Using GPS data to detect critical events in motorcycle rider behaviour

Salvatore Cafiso* and Alessandro Di Graziano

Department of Civil and Environmental Engineering,
University of Catania,
Viale Andrea Doria 6,
I-95125 Catania, Italy
Fax: +390957382247
Email: dcafiso@dica.unict.it
Email: adigraziano@dica.unict.it
*Corresponding author

Oliver Giudice

Department of Mathematics and Computer Science,
University of Catania,
Viale Andrea Doria 6,
I-95125 Catania, Italy
Email: giudice@dmi.unict.it

Giuseppina Pappalardo

Department of Civil and Environmental Engineering,
University of Catania,
Viale Andrea Doria 6,
I-95125 Catania, Italy
Fax: +390957382247
Email: giusy.pap@dica.unict.it

Abstract: A relevant research issue in naturalistic studies is related to analysis and evaluation of the huge data recorded during the experiment. A procedure to identify critical events of PTW is presented and tested using parameters measured or derived by GPS. Raw measures of speed, position and time are copied directly from the data string stored by the GPS in National Marine Electronics Association (NMEA) protocol. From these other measures useful in explaining the rider behaviour were derived. Smoothing cubic spline and Butterworth filters were used to improve the estimation of measures effected by GPS inaccuracy. The Mahalanobis distance was adopted to identify multivariable outliers for each combination of raw and derived measures. The events classified by Mahalanobis distance as outliers from the ‘normal’ rider behaviour were compared with the critical traffic conflicts self-reported by the rider during the experimental test in order to evaluate recall and precision of the method.

Keywords: naturalistic study; motorcycle; GPS; speed; position; video; Mahalanobis distance; time; spline; outlier.


Biographical notes: Salvatore Cafiso has been a member of Teaching and Research Staff since 1987. From 2001, he is a Professor of Highway Engineering at the University of Catania, Italy. From 2002 to 2013, he is a Coordinator of the PhD course in ‘Hydraulic, environmental and transportation infrastructures’ at the University of Catania. He has over 20 years experience of research into highway design, highway safety and pavement management with more than 150 publications in scientific literature. He is a member of national and international scientific committees (e.g., PIARC, TRB) and reviewers of international journals.

Alessandro Di Graziano holds a PhD in Engineering of Road Infrastructures and works since 1997 at the Department of Civil and Environmental Engineering (DICA) of University of Catania, where he has been a Lecturer in Airport and Railway Infrastructure. His scientific activity was mainly carried out on road management with special emphasis on safety and...
1 Introduction

All over the world the number of motorcycles registered powered two-wheelers (PTWs) was constantly increasing. In 2009, there was currently an estimated of 35 million PTWs in Europe: this amount represented about 14% of the entire private vehicles, but they accounted for around 19% of the fatalities (34% in Italy). Crash data showed as PTWs were the most vulnerable of powered transport mode (MAIDS, 2010; Cafiso et al., 2012).

A naturalistic driving study is a study undertaken to provide insight into driver behaviour during everyday trips by recording details of the driver, the vehicle and the surroundings through unobtrusive data gathering equipment and without experimental control. These continuously recorded data for crashes, near crashes and normal driving conditions allow for more sensitive analyses than what is currently available, even after a detailed crash investigation. In this field, a European study was 2-BE-SAFE, officially started on January 2009 as a focused research collaborative project co-funded by European Commission. It was the world’s first naturalistic riding study involving instrumented PTWs (2-BE-SAFE, 2012). In USA, there is one study that is known to be underway with no results being public, that MSF 100 motorcycle naturalistic study (MSF, 2012). When the study will be completed, researchers expect approximately 500,000 miles of riding data.

A relevant research issue in naturalistic studies is related to the data analysis and evaluation. Recording a long period of observation is necessary to limit the analysis only to the critical events (crashes, near crashes, accidents) in a huge set of data records. Correct identification of critical events is related both to quality and evaluation of data.

For this reason, all the referenced studies used high quality instruments. Systems and sensors equipping the vehicle were very advanced but also expensive so that only a limited number of vehicles can be equipped and monitored.

On the other hand, nowadays a large and increasing percentage of vehicles are equipped with in vehicle technologies (IVT) installed for driving assistance, safety and navigation. Despite of differences in the IVT purpose and development, a common characteristic is in the use of GPS as primary source of information for vehicle positioning and speed.

Given the easy availability of GPS, this paper proposed a methodology for identifying ‘critical events’ in naturalistic PTW studies using GPS data as primary source of information.

