Article:

Pariota, L orcid.org/0000-0001-9173-666X, Bifulco, GN, Galante, F et al. (2 more authors) (2016) Longitudinal control behaviour: Analysis and modelling based on experimental surveys in Italy and the UK. Accident Analysis and Prevention, 89. pp. 74-87. ISSN 0001-4575

https://doi.org/10.1016/j.aap.2016.01.007

Longitudinal control behaviour: analysis and modelling based on experimental surveys in Italy and the UK

Luigi Pariota[1], Gennaro Nicola Bifulco[1], Francesco Galante[1], Alfonso Montella[1], Mark Brackstone[2]

Abstract

This paper analyses driving behaviour in car-following conditions, based on extensive individual vehicle data collected during experimental field surveys carried out in Italy and the UK. The aim is to contribute to identify simple evidence to be exploited in the ongoing process of driving assistance and automation which, in turn, would reduce rear-end crashes. In particular, identification of differences and similarities in observed car-following behaviours for different samples of drivers could justify common tuning, at a European or worldwide level, of a technological solution aimed at active safety, or, in the event of differences, could suggest the most critical aspects to be taken into account for localisation or customisation of driving assistance solutions. Without intending to be exhaustive, this paper moves one step in this direction. Indeed, driving behaviour and human errors are considered to be among the main crash contributory factors, and a promising approach for safety improvement is the progressive introduction of increasing levels of driving automation in next-generation vehicles, according to the active/preventive safety approach. However, the more advanced the system, the more complex will be the integration in the vehicle, and the interaction with the driver may sometimes become unproductive, or risky, should the driver be removed from the driving control loop. Thus, implementation of these systems will require the interaction of human driving logics with automation logics and then an enhanced ability in modelling drivers' behaviour. This will allow both higher active-safety levels and higher user acceptance to be achieved, thus ensuring that the driver is always in the control loop, even if his/her role is limited to supervising the automatic logic. Currently, the driving mode most targeted by driving assistance systems is longitudinal driving. This is required in various driving conditions, among which car-following assumes key importance because of the huge number of rear-end crashes.

The increased availability of lower-cost information and communication technologies (ICTs) has enhanced the possibility of collecting copious and reliable car-following individual vehicle data. In this work, data collected from three different experiments, two carried out in Italy and one in the UK, are analysed and compared. The experiments involved 146 drivers (105 Italian drivers and 41 UK drivers). Data were collected by two instrumented vehicles.

Our analysis focused on inter-vehicular spacing in equilibrium car-following conditions. We observed that (i) the adopted equilibrium spacing can be fitted using lognormal distributions, (ii) the adopted equilibrium spacing increases with speed, and (iii) the dispersion between drivers increases with speed. In addition, according to different headway thresholds (up to 1 second) a significant number of potentially dangerous behaviours is observed.

Three different car-following paradigms are also applied to each of the experiments, and modelling parameters are calibrated and compared to obtain indirect confirmation about the observed similarities and differences in driving behaviour.

Keywords: Active road safety, Driving automation, Driving behaviour, Car-following, Instrumented vehicles, Cross-country experiments.

[1]Department of Civil, Building and Environmental Engineering
University of Naples Federico II
Via Claudio 21, 80125 Naples, Italy

[2]TSS-Transport Simulation Systems Ltd, London, United Kingdom
9 Devonshire Square, London, EC2M 4YF, UK

1. Introduction

Reconciling mobility needs with efficient and more sustainable transportation, hence reducing highway crashes, injuries and fatalities, is a key objective in the transportation sector. Given that human factors are considered to contribute to 93% of vehicle crashes (AASHTO, 2010), driving automation at different levels is seen as having the potential to improve highway safety. Such automation levels have been classified in slightly different ways by different organisations. For instance, as reported by ERTRAC (2015), the Society of Automotive Engineers (SAE) considers six automation levels, ranging from 0 to 5. These approximately correspond to the five levels (from 0 to 4) identified by the National Highway Traffic Safety Administration (2013a). SAE level 1 of automation has been deployed for several years (e.g., adaptive cruise control, lane-keeping assistance, etc.), level 2 systems have more recently emerged (e.g., automated parking, adaptive cruise control with stop-and-go and/or truck platooning, etc.) and introduction of level 3 is now under discussion (e.g., combination of adaptive cruise control and lane changing/overtaking systems).

Interestingly, increasing levels of automation require that the driver is even more at the centre of the innovation and design process (Trimble et al., 2014). Indeed, as automation moves from one level to another, the required driving performance shifts from full driver responsibility to co-responsibility with assisting or automation systems. Levels 1 and 2 require that the automation logic interacts with driver behaviour to accomplish complete driving tasks, and higher automation levels also require that the driver is kept in the vehicle control loop, since at level 4 he/she is responsible for the transition from automated to non-automated tasks (or between different automated tasks), and at level 3 he/she can also be required to recall the control of the vehicle should automation fail.

Correct understanding of driving behaviour may allow human errors and mis-intervention to be identified, thus enabling driver intervention to be replaced by automatic control or warning (Muhrer and Vollrath, 2010). Poor interaction between the automation logic and its human counterpart (possibly due to scarce knowledge of driver behaviour) results in poor safety performance.

According to evidence from the scientific literature, drivers may experience problems in regaining control also in the case of moderate automation, possibly due to over-reliance on vehicle systems and/or reduced situational awareness (Vollrath et al., 2011), and the debate on the possible effects of automated driving is far from concluded. As stated in Horrey et al. (2015) and in Lee and See (2004), the discrepancy between the driver’s and the system’s estimates of the state of the world could result in operator-system conflicts, reduced trust in the system and, ultimately, system disuse. That said, the way automation is designed and deployed (e.g., human-like driving assistance systems – Bifulco et al., 2008) strongly impacts on the effects, especially in terms of road safety (Strand et al., 2014). As a consequence, research efforts in this field are being stepped up in every aspect, such as a-priori acceptability (Payre et al., 2014), the driver’s reaction in the event of automation failures (Strand et al., 2014), the influence of automation on the behaviour of other unequipped vehicles (Gouy et al., 2014), and so forth.

Within the vast field of automation, longitudinal control of the vehicle has so far been one of the more widely addressed aspects. This applies to different driving tasks; for instance, Intelligent Speed Adaptation (ISA) mainly works in free-flow conditions, while Adaptive Cruise Control (ACC) and Autonomous Emergency Breaking (AEB) mainly work in car-following conditions. Assisting and automation solutions related to car-following conditions are among the most effective with respect to safety, as they deal with relative speed and spacing (that is with the bumper-to-bumper distance between the leader and the following vehicles and associated time-headway and/or time-to-collision). This affects the occurrence of rear-end crashes, that are mainly caused by the failure of the driver to brake sufficiently early to avoid the collision. The risk of a rear-end crash increases exponentially as the headway time gap decreases. Drivers should maintain a safe headway behind the vehicle ahead, such that they can stop safely in order to avoid a collision if necessary, and most researchers agree that less than two seconds' headway is insufficient and unsafe (a recent review of the literature on this topic can be found in Austroads, 2015, where the poor headway maintenance is evidenced as one of the principal rear-end crash contributory factor). However, many drivers keep headways lower than the time to perceive and react to an unexpected change in traffic ahead (Austroads, 2015), i.e., their perception and reaction time (PRT). As the driver headway is below their PRT, there is a reliance on viewing traffic conditions ahead of the vehicle being followed even though this does not permit the driver to anticipate sudden braking by the vehicle immediately ahead (Hutchinson, 2008). Motorists find it difficult to maintain a sufficient headway in denser traffic, given that lane changes occur frequently. Exacerbating the issue is that most motorists are unaware of what constitutes a safe headway, or how to determine it (Song and Wang, 2010). Following a vehicle with insufficient time headway is defined tailgating and is one of the most dangerous and aggressive driving behaviours, representing a major cause of rear-end crashes.

Crash statistics highlight the major role of both rear-end crashes and the tailgating contributory factor. In Italy, according to the Italian National Institute of Statistics, tailgating was the main contributory factor in 17% of crashes in the three-year period 2011-2013. In the same period, rear-end crashes accounted for 47% of the crashes on motorways and 30% of the crashes on national highways. In the UK, following too close (equivalent to tailgating) was the main contributory factors of 16% of the crashes in the motorways and 9% of the crashes in the A roads in the three-year period 2011-2013 (Department for Transport, 2012, 2013, 2014). In the same period, rear-end crashes accounted for 32% of the crashes in the US (NHTSA, 2012, 2013b, 2014).

We are strongly convinced that in-depth understanding of the peculiarities and criticalities of natural (SAE-level 0, without automation) behaviour of drivers in car-following is a pre-requisite for proper design of driving automation solutions aimed at longitudinal control of vehicles and active-safety systems to prevent rear-end crashes.

In light of the above considerations, the aim of this work is to increase the number of available analyses, based on different experimental surveys, in the long-running discussion on whether and in which way characteristic car-following patterns can be identified in a population of drivers. The paper compares observed characteristic patterns, interpreted within the framework of car-following theories and with reference to two different countries, in an attempt to gain insight into car-following behaviours, as well as their variability across both different drivers of a given population and different populations of drivers. Identification of differences and similarities for different samples of drivers could justify common tuning at a European or worldwide level for a technological solution aimed at active safety, or, in the case of differences, could identify the most critical aspects to be taken into account for localisation/customisation of driving assistance solutions. Without the ambition to be exhaustive, this paper moves a step in the former direction.

We analyse the observed behaviours of different samples of drivers by adopting two different approaches. On the one hand, we directly observe the data collected during car-following sessions, with particular reference to the distribution across different populations of observed adopted spacing (viewed in relation to different speeds). On the other, we adopt some reference models and compare the values of the parameters of these models after calibration against observed data. Comparison of parameters in different samples allows indirect comparison of observed behaviours. In order to enhance the robustness of the model-based analyses, the approach is performed three times, with reference to three different models. The role of car-following models is thereby extended from the traditional field of simulation models for traffic networks, for which they constitute an important part in order to simulate longitudinal movement of vehicles, to the field of driving assistance, oriented to road safety.

The paper is structured as follows. In section 2 the theoretical framework employed to interpret observed behaviours is briefly presented: the main concepts and variables adopted in car-following theories are introduced and discussed with reference to simpler theories, as in Pariota and Bifulco (2015) where the suitability of using simpler paradigms is shown (albeit with reference to a specific theoretical paradigm). In this section the reference models used for comparing observed behaviours via calibrated parameters are also specified. Given that the proposed research is based on actual car-following behaviour rather than on mere theory, field data are of paramount importance and in section 3 the technologies employed for observing car-following behaviours are summarised. Section 4 presents the three experimental surveys carried out and the main characteristics of the resulting datasets, to be analysed and compared. In section 5 direct analysis is performed on longitudinal control behaviour; the method employed to identify the equilibrium conditions is presented, as well as the associated results. In section 6 indirect analysis based on the calibration of the reference models is presented and discussed. Finally, in section 7 conclusions are drawn, comparing equilibrium conditions and modelling parameters for the different experiments in different countries.

