U-Matrix) edge analysis. The two-stage approach has several advantages over a single-stage approach, including computational cost reduction and noise diminution (Vesanto and Alhoniemi, 2000). The principle of the methodology is illustrated in Figure 1 and is outlined below.

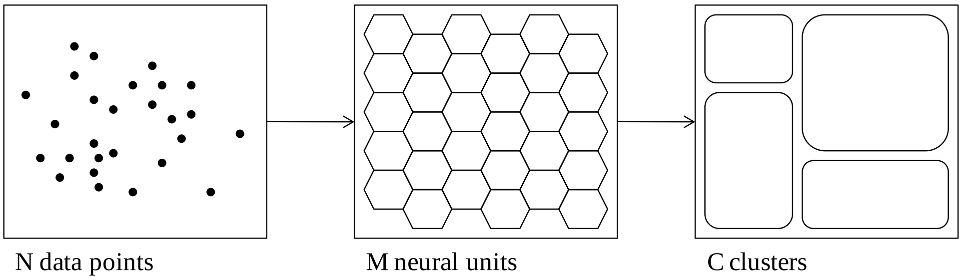


Figure 1 - Two-stage clustering process based on SOM (Source: Vesanto and Alhoniemi, 2000)

At the first stage, the Kohonen's SOM algorithm reduces a high-dimensional manifold of N data points to a two-dimensional array of M neurons (Kohonen, 2001). SOM are a powerful tool for visualizing and examining multi-dimensional data because of their ability to translate complex and nonlinear statistical relationships into simple geometric relationships on a low-dimensional display (Kohonen, 2001). SOM based algorithms have been widely applied in various fields, and their features, their recent variations, and advices for practical application are further detailed by Kohonen (2001).
The SOM algorithm encodes a high-dimensional manifold of $N$ data points onto a two-dimensional array of $M$ neurons (Kohonen, 2001). Each observation $x_i$ ($i=1,2,\ldots,N$) in the data manifold is characterized by a real vector of $K$ attributes $x_i=[x_{i1},x_{i2},\ldots,x_{iK}] \in \Re^K$, and each neuron $j$ ($j=1,2,\ldots.M$) in the array is characterized by a parametric real vector of $K$ scalar weights $m_j=[m_{j1},m_{j2},\ldots,m_{jK}] \in \Re^K$. At the initial stage of the algorithm, the scalar weights are arbitrary and can be drawn either randomly or linearly. Next, at each step of the algorithm a data unit is randomly selected from the input data set and the distance between its attribute values and scalar weights is calculated across $K$ dimensions, with the neuron with the minimal Euclidean distance being chosen as the "best matching unit" neuron for the chosen observation:

$$m_{bmu} \leftarrow \min_j \left\{ \sum_{k=1}^{K} \left( x_{ik} - m_{jk} \right)^2 \right\} \quad (1)$$

Following, a "soft-max" rule is applied to update the scalar weights of the best matching neuron and the topographically closest ones. The update rule for the scalar weight of neuron $j$ at step $t$ is as follows:

$$m_{jk}(t+1) = m_{jk}(t) + m_{bmu}(t)\left[x_{ik}(t) - m_{jk}(t)\right] \quad (2)$$

The function $m_{bmu}(t)$ is a neighbourhood Kernel function:

$$m_{bmu}(t) = \alpha(t)\exp\left(-\frac{\left\|r_{bmu} - r_j\right\|^2}{2\left[\sigma(t)\right]^2}\right) \quad (3)$$

where $\alpha(t)$ is a scalar valued "adaptation gain" ($0 \leq \alpha(t) \leq 1$), $r_{bmu}$ is the vector of coordinates of the "best matching unit" neuron, $r_j$ is the vector of coordinates of neuron $j$, and $\sigma(t)$ is a decreasing function of time.

Upon the SOM formation, cluster analysis is applied to the $M$ neurons in the SOM to form $C$ clusters. The main benefit of clustering the SOM neurons instead of the data units is noise reduction since neurons contain data averages, and hence they are less sensitive to random variations in the original data (Vesanto and Alhoniemi, 2000). The SOM is clustered to $C$ clusters by applying the Neural Gas (NG) algorithm (Martinetz et al., 1993), which has clear advantages in comparison with other clustering methods (i.e., k-means, maximum-entropy and Kohonen's SOM) in terms of convergence speed and accuracy, while its main weakness is the high computational complexity (Martinetz et al., 1993). Consequently, the NG algorithm is applied on the low-dimensional SOM neuron array rather than directly to the data units, in order to gain fast convergence and low distortion error while mitigating computational complexity.

Once the SOM neurons are clustered with the NG algorithm, the $C$ clusters are successively refined by means of Bayesian classification that combines region-growing/merging with edge detection. The general principle involves the comparison of each neuron with its spatial neighbors, and its assignment to the same class based on statistical similarity measures. The number of clusters is determined according to a Bayesian goodness-of-fit criterion (Lee and Crawford, 2005).