2. The use of driving simulators for research in car-following conditions

Vehicle control during driving can be divided into two macro-tasks: longitudinal control and lateral control. Lateral control mainly consists in steering, and letting the vehicle keep its position in the lane and/or perform lane changing. Longitudinal control is mainly actuated by using pedals, and consists in moving the vehicle forward within a lane, applying appropriate acceleration and speed. It can in turn be classified with reference to three main conditions: free-flow (no interaction with vehicles ahead), car-following and emergency braking. The purpose of this work is to analyse the car-following aspects of longitudinal control.

The approach adopted in this work is both to directly analyse observed car-following conditions and to refer to car-following theory in order to identify specific variables and modelling parameters able to represent the phenomenon. Identification of longitudinal control with the mere car-following condition is justified by our field of application, which is the introduction of assisting and automation solutions for road safety. Indeed, many of the driving assistance solutions introduced so far in the automotive arena deal with concepts like the time-headway between a leading and following vehicle in a traffic stream, relative speed, time-to-collision and other variables that are typical of car-following theory and that play a crucial role in road safety, especially for rear-end collisions (Johansson et al., 2004). This is the case, for example, of ACC (Adaptive Cruise Control) where adaptation of the speed is generally based on the time-headway, or AEB (Autonomous Emergency Braking) where emergency braking is activated for critical values of the time-to-collision.

In the absence of automation, car-following models describe the behaviour of a following vehicle as a function of the trajectory of the leading one. Thus, depending on the leader's trajectory, the car-following model can be employed to estimate the kinematic trajectory the driver imposes on the vehicle. According to Saifuzzaman and Zheng (2014), car-following models can be divided into several main categories, such as stimulus-based models, safety distance models and psychophysical or action point models. Other minor streams are represented by so-called linear models and by models based on fuzzy logic. In stimulus-based models the acceleration of the follower is determined by the reaction to the relative velocity and spacing from the leading vehicle. Such models were made popular by the paradigms formulated in the late 1950s as part of the so-called General Motors (GM) model (Gazis et al., 1961), which has continued in use to the present day (Bifulco et al., 2008). The safety distance approach was developed by arguing that a follower maintains a distance and a speed that allow safe braking if the leader slows abruptly. Formulations of this collision-avoidance approach can be found in Kometani and Sasaki (1958) and in Gipps (1981), where multiregime formulations are introduced in order to also deal with non-car-following or different car-following conditions. The psychophysical approach is developed from behavioural analysis on human cognitive and decision-making mechanisms. Such models may be considered to have been initiated by Michaels (1963) who showed that drivers respond to changes in the perceived size of the vehicle ahead, arguing that the response is actuated by drivers (by depressing or releasing the gas pedal and/or the brake) at given thresholds of the perceived variation of the apparent size of the vehicle ahead. These thresholds can differ according to whether the distance from the vehicle ahead is increasing or decreasing. Models developed according to this paradigm were later framed within the well-known action point theory, as developed, among others, by Wiedemann (1974) and Wagner (2011), where car-following behaviour oscillates within the thresholds and, in the event of the leader's steady-state speed, the centre of this oscillation is an equilibrium point with a null relative speed and a given inter-vehicular spacing. Equilibrium is well described (see point A) in Figure 1,where relative speed between vehicles (leader and follower) is plotted against inter-vehicular spacing. This equilibrium point, as argued from the cognitive and decision-making point of view, is not in conflict with the stimulus-based or safety-based theories, and is also confirmed by experimental observations. The distribution of these equilibria is analysed in this work with reference to three different datasets in order to identify recurrent patterns and general behaviours.

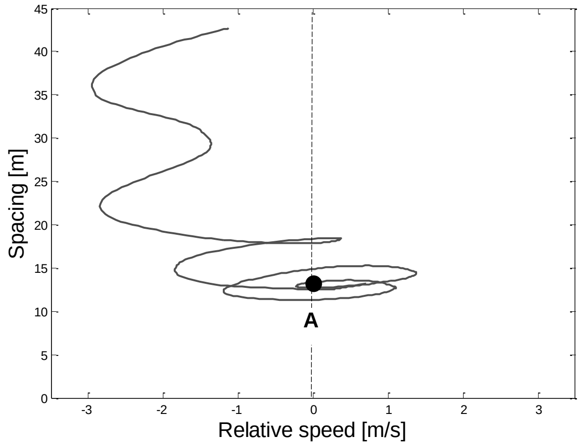

Fig. 1 Car-following oscillations around an equilibrium point

Three interpretative paradigms are also applied to the observed data. These models are based on previous works, namely Pariota et al. (2015 online publication), Bifulco et al. (2013a) and Bifulco et al. (2013b), and allow the characterisation of the observed car-following behaviours. We chose the above three approaches for practical reasons: they adopt as independent variables precisely the data that can be collected by instrumented vehicles, and have few parameters to calibrate, arranged in very simple analytical structures. It is worth pointing out that complex models often fail to outperform simple ones and the adopted approaches have been shown to be no less accurate than others with more complex analytical structures. Moreover, the risk of overfitting (or at least unnecessary over-complexity) is high in the case of models with several parameters (Punzo et al., 2015).

The first interpretative paradigm is based on a state-space model applied to car following. Indeed, car following can be viewed as a time-continuous dynamic process; for instance, it can be represented (see Wilson, 2008) with equation 1), where for the sake of simplicity the so-called additive acceleration (representing a random noise) has been omitted:

$$\dot{v}_n^t = f\left(\Delta x_n^t, \Delta v_n^t, v_n^t\right) \qquad 1)$$

where:

n is the following vehicle and n-1 the leader,

$\dot{v}_n^t$ is the acceleration planned to be applied by the follower as a decision taken at time instant t,

Δx_n^t is the spacing between the leader and the follower at time instant t,

Δv_n^t is the relative speed between the leader and the follower at time instant t,

$f(\cdot)$ is the acceleration function, which formally represents the car-following paradigm.

Considering S^t_n is the absolute (unidirectional) position of the follower and a^t_n is its acceleration, while S^t_{n-1} is the absolute position of the leader and v^t_{n-1} its speed, the definition below holds:

$$\dot{S}_n^t = v_n^t, \qquad \dot{v}_n^t = a_n^t, \qquad \Delta x_n^t = S_{n-1}^t - S_n^t, \qquad \Delta v_n^t = v_{n-1}^t - v_n^t, \qquad \Delta\dot{x}_n^t = \Delta v_n^t$$

From the definition of equilibrium conditions the following results are obtained:

$$\dot{v}_n^* = 0 \qquad \Delta v_n^* = 0 \rightarrow v_n^* = v_{n-1}^* \qquad \Delta x_n^* = g(v_n^*)$$

where the equilibrium spacing (Δx_n^*) has been explicitly assumed as a function of the cruising speed. Adopting for equation 1 a Taylor's expansion at the equilibrium point, it can be written that

$$\dot{v}_n^t = \frac{\partial f(\cdot)}{\partial \Delta x_n}(\Delta x_n^t - \Delta x_n^*) + \frac{\partial f(\cdot)}{\partial \Delta v_n}(\Delta v_n^t - 0) + \frac{\partial f(\cdot)}{\partial v_n}(v_n^t - v_n^*) \qquad 2)$$

Now, given that at equilibrium the cruising speed does not significantly differ from that of equilibrium, the last term of equation (2) can be neglected. Moreover, having defined

$$\frac{\partial f(\cdot)}{\partial \Delta x_n} = \omega_1 \quad \text{and} \quad \frac{\partial f(\cdot)}{\partial \Delta v_n} = \omega_2,$$

and given that $\dot{v}_n^t = \dot{v}_{n-1}^t - \Delta\dot{v}_n^t$, we may write $\Delta\dot{v}_n^t = -\omega_1\,\Delta x_n^t - \omega_2\,\Delta v_n^t + \omega_1\,\Delta x_n^* + \dot{v}_{n-1}^t$

This can be arranged in the classical state-space representation as: $\dot{\bar{x}}_n^t = A\ \bar{x}_n^t + B\ \bar{u}^t$

where $A = \begin{bmatrix} 0 & 1 \\ -\omega_1 & -\omega_2 \end{bmatrix}$ is the state matrix, $B = \begin{bmatrix} 0 & 0 \\ \omega_1 & 1 \end{bmatrix}$ is the input matrix, $\bar{x}_n^t = \begin{pmatrix} \Delta x_n^t \\ \Delta v_n^t \end{pmatrix}$ is the state vector,

$\bar{u}^t = \begin{pmatrix} \Delta x_n^* \\ \dot{v}_{n-1}^t \end{pmatrix}$ is the input vector, and $\dot{\bar{x}}_n^t = \frac{d}{dt}\bar{x}_n^t = \begin{pmatrix} \Delta\dot{x}_n^t \\ \Delta\dot{v}_n^t \end{pmatrix}$ is the first derivative of the status vector.

Once the equilibrium spacing is identified (at a given cruising speed), the values of the parameters of the model (ω_1 and ω_2) depend on the driver's behaviour. The two parameters jointly influence the tendency of the driver to close the gap with the preceding vehicle, while ω_1 alone explains the tendency of the follower to stay far from the leader because of the equilibrium.

The second interpretative paradigm is based on a stimulus-response approach. It has been shown (Bifulco et al., 2013a) to mimic real car-following behaviour once calibrated properly. It aims to enable a human-like control logic (Simonelli et al., 2009) for ACC systems. The time-discrete linear model relates on one side the instantaneous speeds of the leader and the follower and their spacing with, on the other, the target spacing the follower desires for the next simulation time-step. The main equation is:

$$\Delta\xi(k, k+1) = \beta_0 + \beta_1 \cdot \Delta x(k) + \beta_2 \cdot v_n(k) + \beta_3 \cdot v_{n-1}(k) \qquad 3)$$

where $\Delta\xi(k, k+1)$ is the target inter-vehicular spacing estimated at time interval k for time interval k+1, and all other variables are the same as the state-space approach but refer to a discrete-time approach, where k is any reference time interval and k+1 is the subsequent one. Since $s_n(k) = \Delta x(k) + s_{n-1}(k)$ and $v_n(k) = \Delta v(k) + v_{n-1}(k)$, the variables of this second paradigm are no different from those of the first. The linear stimulus-response model is easy to calibrate (identification of the parameters from β_0 to β_3). The most efficient algorithm for this aim is the recursive least squares (RLS) algorithm, which can also be applied at run-time, during data collection. Calibration and validation tests carried out in previous works have shown good agreement between simulated and observed trajectories, with typical errors never exceeding 77%, less than 20% in 60% of cases and in 40% of cases less than 10% (Bifulco et al., 2013a). Interestingly, the model is not restricted to the equilibrium car-following condition and does not require identification of the equilibrium spacing.