Following the clustering of bus crashes on the basis of the characteristics of the accidents, the vehicles and the persons involved, an MLP single-layer classifier validates the correctness of the classification. The MLP is a feed-forward back-propagation neural network that connects an input layer to an output layer through a hidden layer. The term "feed-forward" indicates that the network connects neurons only forward, while the term "back-propagation" indicates that the network performs supervised training by comparing anticipated outputs against the actual outputs of the neural network. Using the cluster belongings of each crash as anticipated outputs and the crash characteristics as the inputs, the "back-propagation" training algorithm calculates the error and adjusts the weights of the single layer backwards from the output layer to the input layer. The algorithm then measures the accuracy of the classification by comparing the actual output from the MLP with the anticipated output of the SOM.

# RESULTS

General characteristics of the accidents describe that, as expected, the majority of the bus accidents occurred during daytime (88.0%) and on weekdays (88.8%). The majority of the bus accidents occurred in multi-lane roads (68.8%) where the speed limit ranged between 40-

70 kilometres per hour. Most of the bus accidents involved two vehicles (78.1%) and about half (52.3%) occurred at intersections. Non-motorists were involved in only a small fraction (4.5%) of the bus accidents. The vast majority of the accidents occurred in favourable road conditions, namely straight road sections (94.3%), at a level grade (85.3%), and dry road surface (79.0%). 28.0% of the accidents involved pick-up trucks, vans and sport utility vehicles and, while about 7.0% of the accidents involved light and heavy trucks. As expected, only a small fraction of the bus accidents involved bus driver charges such as speeding (3.0%), impaired driving or driving under the influence (0.5%). The share of female bus drivers in the data was 39.8%. In terms of age, the share of drivers in the beginning of their career (under 34 years of age) was 19.5%, while the share of elderly drivers (over the age of 64) was 8.1%. The share of young drivers involved as the bus collision partner was 18.7% and the share of elderly drivers was about 7.2%.

In terms of the consequences, 16.6% were single-vehicle accidents, almost a third of the accidents (27.9%) involved rear-end collisions, another third (33.7%) comprised angle collisions, and 16.6% consisted of side-wipes in the same direction. Head-on collisions or side-wipe in the opposite direction comprised only 5.6% of the accidents. 79.7% of the bus accidents involved property damage only, 11.9% involved possible injury, 5.9% resulted in non-incapacitating injury, 2.2% ended in incapacitating injury and 0.2% of the accidents were fatal.

The SOM clustering method was applied to a 20x20 neural network, in order to avoid forced clustering as a result of a small neural network size with respect to the number of observations in the data manifold. The SOM cluster analysis yielded five clusters of bus accidents workers that greatly differ with respect to their main features. Figure 2 compares relevant features across the five clusters. Pearson's chi-square test confirmed the statistical significance of the differences with respect to every feature. The MLP single layer classifier confirmed the usefulness the generated classification. The built classifier based on the SOM predicts well the cluster pertinence (*MSE=4.12E-002*). The usefulness of the classifier is manifested by the confusion matrix presented in Table 2 that shows that the correct prediction rate across clusters is roughly 90%.

Table II – The confusion matrix resulting from the MLP one layer classification

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| Cluster 1 | 92.67% | 2.67% | 0% | 4.67% | 0% |
| Cluster 2 | 4.55% | 90.0% | 0% | 5.45% | 0% |
| Cluster 3 | 0% | 2.18% | 92.14% | 5.68% | 0% |
| Cluster 4 | 3.39% | 0% | 1.69% | 88.14% | 0% |
| Cluster 5 | 0% | 0% | 0% | 0% | 100.00% |

The most important variables for classification seem to be the accident location, the bus service type, number of vehicles and non-motorist involved, the traffic control and speed limit, the bus movement prior to the accidents and the critical event that made the crash imminent, and driver distraction. Visual obstruction, adverse road conditions and driving offences (e.g., speeding or impaired driving) are relatively rare but contribute to the cluster identification. The cluster descriptions are outlined below.