The third interpretative paradigm is based on the definition of car-following waves given in Bifulco et al. (2013b) and on an appropriate representation of action points. If car-following spirals are represented in the plane $-\Delta V/\Delta X$ vs. ΔV and if, according to Pariota and Bifulco (2015), only the action points associated to the speed thresholds are plotted, a linear pattern can be estimated, similar to what is depicted in Figure 2 below, where several (and different) observed car-following trajectories are employed for estimation.

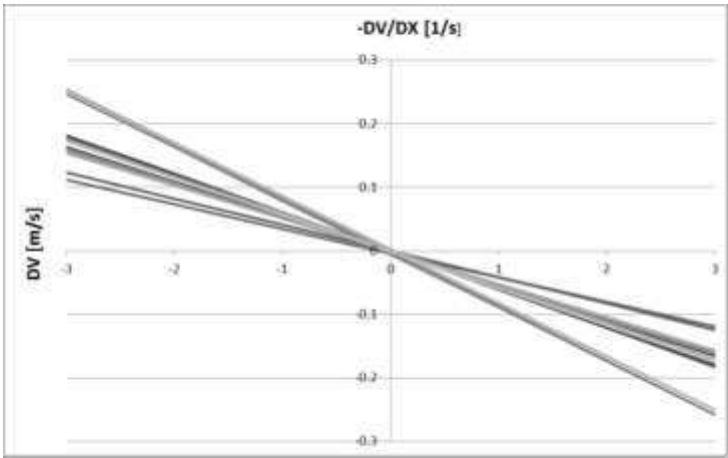

Fig. 2 Linear pattern for action points

Estimation of the linear pattern results in identifying two parameters, with the intercept parameters observed to be null in all estimated cases. Differences in terms of the obtained parameters for different car-following observations can be employed as a measure of the differences between the observed driving patterns. Also in this case the models do not apply only to equilibrium points.

3. Observing car following

Identification of car-following behaviour and calibration of car-following models require the use of car-following data. For many years the availability of enough data, as well as the quality of these data, has been an issue. This was mainly due to the problem of observing and recording disaggregate data. As a result, aggregate traffic flow data (e.g., flows, densities, etc.) were very often used for the calibration of microscopic (disaggregated) traffic models. However, for any research activity in the field of car following, especially if applied to driving assistance or automation, a reasonable quality of microscopic data is a prerequisite. Fortunately, some sources of microscopic car-following data are now much more easily available, mainly thanks to the rapid development of ICT.

One of the new sources of microscopic car-following data is represented by sensing devices placed outside the traffic stream. Some examples in this field are given by the NGSIM project (Alexiadis et al., 2004) in the US and the Motorway Incident Detection and Signalling (MIDAS) data archive collected by the Highways Agency in the UK (Wilson, 2008). In the NGSIM data microscopic traffic observations are available for a 500-metre segment of the I-80 interstate freeway. The I-80 dataset is collected using digital video cameras allowing identification and tracking of each vehicle in the traffic stream with a time-resolution of 0.1 seconds. However, the trajectories observed are quite short, the observations are fixed outside the traffic stream (Eulerian observation approach), and some errors and noise in the observed data have been revealed in recent studies (Hamdar and Mahmassani, 2008). A source of data similar to NGSIM comes from the Seohaean freeway in Korea, where data for estimating rear-end crash potential were prepared and analysed from 1h video images related to the 3-lane segment near the Maesong interchange (Oh and Kim, 2010). The MIDAS system in the UK is based on loop detectors placed on the M25 London Orbital Motorway from junction 10 to junction 16. Strictly speaking, data from MIDAS relate to macroscopic traffic characteristics (average speed, density and flow); however, given the high density of traffic sensors for some segments of the motorway, re-identification algorithms were proposed (Lunt et al., 2006) in order to estimate vehicular trajectories intended as disaggregated data about individual speeds and loop-crossing times.

Another kind of ICT-enabled source of data for car-following behaviour is represented by real-time kinematic (RTK) GPS. Use of such data is widespread in the scientific community (Gurusinghe et al., 2002; Ranjitkar et al., 2003; Brockfeld et al., 2004). However, some issues have to be addressed in using this technique: at least two vehicles equipped with GPS are needed, one acting as the leader and the other (immediately to the rear) as the follower. Maintaining uninterrupted car following between these two vehicles for a suitable duration of time is not trivial and making an effort to do so could introduce some bias into the observed behaviour.

The third kind of data source is the one adopted in this paper: it entails the use of an instrumented vehicle (Boyce and Geller, 2002; Brackstone et al., 2002; Ma and Andreasson, 2007; Brackstone et al., 2009). An instrumented vehicle (IV) is a normal car (similar to any other and as unrecognizable as possible within a traffic stream) that is equipped with a data acquisition system and one or more sensors able to detect neighbouring vehicles (Tarko et al., 2013; Montella et al., 2014). A dedicated on-board computer processes signals and data and stores the measured information, recording real data during driving sessions. The information in question relates to the microscopic characteristics of the traffic flow, such as relative speed and spacing, as well as data associated to the vehicle’s own dynamics or to driving variables detected by querying the on-board devices, electronic control units (ECUs) and commands. However, the data could also include road conditions or, in some cases, the monitoring of the driver’s health or attention or physiological status. A typical architecture for an instrumented vehicle is as described in Figure 3.

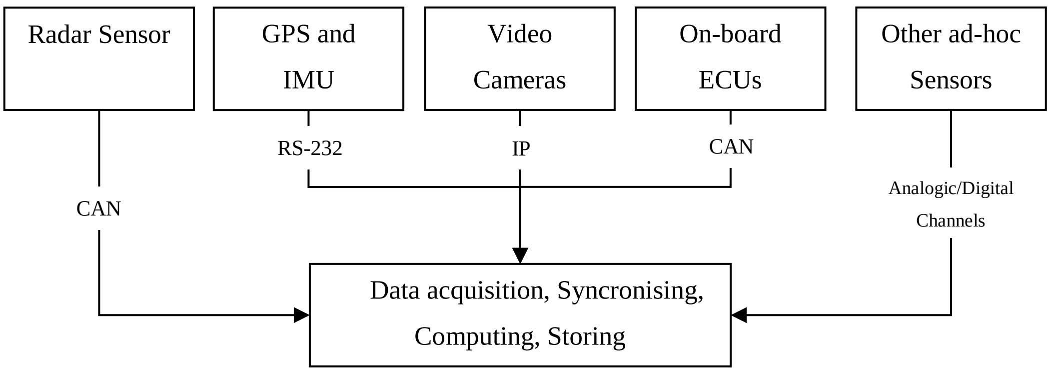

Fig. 3 General architecture of an instrumented vehicle

Different roles can be played by the various on-board sensors of an instrumented vehicle. With respect to the observation of car-following behaviours, GPS and inertial measurement units (IMUs) are mainly adopted for the so-called ego-data, i.e. the data associated to the instrumented vehicle (position, speed, acceleration, roll, pitch, yaw, etc.). Often, the GPS and the IMUs are integrated, as some multi-source algorithms can be conveniently used to merge the absolute position estimates given by the GPS and the dead-reckoning positioning obtained by IMUs (Ma and Andreasson, 2007; Bifulco et al., 2011). Radar devices are mainly adopted for detecting the relative speed and spacing;

they can be mounted in the front or at the rear of the equipped vehicle (or both). In the case of front radar the role of the following vehicle is played by the instrumented one. Hence the driver of the instrumented vehicle is the one whose behaviour is observed. In the case of rear radar the behaviour of other drivers is observed, as the instrumented vehicle plays the role of the leader. In the first case the instrumented vehicle is said to be active, while in the second case it is said to be passive. In active mode, analysts can have greater control over the sample of drivers recruited/involved; however, the drivers are aware of being part of an experiment (even if they do not know exactly which one) and the observation could be subject to bias. As regards cameras, these are adopted in our work only to monitor the road and traffic context and to qualitatively judge the efficiency and accuracy of the other sensors.

In this work we carried out three experimental surveys by using two different instrumented vehicles: one vehicle was equipped at the University of Naples Federico II (Italy) and the other at the University of Southampton (UK). The radars used were the TRW Autocruise AC10 (the AC20 release in the second Italian sample), which proved to be able to detect (and track) a target in a range from 2 to 150 metres at a speed from 40 to 200 km/h (from 10 to 200 km/h, the AC20 model). The Italian vehicle mounted both rear and front radar, while the UK vehicle mounted only rear radar. The radar data exploited in this paper are the inter-vehicular spacing and gap from the vehicle ahead (active mode, Italian experiments) and from the vehicle behind (passive mode, UK experiment). As regards the ego-data, the cruising speed of the instrumented vehicle is mainly exploited in this work. This was computed by validating the speed obtained from the GPS with that obtained by the on-board CAN and adopting the (filtered) GPS speed (sampled at 10 Hz). All data were collected at a 10 Hz frequency, synchronised and recorded on board. They were then processed off-line in order to smooth the signal and to avoid spikes and bias. The data were smoothed using a Kalman Filter, as described in Bifulco et al. (2011), where the state variables are the position, speed, acceleration of both the leader and follower. The measures are the relative speed and distance (from the radar), the speed, and the position of the instrumented vehicle. The represented dynamic system refers to the two vehicles (leader and follower) in instantaneous uniformly accelerated motion (with different accelerations applied at each time step of the process). The matrix of measurement errors was estimated by using known accuracies of the measure instruments. The matrix of the process error was fixed by tuning on the basis of previous experiments: some accurate reference trajectories (obtained by using Differential GPS) were available and discrepancies were estimated with respect to the hypothesis of uniformly accelerated motion. The great advantage of using the Kalman filter is that it allows consistent profiles of speed, accelerations, relative speed and spacing to be obtained.

4. Experimental surveys and collected data

Three independent field surveys were carried out: two in Italy, and the other in the UK. The datasets are identified as IT_1, UK and IT_2.

The IT_1 dataset was collected on the SS7 quater Domitiana (Bifulco et al., 2013a), a divided rural highway with two lanes in each direction and speed limit of 90 km/h, largely not respected by drivers due to low enforcement and medium-high road quality that enables higher speeds. Experiments were carried out in the morning, in calm weather conditions. In this experiment, the drivers were asked to follow a corporate vehicle; the driver of the corporate vehicle was an experimenter without any particular experimental protocol and required to drive in a natural/spontaneous way in the traffic stream. In this experiment voluntary drivers were recruited from students at the university.

For the UK dataset (Brackstone et al., 2002) experiments were performed during October on the M3 three-lane motorway (70 mph speed limit, equivalent to about 110 km/h) between junctions 2 and 4a (a total of 22.2 km), during the morning peak between 7:30 and 8:30 AM. Data were collected using the passive mode, the radar was fitted facing rearward and observation made of following drivers selected at run-time in the traffic stream. Gender and age of the monitored drivers were inferred by video in this case.