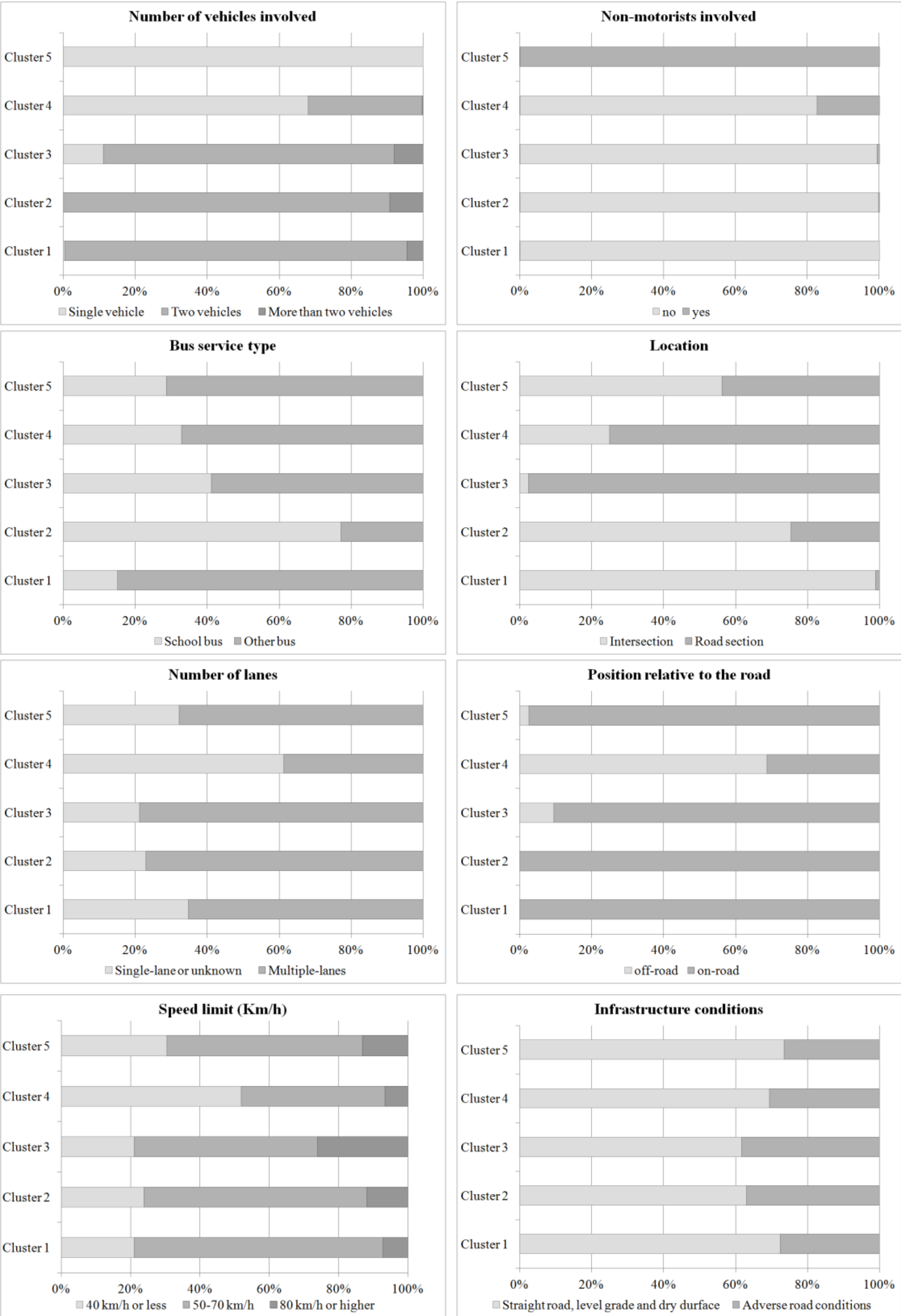


Figure 2 - Cluster comparison by feature

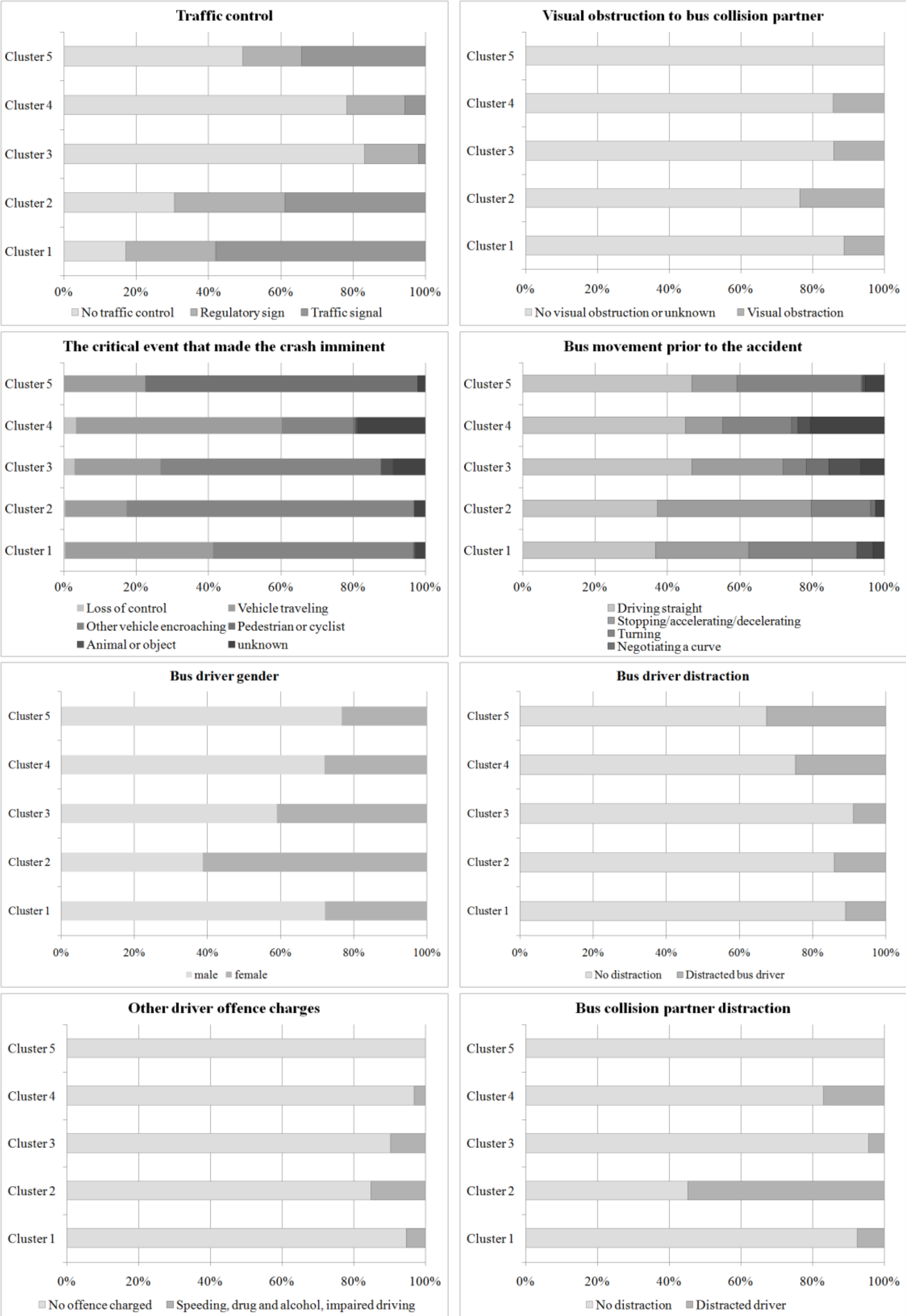


Figure 2 (continued) - Cluster comparison by feature

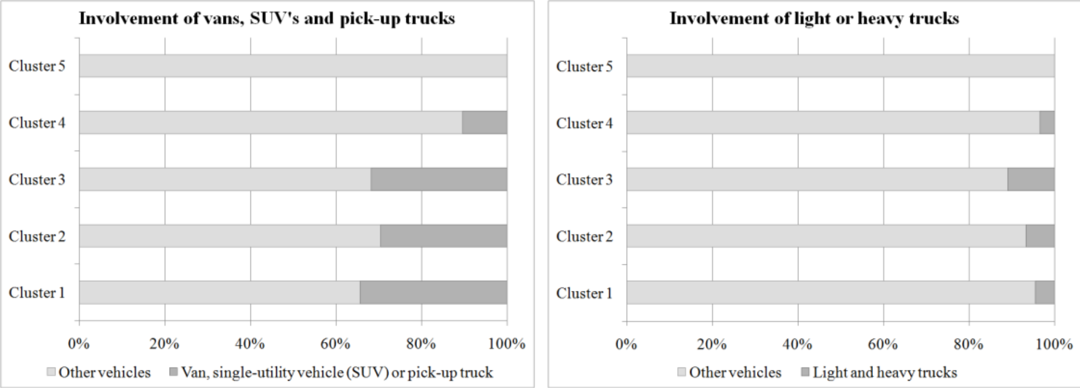


Figure 2 (continued) - Cluster comparison by feature

*Multi-vehicle collisions at intersections: vehicle encroaching or travelling* (cluster 1) –26.2% of the accidents. Accidents involved two vehicles (95.1%) and occurred at intersections (98.9%) mainly on multi-lane roads (65.2%). Most of the accidents in this cluster took place at signalized intersections (57.0%) or intersections with regulatory signs (24.8%). 71.8% of the accidents happened where the speed limit was between 40-70 kilometres per hour. 55.4% of the accidents occurred as a result of another vehicle encroaching, and 40.8% took place as a result of the bus travelling. 47.5% of the vehicle encroaching accidents occurred while the bus was going straight and 38.6% occurred while the bus was accelerating, decelerating or stopping in the traffic lane. 56.3% of the vehicle travelling accidents took place while the bus was turning, 22.6% occur took place the bus was going straight and another 10.3% took place when the bus was overtaking another vehicle. Interestingly, only a small fraction of these accidents involved visual obstruction (11.2%) distracted drivers (7.5%), speeding or impaired driving (6.7%). Notably, only 15.2% of these accidents involved school buses.

*Multi-vehicle collisions with school bus at intersection: distracted drivers* (cluster 2) – 21.5% of the accidents. Accidents typically involved two vehicles (90.8%) and occurred at intersections (75.4%) on multi-lane roads (76.9%). Most of the accidents in this cluster happened at signalized intersections (37.5%) or at intersections with regulatory signs (24.5%). 64.3% of the accidents took place where the speed ranged between 40-70 kilometres per hour. 77.2% of the involved buses were school buses (77.2%). The critical event that made the crash imminent was another vehicle encroaching into the lane (79.4%). In more than half of the accidents (54.8%), the bus collision partner was distracted, which was almost triple the average rate of driver distractions in the data (17.8%). The bus driver was distracted in 13.1% of the cases. A large share of the accidents took place while the bus was stopping, accelerating or decelerating in traffic lane (42.4%), and while the bus was going straight (37.4%). Another 16.5% of the accidents occurred during bus turning movements. 23.4% of the accident involved visual obstruction to bus collision partner and 36.9% involved difficult road conditions (i.e., curve, grade and wet surface). In 16.6% of the cases, the collision partner was charged with speeding or impaired driving, which was much higher than the average share of speeding or impaired driving in the data (9.2%). Interestingly, the share of female bus drivers in this cluster was 61.3%, which was about 50% higher than their average share in the data.

*Multi-vehicle collisions in road sections: infrastructure and traffic* (cluster 3) – 35.9% of the accidents. Accidents typically involved two vehicles (80.7%) and happened on road sections

(97.4%) of multi-lane roads (78.8%). Both school buses (41.3%) and other buses (58.8%) were involved in this type of accidents. 26.0% of the accidents occurred on roads where the speed limit was 80 kilometres per hour or higher, which was 73% higher than the share of such accidents in the data. 38.2% involved difficult road conditions (i.e., curve, grade and wet surface). The critical event that made the crash imminent was another vehicle encroaching in almost two third of the cases (60.9%) and another vehicle travelling in one fifth of the accidents (23.8%). Although most critical events occurred when the bus was either going straight (46.7%), accelerating or decelerating in traffic lane (25.3%), some accidents took place when the bus was turning (6.5%), overtaking (8.8%), or negotiating a curve (6.2%). In 14.5% of the cases, the collision partner was charged with speeding or impaired driving, which was much higher than the average share of speeding or impaired driving in the data. The share of bus collisions with trucks (10.9%) was higher than the average in the data.

*Single-vehicle bus accidents off-road: bus travelling and bus driver distraction at low speeds* (cluster 4) – 14.6% of the accidents. These were mostly single-vehicle accidents (68.1%) that occurred on road sections (74.9%) of single lane roads (61.4%). Non-motorists were involved in 17.3% of the accidents in this cluster. In contrast to the other clusters, most of the accidents happened off-road (68.8%) in areas where the speed limit was 40 kilometres per hour or less (51.8%). Only 33.0% of the buses involved are school buses. The critical event that made the crash imminent was mostly a vehicle travelling (56.9%), and in one quarter of the accidents (24.5%) the bus driver was distracted, which was almost twice the average share of distracted drivers in the data (13.1%). The bus movement prior to the critical event was mostly going straight (45.0%), turning (19.0%), and to a lesser extent stopping, accelerating or decelerating in traffic lane (10.3%) or parking (9.8%).

*Single-vehicle collisions with non-motorists: pedestrian and cyclists* (cluster 5) – 1.7% of the accidents. These were single-vehicle accidents (100.0%) involving non-motorists (100.0%). The accidents occurred mainly on multi-lane roads (67.7%) both at intersections (56.3%) and road sections. 28.8% of the accidents involved school buses. The critical event that made the crash imminent concerned pedestrians (50.5%) and cyclists (24.8%) on the road. The bus driver was distracted in one third or the accidents (32.5%), a triple rate than the average share of distracted bus drivers in the data (13.1%). The main bus movements prior to the accidents were going straight (46.8%), turning (34.4%), stopping, accelerating or decelerating in traffic lane (12.6%). The representation of turning movements was twice the average share in the data (17.1%). Interestingly, visual obstruction was not reported in any of the accidents. The share of female bus drivers in this cluster was 23.3%, which was roughly 50% lower than their average share in the data.

Figure 3 presents the comparison of the clusters by their severity and the share of each cluster to in the accident severity categories. Cluster 5, the collision of single-vehicle with a non-motorist bears the highest severity levels, as most of the accidents in this cluster result in non-capacitating and incapacitating injuries. In terms of the share of each cluster in the accident severity categories, 36.0% and 42.5% of the accidents are included in cluster 3 and cluster 5, respectively. 30.4% of the incapacitating injuries occur in cluster 3, while another 60% are almost equally divided among cluster 1, cluster 2 and cluster 5. One third of the non-incapacitating and possible injuries are included in cluster 1, and about 20% of these accidents are included in cluster 2. Cluster 3 generates 28.0% and 37.0% of the non-incapacitating and possible injuries, respectively.