The IT_2 dataset was collected within the Italian research project DRIVEIN2 (Bifulco et al., 2012; Bifulco et al., 2014) on the Motorways A1 and A30, which are divided highways with three lanes in each direction, and speed limits in the study sections of 100 km/h on the A1 and 130 km/h on the A30. In the IT_2 experiment, the drivers were asked to adopt different experimental protocols: on the A1, they were asked to drive spontaneously; on the A30, they were asked to drive spontaneously in the first part of the journey and to follow a corporate vehicle in the second part. The experimenter of the corporate vehicle applied, in turn, a precise experimental protocol: (a) he drove at 80 km/h for 2 km, then (b) he sped up to 100 km/h for 3 km, and finally (c) he sped up to 120 km/h for 3 km. Participants in the experiments were recruited in order to reproduce in our sample the main characteristics (in terms of gender, age, etc.) of the Italian population of drivers.

TABLE 1 Main characteristics of the Italian and UK datasets

Dataset	Drivers			Driving sessions in car-following conditions						
	N	Female (%)	Age < 26 (%)	Number of trajectories	Total duration (min)	Total length (km)	Time (s) Mean	Max	Min	Average speed (km/h)
IT_1	13	46	85	13	110	88	527.5	1220	234.0	58
UK	31	20	80	54	82	134	178.7	562	30.6	66
IT_2	92	43	27	759	441	576	54.4	400	37.2	92

Table 1 above summarises the main characteristics of all the IT and UK datasets. In all cases the traffic conditions may be described as medium congestion, without stop-and-go phenomena.

In all the experiments data related to different drivers were collected as different driving sessions. These were split into trajectories, each characterized by a unique leading vehicle and by uninterrupted car-following (CF) conditions. As a result, non-CF conditions were discarded, and 826 trajectories were obtained. Given that the instrumented vehicles adopted in all experiments were similarly equipped and the experimental procedures were similar, in all trajectories the same variables were logged, at a 10 Hz frequency. Among these we employed for our analyses:

- the speed of the follower;
- the relative speed between the leader and the follower;
- the relative spacing between the leader and the follower.

5. Identification of car-following equilibria

As stated in section 2, the approach followed here is twofold. First the equilibrium conditions of observed car-following are directly analysed. Then the observed driving behaviour are indirectly characterised via identification of the modelling parameters for three different car-following paradigms (each applied to all of the three different samples).

The first step to carry out is to identify equilibrium conditions in observed car-following trajectories. The purpose is to ascertain whether the observed car-following situation in each trajectory is homogeneous, and to assign to each of them both a reference speed, and an equilibrium/desired spacing. Searching for homogeneity is very important in our case. Indeed, as shown above, each dataset comprises several trajectories and car-following conditions can be observed to change both across trajectories and within a given trajectory. In the latter case the observed car-following episodes could be representative of more than one equilibrium condition. An example is depicted in Figure 4, which is a very particular case taken from the UK dataset, where the car-following process occurs at three different speeds.

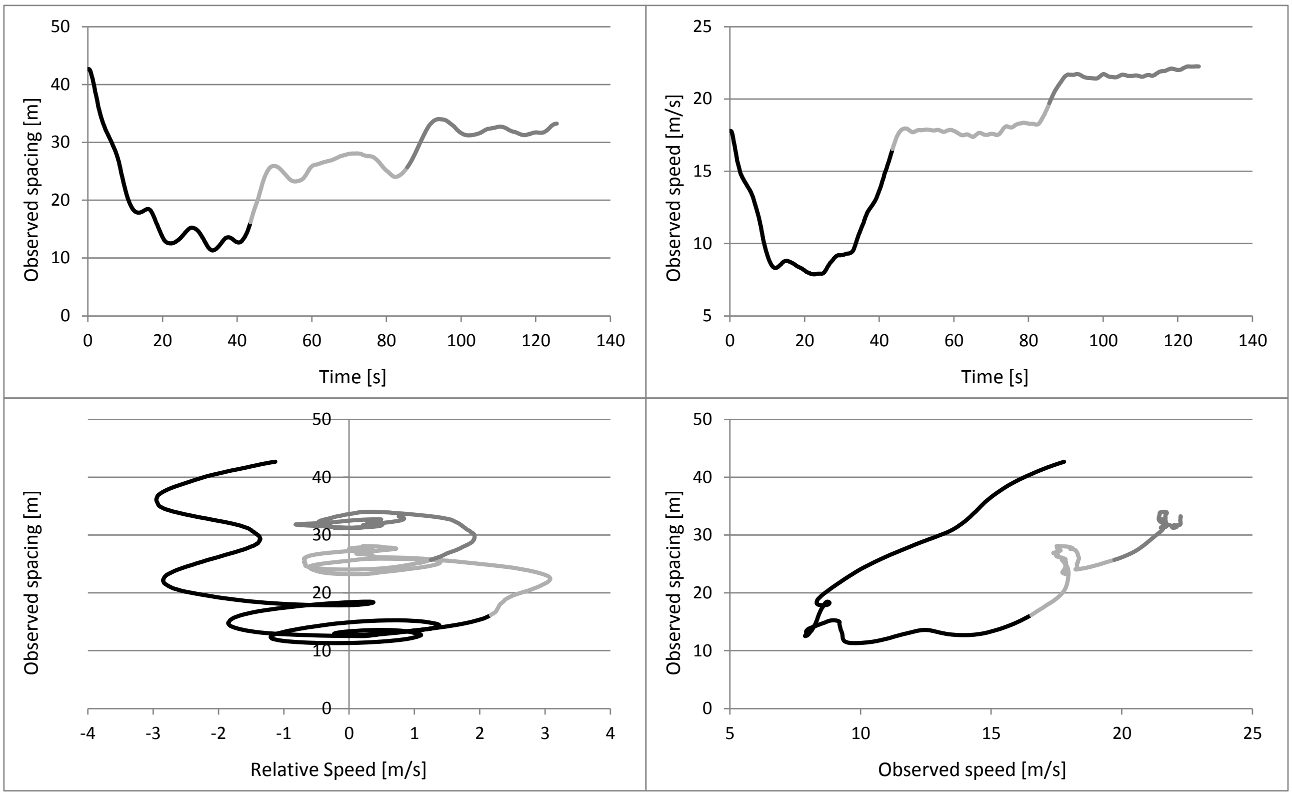

Fig. 4 An example of cluster analysis applied on one trajectory of the UK dataset

In the first 40 seconds the following vehicle approaches a slower leader, until it reaches a speed at about 9 m/s and the observed spacing is around 15 m. The leading vehicle then increases its speed, and a new equilibrium condition is reached, with about 28 m of spacing. In the last part of the trajectory the speed reaches about 22 m/s, and an equilibrium condition is reached for the third time at a spacing value of 32 m. As a consequence, this trajectory has to be split into three subsets and one point representing the equilibrium condition reached has to be assigned to each. These subsets are here called clips. Of course, if the trajectory is homogeneous throughout its duration, the clip coincides with the trajectory itself.

Importantly, once the phenomenon is analysed in the relative speed-spacing and in the follower's speed-spacing planes, the points gather in specific zones of the two planes, one for each equilibrium condition. This circumstance is of great help when equilibrium conditions have to be searched. This can be observed in the second row of Figure 4, where the three equilibrium conditions described above can be clearly identified.

Given the huge quantity of data treated in the paper the procedure for verifying homogeneity and, if need be, for clip generation was based on a machine learning approach. The machine learning approach ensures the possibility of automating the process and establishing an objective-function and systematic criterion for clip generation.

In particular, for each trajectory a vector is defined using three variables: spacing, the follower's speed and the time elapsed; each vector has n components, corresponding to the single observation records in the trajectory. The vector is then analysed using a k-means clustering algorithm.

K-means clustering aims to divide n observations into k clusters in which each observation belongs to the cluster with the nearest mean, which represents the centroid of the cluster (Galante et al., 2010; Montella et al., 2010; Montella et al., 2011). Given a set of observations (x_1, x_2, …, x_n), where each observation is a d-dimensional real vector, k-means clustering aims to partition them into k ($\leq$ n) sets S = {S_1, S_2, …, S_k} so as to minimise the within-cluster sum of squares (WCSS). In other words, the objective function is:

$$\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2$$

where μ_i is the mean of points in S_i.

The problem here is that the number of clusters k has to be defined a priori. Thus a criterion is required in order to verify that k has been adequately chosen. In our case the gap statistic (Tibshirani et al., 2001) was used to identify k. This technique uses the output of any clustering algorithm, comparing the change in within-cluster dispersion with that expected under an appropriate reference null distribution, and can be used in an automatic process as follows:

- cluster the observed data, varying the number of clusters from 1 to n_c, which is the predefined max number of accepted clusters;
- compute the gap statistic;
- choose the k corresponding to the maximum value of the gap statistic.

In our procedure the maximum number of accepted clusters n_c was set equal to 4, given that the procedure is computationally demanding, and on the basis of previous experience n_c was < 4 in 92% of the cases in our datasets.

In summary for each trajectory:

- the optimal number of clusters was chosen using the gap statistic;
- k clips were created using the k-means algorithm;
- the k centroids were considered as representative of the equilibrium condition in the clip.

It is worth noting that the follower's speed and spacing were considered for the clustering procedure because the two variables represent the main interest of our study. The elapsed time was also added in order to ensure that the points selected for each cluster refer to close instants. The outcome of the procedure described hereinis reported in Figure 4 above by using different grey tones.

The clustering results were also checked with respect to the equilibrium spacing computed by using the procedure described in Pariota (2013). Pariota demonstrated that a good estimation of the equilibrium spacing can be carried out once relative speed action points are selected in a clip and a linear regression is made using the relative speed as independent variable, and the spacing as dependent; the intercept term of the regression line is the equilibrium spacing. The two procedures were compared, and ascertained to be equivalent. However, in this work the machine learning approach is preferred, as it automatically allows a reference speed to be associated to the selected equilibrium spacings.

6. Presentation and discussion of the results

Direct analysis of the identified equilibrium points is first carried out. Based on this analysis two conjectures are discussed: i) the adopted equilibrium spacing increases with speed; ii) the across-drivers dispersion of the observed behaviour increases with speed.