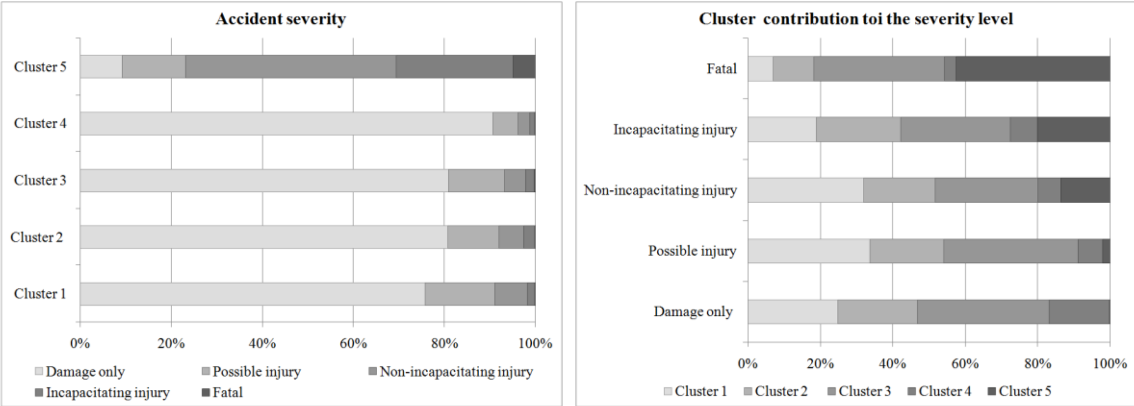


Figure 3 - Cluster comparison by severity (right); accident severity level by clusters (left)

# DISCUSSION AND CONCLUSIONS

The current study contributes to the body of knowledge about bus accidents by presenting a data driven taxonomy of bus accidents, namely the classification of bus accidents according to their recurrent features that emerge from the data. This issue is largely unexplored since thus far the study of Af Wåhlberg (2002, 2004) in Sweden is the only study that explored the typology of bus accidents. The bus accident patterns were extracted by employing innovative pattern recognition and classification methodologies. Specifically, the accident patterns were extracted from the data by using the SOM cluster analysis, which is an unsupervised learning neural network, and the produced clusters were confirmed by employing the ML one-layer, which is a supervised learning algorithm. The study provides an integrative and multi-faceted map containing five bus accident clusters that largely differ in their characteristics and their level of severity. The clusters are: (i) multi-vehicle collisions at intersections: vehicle encroaching or travelling (26.2%), (ii) multi -vehicle collisions with school bus at intersection: distracted drivers (21.5%), (iii) multi -vehicle collisions in road sections: infrastructure and traffic (35.9%), (iv) single-vehicle bus accidents off-road: bus travelling and bus driver distraction at low speeds (14.6%), and (v) single-vehicle collisions with non-motorists: pedestrian and cyclists (1.7%).

*Multi-vehicle collisions at intersections: vehicle encroaching or travelling*. These accidents involved two vehicles and occurred on multi-lane roads at signalized or regulated intersections where the speed limit was between 40-70 kilometres per hour. The two main critical events relevant to the accidents in this cluster were another vehicle encroaching while the bus was going straight or changing velocity and vehicle travelling while the bus was turning. Only a small fraction of these accidents involved visual obstruction, distracted drivers, speeding or impaired driving. Less than one sixth of these accidents involve school buses.

*Multi-vehicle collisions with school bus at intersection: distracted drivers*. These accidents typically involved two vehicles and occurred at signalized or regulated intersections on multi-lane roads, where the speed limit ranged between 40-70 kilometres per hour. The accidents mainly involve school buses and another vehicle encroaching into the lane as the critical event that made the crash imminent. A large share of the accidents took place while the bus was stopping, accelerating or decelerating in the traffic lane. In more than half of the accidents the bus collision partner was distracted, while the bus driver was distracted in less than a sixth of the cases. Visual obstruction was reported in roughly one fifth of the accidents and in one sixth of the cases the bus collision partner was charged with a risky behaviour. Female bus drivers were involved in almost two thirds of these incidents.

*Multi-vehicle collisions in road sections: infrastructure and traffic*. This cluster comprises slightly more than one third of the accidents and is a major contributor of injuries and fatalities to motorists. The accidents typically involved two vehicles on multi-lane road sections and were associated, though not exclusively, with high speed limits and adverse road conditions. Both school buses and other buses were involved in this type of accidents. The critical event that made the crash imminent was another vehicle encroaching in almost two third of the cases, and the bus was either going straight or changing acceleration. In slightly less than one sixth of the incidents risky behaviour of the bus collision partner was reported. The share of bus collisions with trucks was higher than the average in the data.

*Single-vehicle bus accidents off-road: bus travelling and bus driver distraction at low speeds*. This cluster mostly resulted in damage only accidents. These accidents occurred mainly off the roadway on single lane roads where the speed limit was fairly low. The critical event that made the crash imminent was typically another vehicle travelling, mostly during going straight and turning movements, but also during parking and changes in acceleration. In one quarter of these accidents the bus driver was distracted.

*Single-vehicle collisions with non-motorists: pedestrian and cyclists*. This cluster comprises a small fraction of the bus accidents, although it is the main contributor to severe injuries and fatalities. The accidents occurred mainly on multi-lane roads at intersections following the critical event of pedestrian or cyclist movement mainly when the bus was going straight or turning in areas where there was a clear line of sight. The accidents involved a distracted bus driver in one third of the accidents. Female drivers and school bus had a low rate of involvement in this type of accidents.