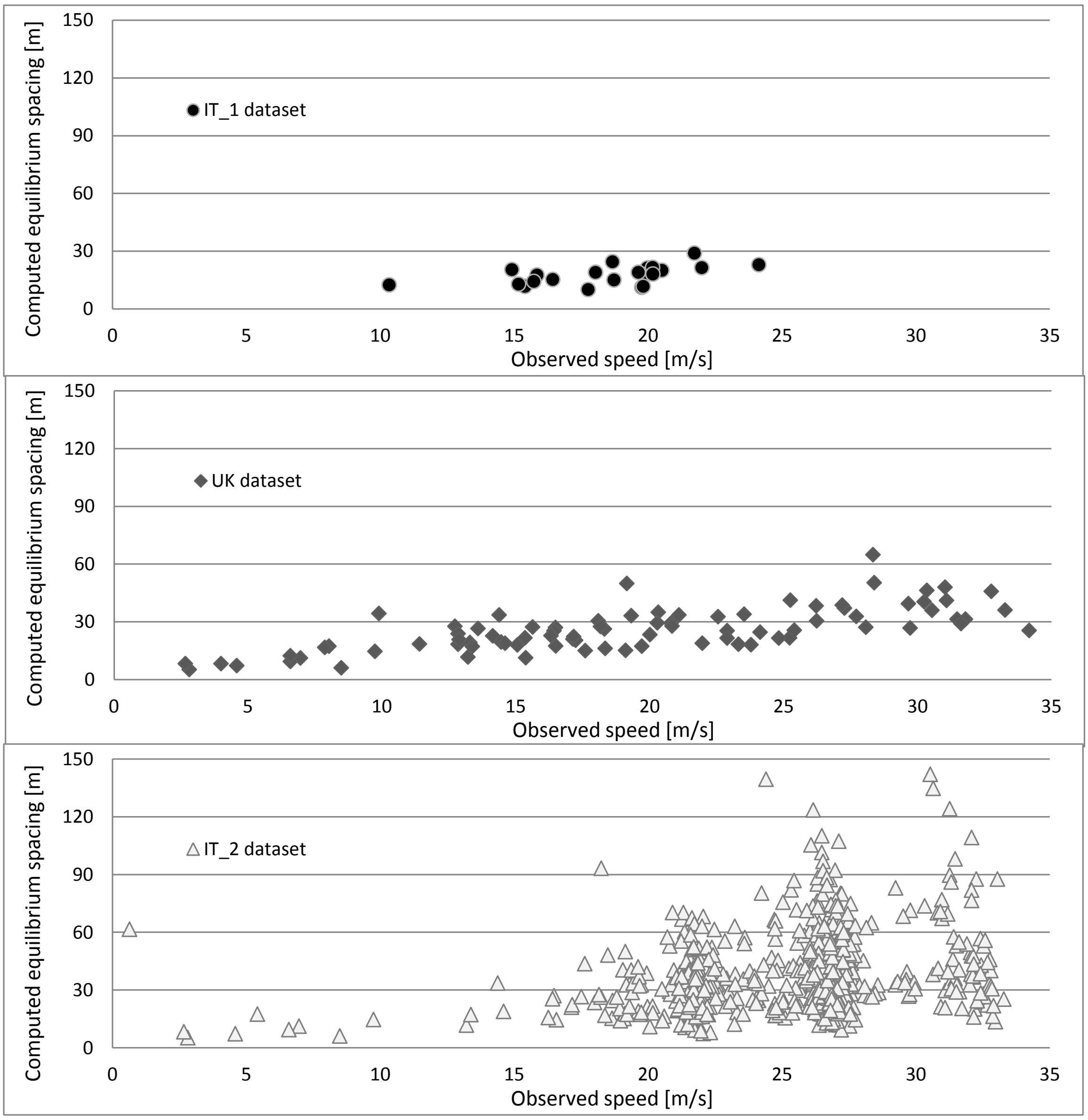

Fig. 5 The scatter plot of the equilibrium conditions evaluated in the three datasets: spacing vs. follower's speed

Figure 5 presents a scatter plot of the identified equilibrium conditions in the follower's speed-spacing plane. The different size of the three datasets is evident, and the dispersion of data seems to be different too, and influenced by the number of points. In particular, equilibrium conditions identified in the three datasets numbered 22 for IT_1, 110 for the UK data set and 1874 for IT_2. Another point of difference concerns the range of speed covered in the three datasets: the two Italian datasets show most of the points in more limited speed ranges (10-25 m/s for IT_1, 15-35 m/s for IT_2), while the UK experiment covers a greater speed range. However, the common point in the three datasets is the growing trend of the equilibrium spacing with respect to the speed. The average values increase with speed in all the datasets (albeit assuming different values). This result is intuitive, and also confirms references from the literature. Indeed, considering studies carried out in different countries some analogies can be found with respect to our collected data. In particular, Fig. 6 compares the results of this work with the findings of Parker (1996) obtained in England, those of Huddart and Lafont (1990) concerning French drivers, and the indications of the California Code reported by Chandler et al. (1958).

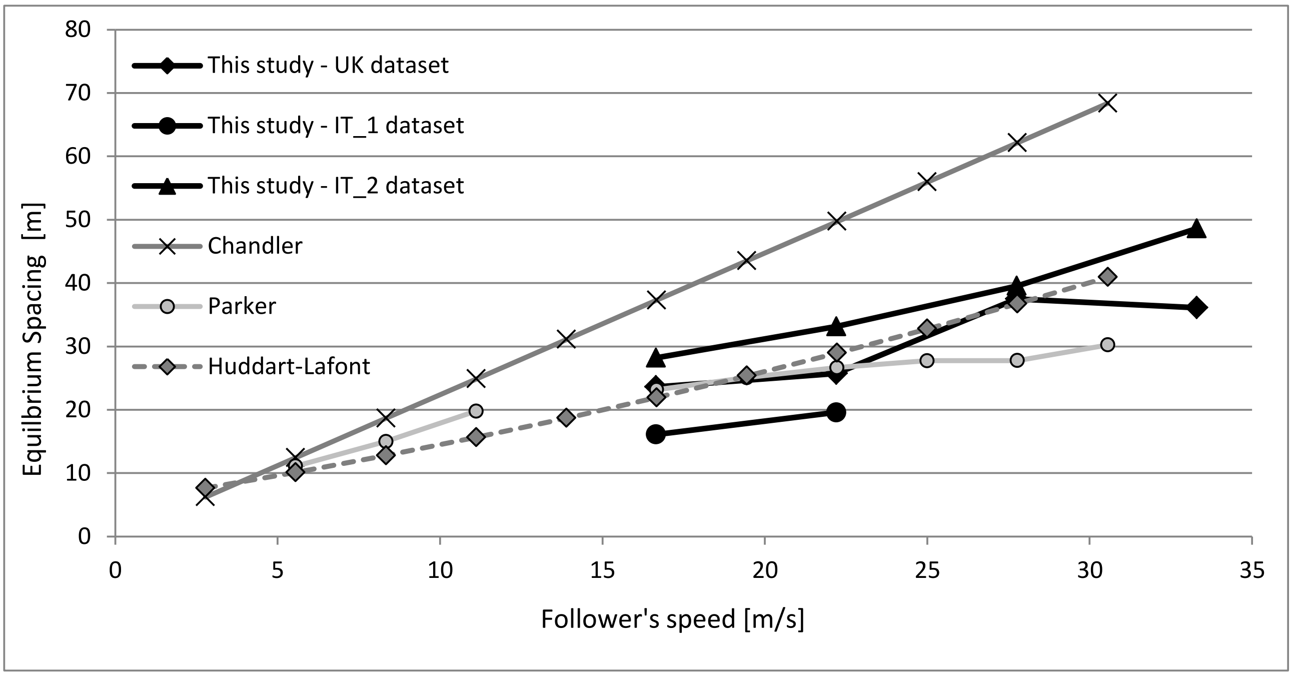

Fig. 6 Distribution of the equilibrium spacing computed in different datasets with respect to the follower's speed

Note that our data are in accordance with other observer data such as those of Huddart-Lafont and Parker. The two Italian datasets are aligned on different sides with respect to the UK dataset. This does not appear to highlight a systematic effect of drivers' nationality nor of the time the data were collected (the UK data are much less recent than the Italian data). It can be noted that the larger the observed dataset (from IT_1 to UK and then to IT_2), the greater the average spacing (for any given follower speed). This tendency of the data to be dependent on sample size is discussed below.

Other more formal analyses can be carried out to identify the dispersion of the data observed in the three experiments. Our analyses start from the IT_2 dataset. Given the particular experimental conditions, the detected points were grouped into four ranges of speed, equally spaced, from 16.65±2.78 to 33.30±2.78 m/s (which corresponds to covering the interval from 50 to 130 km/h with one speed class every 20 km/h). This choice was influenced by the particular experimental conditions in the IT_2 dataset, where during much of the experiment each driver was committed to following a corporate vehicle at 80, 100 and 120 km/h. Points in each of the classes were used to build an empirical probability density function (PDF). The empirical distributions of the four classes are plotted in Figure 7. Once selected, points in each class were used to fit a lognormal distribution. The fitted distributions, as well as the relative Q-Q plots, are also plotted in Fig. 7. The Q-Q plot is a plot of the percentiles of a standard log-normal distribution against the corresponding percentiles of the observed data. It shows the expected log-normal value on the X axis plotted against the observations on the Y axis. If the data are log-normally distributed, the data points are close to the diagonal. If the data points stray from the line in an obvious non-linear fashion, the data are not log-normally distributed (Montella et al., 2015a, 2015b).

Parameters of the four distributions and their statistics (mean, standard deviation and coefficient of variation) are given in Table 2. The same analyses were also repeated for the other two datasets. For comparison, they are limited to the four speed classes used for the analysis of the IT_2 dataset; all the results are reported in Table 2 as well. Also in these cases the lognormal distributions fit the observed data well.