The analysis of the current study is essentially different from the analysis of Af Wåhlberg (2004), which was conducting by focusing on the crash results (e.g., injury in bus, injury outside of bus, hit object, etc.) and hence is not directly comparable. However, it seems that, while some similarities do exist, bus accident taxonomy is not easily transferable across world regions. For example, the study of Af Wåhlberg (2004) linked the event type “bus shunt by another vehicle” with car (63.4%) and bus (29.3%) as the other vehicle, bus stop (62.8%) as the location, and slippery road (40.9%) as the environmental condition. The current analysis shows that bus movements for stopping, accelerating and decelerating at traffic lanes are indeed important factors in vehicle encroaching incidents, in addition to bus turns and straight movement. The incidents occurred mainly at signalized or regulated intersections, but the location of the bus stop in the current data set was unclear. Cars served as collision partners in about 60.0% of the incidents, but the main collision partners in the remaining incidents were pick-up trucks, SUV’s and vans, rather than other buses. Adverse road conditions indeed existed in about 40% of these accidents, but this share did not vary greatly across different accident types. Additional risk-factors identified by the current study are driver distraction, visual obstruction, and risky road behaviour of the bus collision partner. Naturally, due to differences in the crash reporting system, infrastructure, scale and employed methodologies, further research is needed to establish the comparability of accident patterns across countries.

Nevertheless, both studies demonstrate the importance of country-specific identification of the prevailing bus accident patterns.

The results of the current analysis indicate several main research directions towards improving the safety of transit, coach and school bus operations.

Firstly, the current study indicates that non-motorists injuries and fatalities are mainly associated with conflicts between the bus and non-motorized road users and are not related to the bus conflicts with other vehicles. A similar conclusion is reached also in the former study of Af Wåhlberg (2004). Although the share of these accidents is fairly small, they bear severe consequences in terms of injuries and fatalities. Hence, essential steps towards improving the safety of bus operations should include in-depth analysis of the conflicts between bus and non-motorist at intersections, investigating the effectiveness of road design elements for mitigating these conflicts, and exploring the effectiveness of electronic warning systems in increasing driver awareness of these conflicts.

Secondly, the current analysis indicates that accidents involving another vehicle encroaching while a school bus is stopping or changing velocity at intersections on-multilane roads are related to bus collision partner distraction. This issue is puzzling since school buses in the United States differ from other buses in their colour and shape, thus increasing public awareness of their presence. Moreover, the element of driver distraction does not occur in accidents involving another vehicle encroaching on regular bus services. Hence, it is important to further explore the exact reasons for driver distraction as well as to attentively explore the difference in the bus movements versus the general traffic in order to identify the possible reasons underlying this problem. Also, it is important to explore manners for further increasing the attention of drivers in the general traffic to school bus movements.

Thirdly, the current analysis indicates that the main critical events in two vehicle bus accidents at intersections involving regular bus services are and are another vehicle encroaching while the bus is going straight or changing velocity and bus travelling while turning. Interestingly, the data does not reveal any particular behavioural reason for this phenomena (i.e., vehicle speeding, driver distraction, impaired driving). Hence, the risk factor underlying these incidents including intersection design, and specific conflicts due differences between the bus movements and the general traffic should be explored.

Fourthly, while the share of distracted bus drivers in the sample was 13%, driver distraction bears serious consequences in terms of accident severity, since the rate of distracted drivers in bus non-motorists conflicts was 32.5%. Consequently, bus operators should explore the reasons for bus drivers' lack of attention such as cell phone use by drivers and passengers, listening to the radio and conversations between drivers and passengers, as well as explore various measures to increase the driver attention under different circumstances.

Fifthly, the current analysis identifies a problem of multi-vehicle collisions on road sections. This cluster is significant both in terms of size and severity consequences, but seems to be rather heterogeneous in terms of bus service type involvement, accident characteristics, infrastructure, and behavioural elements. Hence, it seems that this type of accidents should be further investigated.

Sixthly, roughly 30% of the multi-vehicle bus accidents involved pick-up trucks, sport utility vehicles or vans, while an additional 10% of the accidents on road sections involved light and heavy trucks. Hence a possible research direction would be to identify possible reasons for conflicts among buses and these vehicles, such as perceptions regarding the right of way, and inherent visual obstruction related to vehicle design features.

Seventhly, bus drivers in the beginning of their career (until the age of 34) comprised roughly 20% of the bus accidents, and the share of young drivers among the bus collision partner was about 25%. These results indicate that novice young drivers are related to problems in car-bus conflict avoidance and management. Moreover, while these shares do not vary much across clusters, but may be related to accident severity within clusters. Hence the role of bus driver age in bus car conflict management and its potential impact on accident severity should be further explored.