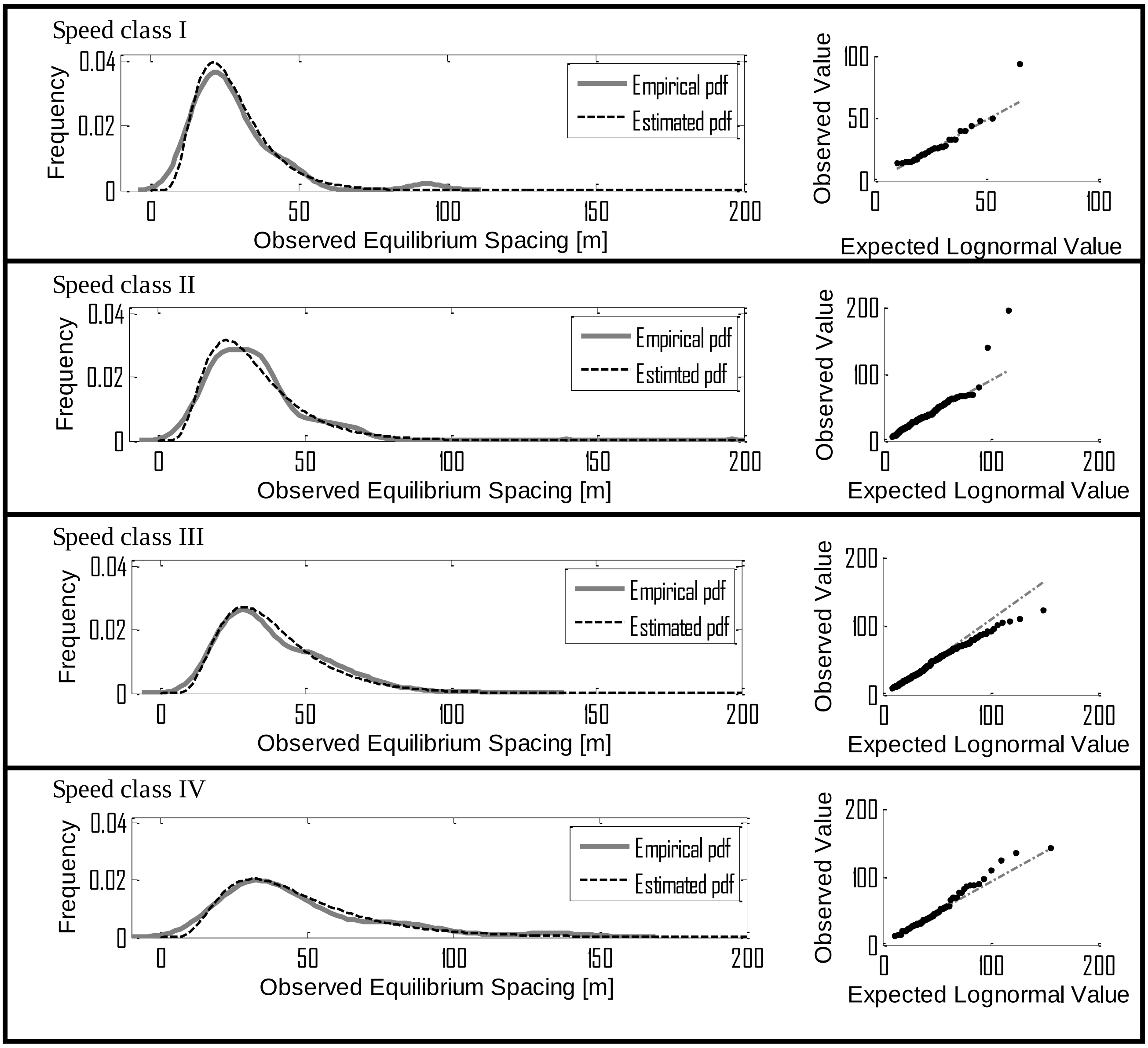

Fig. 7 Estimated lognormal distribution for the four speed classes in the IT_2 dataset, compared with the empirical distribution

TABLE 2 Parameters of the lognormal distributions estimated in the three datasets and their statistics

Dataset	Speed class	Lognormal spacing distribution parameters				
		m	s	Mean	SD	CV (SD/Mean)
IT_1*	I: 50-70 km/h	2.74	0.27	16.12	4.437	0.275
	II: 70-90 km/h	2.94	0.28	19.58	5.546	0.284
	III: 90-110 km/h	-	-	-	-	-
	IV: 110-130 km/h	-	-	-	-	-
UK	I: 50-70 km/h	3.11	0.32	23.63	7.787	0.330
	II: 70-90 km/h	3.22	0.24	25.74	6.373	0.248
	III: 90-110 km/h	3.58	0.29	37.50	10.95	0.292
	IV: 110-130 km/h	3.56	0.21	36.13	7.740	0.214
IT_2	I: 50-70 km/h	3.25	0.43	28.18	12.79	0.454
	II: 70-90 km/h	3.38	0.48	33.12	16.94	0.512
	III: 90-110 km/h	3.57	0.46	39.52	19.14	0.484
	IV: 110-130 km/h	3.74	0.53	48.60	27.94	0.575

*The IT_1 dataset does not present points in speed classes III and IV

The hypotheses on the significance of distributions were also tested by using the Kolmogorov-Smirnov and the Chi-square non-parametric tests. All data significantly fit the lognormal distribution, with all values of the tests less than

0.05. Handy to note, the values of the estimated parameters of the spacing distribution differ in the three datasets. Interestingly, the variances of the equilibrium spacing within each class increase with speed; this is fully true for IT_1 (even if only two classes are evaluated) and especially for IT_2, while it is controversial for the UK sample. Note that the values of the standard deviations of observed equilibrium spacing significantly increase according to the size of the sample. This is a further indication that some differences in the observed behaviours can also be explained by the greater variability (and number) of observed driving conditions and behaviours.

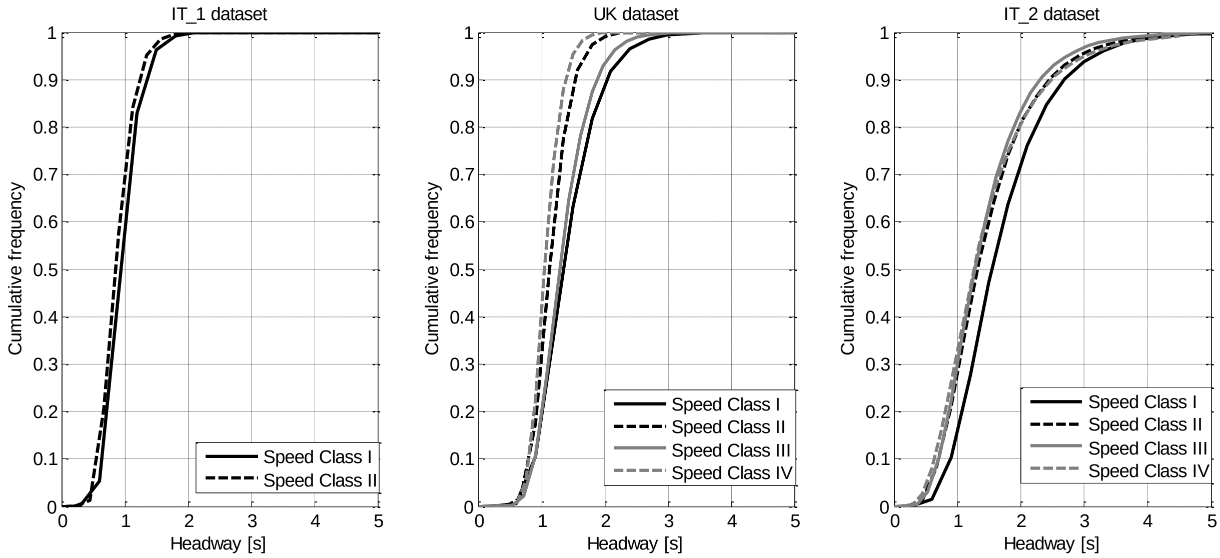

Fig. 8 Estimated cumulative distribution of the observed headway for all the speed classes and all the datasets

Considering the headway cumulative distributions depicted in Figure 8, it is possible to carry out an evaluation of the observed safety conditions. It is worth noting that the cumulative distribution of the observed headways are obtained as a linear transformation of those estimated for the spacing (each of the spacing distribution parameterized in Table 2 has been scaled using the corresponding speed class value). Once an headway threshold is chosen, the corresponding percentile in the cumulative distribution can be computed. Of course, the evaluation can be made with respect to different thresholds. As an example a first threshold of 2 seconds can be fixed according to the perception-reaction time given in the geometric design standards from Australasia (Austroads, 2010) and France (SETRA, 2001). It is worthy to note that in North America the perception-reaction time is assumed 2.5 seconds (AASHTO, 2011; TAC, 1999). The probability of observing a headway greater than the 2 seconds threshold is very low in the IT_1 and UK datasets. It is only slightly greater in the IT_2 dataset, with probabilities ranging from 0.17 to 0.28 for the four speed classes. For an insight on the phenomenon, a second threshold of 1 second can be fixed according to Vogel (2003). In this case, the probability of observing headways greater than the threshold ranges from 0.28 to 0.39 in the IT_1 dataset, and from 0.60 to 0.84 in the other two datasets. However, this means that a high frequency of headways lower than the second threshold fixed can be still observed in all the samples. These data clearly show that several following too close behaviours are observed in all the three samples. As a consequence, there is a greater potential for reduction of rear-end crashes by introducing proper driving automation solutions aimed at longitudinal control of vehicles.

A difference in the values of estimated parameters could be reasonably explained by different behaviours being actually observed in the three datasets. However, in order to better discuss this result, note that elaborations in Table 3 show some interesting analyses about the statistical similarity of the different samples of drivers. Indeed, for the points in the UK and IT_1 datasets the null hypothesis is tested that they can be considered samples of the distributions estimated (for the four speed classes) using points of the IT_2 dataset. The findings may be considered quite controversial. The null hypothesis is always rejected (p-values less than 0.05) for the IT_1 dataset, so that we can be more confident on the hypothesis that the two datasets actually present different behaviours. Differently, the hypothesis is not rejected (except for the Kolmogorov-Smirnov test in the second speed class) in the UK dataset. In other words, if the distribution that results from IT_2 is sampled, then the UK dataset could be obtained. Put simply, there is more dispersion within the IT_2 dataset than across the drivers of the UK and the IT_2 datasets.

The previous issue is interesting also with reference to the adoption of datasets collected in different periods. Indeed UK experiments were performed about 10 years before those carried out in Italy, but no significant difference was found between the UK and IT_2 datasets, in contrast with differences emerging between IT_1 and IT_2 data that were both collected in Italy, with the same vehicle, and more recently (refer to figure 6, and Tables 2 and 3). This result reassured us about going ahead with the use of all three datasets also for the indirect comparisons carried out in the rest of the paper.

TABLE 3 Testing the hypothesis that datasets IT_1 and UK can be considered random draws from the distribution estimated for dataset IT_2

Dataset	Speed class	Kolmogorov-Smirnov	Chi-square
		p-value	p-value
IT_1*	I: 50-70 km/h	**<0.01**	**<0.01**
	II: 70-90 km/h	**<0.01**	**<0.01**
	III: 90-110 km/h	-	-
	IV: 110-130 km/h	-	-
UK	I: 50-70 km/h	0.07	0.13
	II: 70-90 km/h	**0.02**	0.06
	III: 90-110 km/h	0.69	0.55
	IV: 110-130 km/h	0.08	0.08

*The IT_1 dataset does not present points in speed classes III and IV

TABLE 4 Parameter distributions in the three datasets

Model	Parameter	Dataset								
		IT_1			UK			IT_2		
		25th	50th	75th	25th	50th	75th	25th	50th	75th
State-space	ω_1	0.011	**0.044**	0.147	0.015	**0.050**	0.101	0.019	**0.058**	0.199
	ω_2	0.332	**0.641**	0.942	0.514	**0.709**	1.185	0.222	**0.575**	1.507
Stimulus-response	β_0	0.031	**0.082**	0.167	-0.041	**0.238**	0.985	-0.952	**0.938**	5.712
	β_1	-0.007	**-0.004**	-0.002	-0.052	**-0.025**	-0.012	-0.026	**-0.008**	-0.003
	β_2	0.572	**0.586**	0.591	0.368	**0.446**	0.496	0.551	**0.594**	0.621
	β_3	-0.004	**0.002**	0.005	-0.011	**0.012**	0.040	-0.179	**-0.015**	0.065
Waves	a_1	-0.0008	**0.0002**	0.0013	-0.0003	**0.0004**	0.0020	-0.0006	**-0.0005**	0.0003
	a_2	-0.076	**-0.055**	-0.048	-0.054	**-0.038**	-0.029	-0.037	**-0.027**	-0.019

Indeed, indirect analyses were also carried out. The three datasets were compared by using the interpretive models introduced in Section 2. The result of the identification of the modelling parameters is reported in Table 4.

Identification of the state-space model was carried out using the Matlab-System Identification Toolbox; for the calibration of the parameter of the other two models the Matlab Statistical Toolbox was used. With respect to the stimulus-response model and that based on waves, good fitting of the data was obtained, as ensured by the computed r-square statistics; they are not reported in Table 4 because for all the clips the values assumed were between 0.85 and 1, without any significant difference between the datasets. For each of the modelling parameters the respective 25th, 50th, and 75th percentile of the values assumed in the dataset are shown (median values are in bold); parameters in the three datasets are fully comparable. With reference to the state-space model, positive values confirm the rational driving behaviour hypothesis. In order to maintain safe car-following conditions drivers should increase their acceleration when there is an increase in spacing (the actual spacing becomes greater than the desired one), and decrease it in the opposite situation. The same should happen with respect to relative speed (Wilson, 2008): this happens when $\omega_1, \omega_2 \geq 0$. In the stimulus-response model, parameters β_1 and β_3 assume quite low values. Hence the main stimulus is that associated to β_2, the relative speed. The wave-based model shows the main differences across the three datasets: on the one hand, a_1 is negligible (as expected, the value of a_1 should be zero) for all three cases; on the other, the value of a_2 decreases (hence the slope of the trend curve) from IT_1 to IT_2. If considered in the quadrant with negative DV, the slope a_2 (see also Figure 2 in section 2) interpolates points where an action is performed by the driver in correspondence to a given relative speed and a corresponding (inverse of) time-to-collision. Given the relative speed, the (inverse) time-to-collision is strictly related to the relative spacing at which the action is made. In conclusion, smaller slopes correspond to larger relative spacing for any given relative speed, that is more precautionary (less risk-prone) behaviours. Thus, values a_2 in Table 4 suggest that the IT_2 dataset is, overall, more risk-averse than the UK dataset and, in turn, the UK dataset is more risk-averse than the IT_1 dataset; this result confirm safety considerations drawn from the analysis on the observed headway described for Figure 8. The difference in the observed behaviours can be interpreted in two ways: the behaviours could be considered different because, say, different driving styles may be due to some systematic difference in the datasets; or, as has been already noted, the observed datasets differ in terms of dispersion of observed behaviours, and the larger the sample, the greater is the probability of observing uncommon behaviours. Viewed from the latter standpoint, the decreasing a_2 slope indicates that uncommon behaviours could be more risk-averse than common ones and that the former appear more probably in larger datasets. To confirm the different structure of the dispersion as the size of the dataset increases, note that in Table 4 the inter-quartile range (IQR) almost always increases from the IT_1 to the IT_2 dataset. It should also be noted that in the IT_2 dataset drivers were recruited in order to reproduce the main characteristics (in terms of gender, age, etc.) of the Italian population of drivers. In summary, larger (and more representative) samples allow observation of a wider range of behaviours which adds non-negligible information to the collected datasets. This, in our opinion, is an expected (but interesting) finding that emerges from our analyses.

7. Conclusions and future research directions

The results presented in section 6 refer to direct observations of driving behaviour in three different experimental datasets, as well as indirect comparisons allowed by the identification of modelling parameters able to describe the observations. Data were obtained thanks to the development of a machine learning approach, able to identify and select equilibrium conditions in automatically collected car-following data. Albeit employed here in an off-line approach, once implemented in real time the method is a useful tool for on-board identification of equilibrium behaviours desired/applied by drivers. Thus the analyses here applied off-line can be further developed to be applied on-line. In this way off-line characterisation of driving behaviours and on-line testing of actually observed behaviour can be combined to set up customised/localised driving assistance solutions, oriented to active safety systems. For instance, the on-line observed driving behaviour, once analysed in terms of equilibrium spacing actually applied by the driver, can be compared with the off-line estimated statistical distributions; if the actual spacing is shorter than (say) the tenth-percentile of the observed distribution, the driver could be warned for applying too aggressive (and thus dangerous) driving behaviour. This could be a practical enhancement of traffic safety enabled by our study, given the impact of poor headways on rear-end crashes, as discussed in Austroads (2015). The implementation of this solution will be addressed in future works.

Our results herein concern off-line observations and the characterisation and comparison of different datasets, collected in different experiments in different years and countries. The analyses show that the lognormal distributions fit the observed data well, and confirm the findings of previous studies. Greenberg (1966) in his early paper introduced the hypothesis of lognormal distribution of the headway. This hypothesis was confirmed several times in subsequent years (e.g., Piao et al., 2003; Li et al., 2010) up to the recent paper of Jiang and Lu (2015).

It is also evident from our analyses that the variances of the equilibrium spacing within each class increase with speed, showing that drivers' behaviours are less dispersed at lower speeds. This was also confirmed by Jiang and Lu (2015) with respect to Chinese drivers. The main reason we would suggest for such a result is that at higher speeds drivers' aggressiveness tends to produce greater effects: the more aggressive a driver is at higher speeds, the closer the spacing he/she adopts with respect to non-aggressive drivers. This effect is less evident at lower speeds, where aggressive and non-aggressive drivers tend to adopt more similar spacing (Elefteriadou, 2014). As regards the comparison across the three different datasets, the average influence of speed on equilibrium conditions is similar (piecewise linear) and in line with previous studies (Figure 6): it all falls in quite a narrow band (except for the curve obtained by Chandler which seems very conservative).

The equilibrium spacing results to be differently dispersed in the three datasets with respect to both the estimated lognormal distributions and to the inter-quartile range values of the estimated parameters of the considered models. This result is correlated to the different size of the three datasets: increasing the number of drivers observed (as well as the duration of the driving sessions) increases the probability of observing different (and even uncommon) behaviours. Moreover, as discussed at the end of section 6, the values of the a_2 slope decreases if the size of the sample increases, and this indicates that uncommon behaviours are more risk-averse than common ones. Put simply, studies based on a relatively small (but statistically significant) sample can be useful for the observation of the average phenomenon, while an increased sample allows in-depth understanding, especially in the case of car-following behaviour, which is confirmed to be a complex phenomenon.

Indirect comparison of observed behaviours, based on behavioural models, shows the rationality of the parameters obtained. Moreover, similar average distributions of the parameters are observed across the three datasets. Hence reactions to stimuli are similar in the different (observed) groups of drivers.

On the other hand it should be considered that although the three different field experiments have been carried out in similar contexts and speed ranges, they exhibit a variability in both the detected equilibrium conditions and the parameters of the models. The results are probably influenced by the differences in the sample compositions in terms of gender, age and other personal characteristics of the drivers. That is, important lower-order factors may have been neglected due to the very different size of the datasets. Also within-driver variability (Wagner, 2012) could represent another significant source of randomness; the latter aspects are even more difficult to address, considering how difficult it is to carry out larger naturalistic observations of driving behaviour. The analysis of within-driver variability and of the impact of composition in terms of age and gender is the main goal the authors set for themselves as a further research perspective. These aspects are of ever greater importance, as it is increasingly clear that not only do they affect models aimed at (microscopic) traffic simulation, but strongly impact in this case of driving assistance/automation solutions. Indeed, the desired driving behaviour (with which the automation system has to interact) can be observed to vary for a given driver, depending on the context, purpose and duration of the trip or on other parameters specific to each trip, which are often very difficult to fully understand. The availability of a large amount of vehicle trajectory data (collected at the road-side or on board), such as what we have, can be successfully treated with appropriate algorithms (e.g., Taylor et al., 2015) in order to investigate within-driver variability at a future stage.

Acknowledgements

The work reported in this paper was partially supported by the App4Safety research project (B61H110004000005/PON01_00744) under a grant from the Italian Ministry of Education (MIUR). Italian data used in the work were collected during the research project DRIVEIN2 (B61H110004000005/PON01_00744) also under grant from the MIUR. United Kingdom data were supplied by the University of Southampton's Transportation Research Group (TRG). The views expressed in this paper should not be taken as necessarily representing those of TSS Ltd.

REFERENCES

AASHTO, 2010. Highway Safety Manual, 1st ed., Washington D.C.

AASHTO, 2011. A Policy on Geometric Design of Highways and Streets, 6th ed., Washington, D.C.

Alexiadis V, Colyar J, Halkias J, Hranac R, McHale G (2004) The next generation simulation program. ITE Journal, 74(8): 22-26.

Austroads, 2010. Guide to Road Design Part 3: Geometric Design. Austroads Publication AGRD03-10, Sydney, New South Wales.

Austroads, 2015. Investigation of Key Crash Types: Rear-end Crashes in Urban and Rural Environments. Austroads Publication AP-R480-15, Sydney, New South Wales.

Bifulco G N, Pariota L, Simonelli F, Di Pace R (2011) Real-time smoothing of car-following data through sensor-fusion Techniques. Procedia Social and Behavioural Sciences, 20: 524–535.

Bifulco GN, Galante F, Pariota L, Russo Spena M, Del Gais P (2014) Data collection for traffic and drivers' behaviour studies: A large-scale survey. Procedia-Social and Behavioural Sciences, 111: 721-730.

Bifulco GN, Pariota L, Brackstone M, McDonald M (2013b) Driving behaviour models enabling the simulation of Advanced Driving Assistance Systems: revisiting the Action Point paradigm. Transportation Research Part C: Emerging Technologies, 36: 352-366.

Bifulco GN, Pariota L, Galante F, Fiorentino A (2012) Coupling instrumented vehicles and driving simulators: opportunities from the DRIVE IN2 project. IEEE Conference on Intelligent Transportation Systems (ITSC2012): 1815-1820, IEEE press. IEEE Catalogue number CFP12ITS-USB. ISBN: 978-1-4673-3026-6

Bifulco GN, Pariota L, Simonelli F, Di Pace R (2013a) Development and testing of a fully Adaptive Cruise Control system. Transportation Research Part C: Emerging Technologies, 29: 156-170. doi: 10.1016/j.trc.2011.07.001

Bifulco GN, Simonelli F, Di Pace R (2008) Experiments toward a human-like Adaptive Cruise Control. 2008 IEEE Intelligent Vehicles Symposium: 919-924.

Boyce TE, Geller SE (2002) An instrumented vehicle assessment of problem behaviour and driving style: do younger males really take more risk? Accident Analysis & Prevention, 34(1): 51-64.

Brackstone M, Sultan B, McDonald M (2002) Motorway driver behaviour: studies on car following. Transportation Research Part F: Traffic Psychology and Behaviour, 5: 31-46.