Last, although off-road single-vehicle bus accidents at low speeds are mostly damage only accidents and hence may seem less interesting, they might be an indicator to deficiencies in the organizational safety policy of bus operators. An interesting research direction would be to explore the linkage between this type of accidents and other types of accidents for bus operators.

# REFERENCES

Af Wåhlberg, A.E. (2004a). Characteristics of low speed accidents with buses in public transport: part II. Accident Anal Prev, 36, 63-71.

Af Wåhlberg, A.E. (2004b). The stability of driver acceleration behavior, and a replication of its relation to bus accidents. Accident Anal Prev, 36, 83-92.

Af Wåhlberg, A.E. (2007). Aggregation of driver celeration behavior data: Effects on stability and accident prediction. Safety Sci, 45, 487-500.

Af Wåhlberg, A.E. (2008a). The relation of non-culpable traffic incidents to bus drivers' celeration behavior. Journal of Safety Research, 39, 41-46.

Af Wåhlberg, A.E. (2008b). If you can't take the heat: Influences of temperature on bus accident rates. Safety Science, 46, 66-71.

Af Wåhlberg, A.E., and L. Dorn (2009). Absence behavior as traffic crash predictor in bus drivers. J Safety Res, 40, 197-201.

Björnstig U., P. Albertsson, J. Björnstig, P-O Bylund, T. Falkmer, and J. Petzäll (2005). Injury events among bus and coach occupants: non-crash injuries as important as crash injuries. IATSS Res, 29, 79-87.

Blower, D., and P.E. Green (2010). Type of motor carrier and driver history in fatal bus crashes. Transport Res Rec, 2194, 37-43.

Chang, H.L., and C.C. Yeh (2005). Factors affecting the safety performance of bus companies - The experience of Taiwan bus deregulation. Safety Sci, 43, 323-344.

Estache, A., and A. Gómez-Lobo (2005). Limits to competition in urban bus services in developing countries. Transport Rev, 25, 139-158.

Jayatilleke, A.U., S. Nakahara, S.D. Dharmaratne, A.C. Jayatilleke, K.C. Poudel, and M. Jimba (2009). Working conditions of bus drivers in the private sector and bus crashes in Kandy district, Sri Lanka: a case-control study. Inj Prev, 15, 80-86.

Kohonen, T. (2001). Self-Organizing Maps, 3rd edition. Springer-Verlag, Berlin, Heidelberg, Germany.

Lee, S., and M.M. Crawford (2005). Unsupervised multistage image classification using hierarchical clustering with a Bayesian similarity measure. IEEE T Image Process, 14, 312-320.

Martinetz, T.M., S.G. Berkovich, and K.J. Schulten (1993). Neural-gas network for vector quantization and its application to time-series prediction. IEEE T Neural Networ, 4, 558-569.

Mirza, S., M. Mirza, H. Chotani, and S. Luby (1999). Risky behavior of bus commuters and bus drivers in Karachi, Pakistan. Accident Anal Prev, 31, 329-333.

National Highway and Traffic Safety Administration (2010). National Automotive Sampling System (NASS) General Estimates System (GES): Analytical Users Manual 1988-2009. Report number DOT HS 811-355.

National Transportation Safety Board (2011). Most wanted list – Bus occupant safety. http://www.ntsb.gov/safety/mwl.html.

Pearce, T., D.A.C. Maunder, T.C. Mbara, D.M. Babu, and T. Rwebangira (2000). Bus accidents in India, Nepal, Tanzania, and Zimbabwe. Transport Res Rec, 1726, 16-23.

Peltarion (2012). Synapse software, http://www.peltarion.com/products/synapse/.

Prato, C.G., V. Gitelman, and S. Bekhor (2012). Mapping patterns of pedestrian fatal accidents in Israel. Accident Anal Prev, 44, 56-62.

Razmpa, E., K.S. Niat, and B. Saedi (2011). Urban bus drivers' sleep problems and crash accidents. Indian J Otolaryngol, 63, 269-273.

Salminen, S., M. Vartia, and T. Giorgiani (2009). Occupational injuries of immigrant and Finnish bus drivers. J Safety Res, 40, 203-205.

Shahla, F., A.S. Shalaby, B.N. Persaud, and A. Hadayeghi (2009). Transport Res Rec, 2102, 108–114.

Rahman, M., L. Kattan, and R. Tay (2011). Injury risks in collisions involving buses in Alberta. Proceedings of the 91st TRB Annual Meeting, Washington, D.C.

Vesanto, J., and E. Alhoniemi (2000). Clustering of the self-organizing map. IEEE T Neural Networ, 11 (3), 586-600.

Yang, J., C. Peek-Asa, G. Cheng, E. Heiden, S. Falb, and M. Ramirez (2009). Incidence and characteristics of school bus crashes and injuries. Accident Anal Prev, 41, 336-341.

Zegeer, C.V, H. Huang, J. Stutts, E. Rodgman, and E. Hummer (1995). Commercial bus accident characteristics and roadway treatments. Transport Res Rec,1467, 14-22.