Brackstone M, Waterson B, McDonald M (2009) Determinants of following headway in congested traffic. Transportation Research Part F: Traffic Psychology and Behaviour, 12: 131-142.

Brockfeld E, Kühne RD, Wagner P (2004) Calibration and validation of microscopic traffic flow models. Compendium of Papers of the 83rd Annual Meeting, Transportation Research Board, Washington.

Department for Transport, 2012. Reported Road Casualties, Great Britain 2011. London.

Department for Transport, 2013. Reported Road Casualties, Great Britain 2012. London.

Department for Transport, 2014. Reported Road Casualties, Great Britain 2013. London.

Elefteriadou L (2014) An Introduction to Traffic Flow Theory. Springer Optimization and Its Applications, 84. Springer, Berlin.

ERTRAC (2015) Automated Driving Roadmap: version 3.0 16/Feb/2015.

Galante F, Mauriello F, Montella A, Pernetti M, Aria M, D'Ambrosio A (2010) Traffic calming along rural highways crossing small urban communities: Driving simulator experiment. Accident Analysis and Prevention, 42(6): 1585-1594. doi: 10.1016/j.aap.2010.03.017

Gazis R, Herman R, Rothery RW (1961). Nonlinear follow the leader models of traffic flow. Operations Research, 9: 545-567.

Gipps PG (1981) A behavioural car-following model for computer simulation. Transportation Research Part B: Methodological, 15: 105-111.

Gouy M, Wiedemann K, Stevens A, Brunett G, Reed N (2014) Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers' longitudinal control? Transportation Research Part F: Traffic Psychology and Behaviour, 27(B): 264-273. http://dx.doi.org/10.1016/j.trf.2014.03.003

Greenberg I (1966) The log normal distribution of headways. Australian Road Research, 2(7).

Gurusinghe GS, Nakatsuji T, Azuta Y, Ranjitkar P, Tanaboriboon Y (2002) Multiple car-following data with real-time kinematic Global Positioning System. Transportation Research Record: Journal of the Transportation Research Board, 1802: 166-180.

Hamdar SH, Mahmassani HS (2008) Driver car-following behaviour: From discrete event process to continuous set of episodes. Compendium of Papers of the 87th Annual Meeting, Transportation Research Board, Washington .

Horrey WJ, Lesch MF, Mitsopoulos-Rubens E, Lee JD (2015) Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety, Accident Analysis & Prevention, 76: 25-33.

Huddart KW, LaFont R (1990) Close driving-hazard or necessity? Proceedings of the 18th PTRC meeting: 205–217, Seminar G-P334

Hutchinson, T.P., 2008. Tailgating. Report CASR046, Centre for Automotive Safety Research, Adelaide, South Australia. Jiang J, Lu J (2015) Car-following behaviour of Chinese drivers. Compendium of Papers of the 94th Annual Meeting, Transportation Research Board, Washington.

Johansson E, Engström J, Cherri C, Nodari E, Toffetti A, Schindhelm M, Gelau C (2004) Review of existing techniques and metrics for IVIS and ADAS assessment. AIDE Deliverable 2.2.1.

Kim T, Lovell DJ, Park Y (2007) Empirical analysis of underlying mechanisms and variability in car-following behavior. Transportation Research Record: Journal of the Transportation Research Board, 1999: 170-179.

Kometani E, Sasaki T (1958) On the stability of traffic flow. Journal of Operations Research Japan, 2: 11-26.

Lee JD, See KA (2004) Trust in automation: Designing for appropriate reliance. Human Factors, 46 (1): 50-80.

Li L, Fa W, Rui J, Jian-Ming H, Yan J (2010) A new car-following model yielding log-normal type headways distributions. Chinese Physics B, 19(2).

Lunt G, Day M, Wilson RE (2006) Enhancing motorway traffic data with novel vehicle re-identification algorithms. Proceedings of the 13th ITS World Congress. London (UK).

Ma X, Andreasson I (2007) Behaviour measurement, analysis, and regime classification in car following. IEEE Transactions on Intelligent Transportation Systems, 8(1): 144-156.

Michaels RM (1963) Perceptual factors in car following. Proceedings of the Second International Symposium on the Theory of Road Traffic Flow: 44-59, Paris.

Montella, A., Aria, M., D'Ambrosio, A., Galante, F., Mauriello, F., Pernetti, M., 2010. Perceptual measures to influence operating speeds and reduce crashes at rural intersections: Driving simulator experiment. Transportation Research Record: Journal of the Transportation Research Board, 2149: 11-20, http://dx.doi.org/ 10.3141/2149-02.

Montella, A., Aria, M., D'Ambrosio, A., Galante, F., Mauriello, F., Pernetti, M., 2011. Simulator evaluation of drivers' speed, deceleration and lateral position at rural intersections in relation to different perceptual cues. Accident Analysis and Prevention, 43(6). 2072-2084. http://dx.doi.org/10.1016/j.aap.2011.05.030.

Montella, A., Pariota, L., Galante, F., Imbriani, L.L., Mauriello, F., 2014. Prediction of drivers' speed behaviour on rural motorways based on an instrumented vehicle study. Transportation Research Record: Journal of the Transportation Research Board. 2434, 52-62, http://dx.doi.org/10.3141/2434-07

Montella, A., Imbriani, L.L., Marzano, V., Mauriello, F., 2015a. Effects on speed and safety of point-to-point speed enforcement systems: evaluation on the urban motorway A56 Tangenziale di Napoli. Accident Analysis and Prevention 75, 164-178, http://dx.doi.org/10.1016/j.aap.2014.11.022.

Montella, A., Punzo, V., Chiaradonna, S., Mauriello, F., Montanino, M., 2015b. Point-to-Point speed enforcement systems: Speed limits design criteria and analysis of drivers' compliance.. Transportation Research Part C: Emerging Technologies 53, 164-178, http://dx.doi.org/10.1016/j. trc.2015.01.025.

Muhrer E, Vollrath M (2010) The consequences for driving behaviour in a simulated driving task. Accident Analysis & Prevention, 42(6): 2158-2164.

NHTSA, 2012. Traffic Safety Facts 2011. Report DOT HS 811 754, Washington, D.C.

NHTSA, 2013a. Preliminary Statement of Policy Concerning Automated Vehicles System. Washington, D.C.

NHTSA, 2013b. Traffic Safety Facts 2012. Report DOT HS 812 032, Washington, D.C.

NHTSA, 2014. Traffic Safety Facts 2013. Report DOT HS 812 139, Washington, D.C.

Oh C, Kim T (2010) Estimation of rear-end crash potential using vehicle trajectory data. Accident Analysis & Prevention, 42(6): 1888-1893.

Pariota L (2013) Driving behaviour for ADAS: theoretical and experimental analyses. Dissertation, University of Naples. http://www.fedoa.unina.it/9074/1/Pariota_Luigi_25.pdf. Accessed 9 February 2015

Pariota L. Bifulco, G.N. (2015) Experimental evidence supporting simpler Action Point paradigms for car-following. Transportation Research Part F: Traffic Psychology and Behaviour, 35:1-15, doi:10.1016/j.trf.2015.08.002.

Pariota L, Bifulco G.N., Brackstone M (2015, on-line pub.) A linear dynamic model for driving behaviour in car-following. Transportation Science (articles in advance), doi: 10.1287/trsc.2015.0622

Parker MT (1996) The effect of heavy good vehicles and following behaviour on capacity at motorway roadwork sites. Traffic Engineering and Control, 37: 524–531.

Payre W, Cestac J, Delhomme P (2014) Intention to use a fully automated car: Attitudes and a priori acceptability. Transportation Research Part F: Traffic Psychology and Behaviour, 27(B): 252-263. http://dx.doi.org/10.1016/j.trf.2014.04.009

Piao J, McDonald M (2003) Low speed car following behaviour from floating vehicle data. Proceedings Intelligent Vehicles Symposium: 462-467.

Punzo, V., Montanino, M., Ciuffo, B. (2015). Do We Really Need to Calibrate All the Parameters? Variance-Based Sensitivity Analysis to Simplify Microscopic Traffic Flow Models. Intelligent Transportation Systems, IEEE Transactions on, 16(1), 184-193.

Ranjitkar P, Nakatsuji T, Azuta Y, Gurusinghe GS (2003) Stability analysis based on instantaneous driving behaviour using car-following data. Transportation Research Record: Journal of the Transportation Research Board, 1852: 140-151.

Saifuzzaman M, Zheng Z (2014) Incorporating human-factors in car-following models: A review of recent development and research needs. Transportation Research Part C: Emerging Technologies, 48: 379-403.

SETRA, Roads and Motorways Engineering Department, 2001. National Instruction on Technical Design Requirements for Rural Motorways.

Simonelli F, Bifulco GN, De Martinis V, Punzo V (2009) Human-like adaptive cruise control systems through a learning machine approach. Application of Soft Computing. Advances in Soft Computing, 52: 240-249. doi 10.1007/978-3-540-88079-0_2

Song, M., Wang, J.H., 2010. Studying the tailgating issues and exploring potential treatment. Journal of the Transportation Research Forum 49 (3), 69–86.

Strand N, Nilsson J, Karlsson ICM, Nilsson L (2014) Semi-automated versus highly automated driving in critical situations caused by automation failures. Transportation Research Part F: Traffic Psychology and Behaviour, 27(B): 218-228. http://dx.doi.org/10.1016/j.trf.2014.04.005

TAC, Transportation Association of Canada, 1999. Geometric Design Guide for Canadian Roads, Ottawa.

Tarko A, Boyle L, Montella A (2013) Emerging research methods and their application to road safety. Accident Analysis and Prevention, 61: 1-2. doi: 10.1016/j.aap.2013.07.006

Taylor J, Zhou X, Rouphail NM, Porter RJ (2015) Method for investigating intradriver heterogeneity using vehicle trajectory data: A Dynamic Time Warping approach. Transportation Research Part B: Methodological, 73: 59-80.

Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2): 411-423.

Trimble TE, Bishop R, Morgan JF, Blanco M (2014). Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts. Report No. DOT HS 812 043. Washington, DC: National Highway Traffic Safety Administration.

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. Accident analysis & prevention, 35(3), 427-433.

Vollrath M, Schleicher S, Gelau C (2011). The influence of cruise control and adaptive cruise control on driving behaviour – A driving simulator study. Accident Analysis and Prevention, 43(3), 1134–1139.

Wagner P (2011) A time-discrete harmonic oscillator model of human car-following. The European Physical Journal B-Condensed Matter and Complex Systems, 84(4): 713-718,

Wagner P (2012) Analyzing fluctuations in car-following. Transportation Research Part B: Methodological, 46(10): 1384-1392.

Wiedemann R (1974) Simulation des Strassenverkehrsflusses. Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, Band 8, Karlsruhe, Germany.

Wilson RE (2008) Mechanisms for spatio-temporal pattern formation in highway traffic models. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366(1872): 2017-2032.