The recall and precision of the procedure in identifying actual near crashes will be tested comparing the results with the subjective identification and classification of the critical events provided directly by the PTW’s rider during the experiment.

2 The experiment

2.1 Instrumentation

In order to acquire a significant amount of data, a short term naturalistic study was carried out using the following instrumentation:

- A GPS, with a sampling frequency of 10 Hz (MTK II, L1 frequency, 66 channels, NMEA-0183). The GPS collected information on vehicle position and time. The receiver is also equipped with an ‘incident’ push-button that the driver can press to mark an event in the data storage.
- A digital camera with fixed focal lengths (720 × 480 pixel), synchronised with the GPS in order to provide a continuous window into the happenings around the motorcycle.
Using GPS data to detect critical events in motorcycle rider behaviour

2.2 Test course

The test course, approximately 40 minutes long, was selected in the urban area of Catania (Italy) (Figure 1). The test was performed in daylight and good weather conditions avoiding congestion during peak traffic hours.

No particular training was required and drivers were instructed to drive normally using their own motorcycle. They were asked only to mark the events that they consider a severe traffic conflict by pushing the button positioned close to the right handlebar grip.

2.3 Riding parameters

Riding parameters, recorded by the instrumentation, can be divided into raw and derived measures.

1 Raw measures

During tests, GPS measures are copied directly from the data string stored in accordance with NMEA protocol. The raw data used in this study were:

- Vehicle’s speed over ground. Accuracy is estimated in 0.1 m/s.
- Latitude (N) and Longitude (E). Absolute position accuracy is 3.0 m (2D-RMS).
- HDOP: horizontal dilution of precision.
- Time is received as GPS time and recorded in the NMEA data in UTC time. Time is accurate to about 50 nanoseconds.

2 Derived measures

From these raw data other measures useful in explaining the rider behaviour can be derived.

Usually, it is assumed that it is possible to detect safety critical driving behaviour by studying fluctuations in the longitudinal acceleration profiles (2-BE-SAFE, 2012; Cafiso et al., 2005). So the acceleration \( (A) \) and jerk \( (J) \) are calculated as the 1st and 2nd derivative of the speed with respect to time.

The parameter selected to describe vehicle horizontal path was the instantaneous curvature \( (1/R) \).

Other derived measures were: the centripetal acceleration, dependent on speed and curvature: \( AC = \frac{V^2}{R} \) and the angular change in the direction of trajectory \( \Delta \theta \) with respect to time \( V_A = \Delta \theta \Delta T \) and space \( DS = \Delta \theta \Delta S \) (\( \Delta T \) = time interval, \( \Delta S \) = travelled distance in the time interval).

These riding parameters (Table 1) were only a portion of those collected in the main PTW naturalistic studies previously introduced.

However, main parameters directly collected by the high resolution equipment can be represented by surrogate measures in our experiment (e.g., longitudinal speed is a surrogate of speed of rear wheel, curvature \( 1/R \) is a surrogate of steering angle).
Table 1  Comparison between monitored variables in 2-BE-SAFE (all direct measures) and in proposed study

<table>
<thead>
<tr>
<th>2-BE-SAFE</th>
<th>Proposed study</th>
<th>Type of measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal speed (km/h)</td>
<td>Longitudinal speed (km/h)</td>
<td>D</td>
</tr>
<tr>
<td>Longitudinal acceleration (g)</td>
<td>Longitudinal acceleration (m/s²)</td>
<td>D</td>
</tr>
<tr>
<td>Lateral acceleration (g)</td>
<td>Centripetal acceleration (m/s²)</td>
<td>S</td>
</tr>
<tr>
<td>Vertical acceleration (g)</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td>Yaw rate (deg/s)</td>
<td>Angular change rate (deg/sec)</td>
<td>S</td>
</tr>
<tr>
<td>Pitch rate (deg/s)</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td>Roll rate (deg/s)</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td>Throttle position (%)</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td>Brake rear (%)</td>
<td>Longitudinal acceleration (m/s²)</td>
<td>S</td>
</tr>
<tr>
<td>Steering angle (%)</td>
<td>Path curvature (1/m)</td>
<td>S</td>
</tr>
<tr>
<td>Brake activation front (0/100)</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td>Brake activity rear (0/100)</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td>Speed of the rear wheel (km/h)</td>
<td>Longitudinal speed (km/h)</td>
<td>S</td>
</tr>
</tbody>
</table>

Notes: D = direct measure; S = surrogate measure; M = missing measure

Obviously, this approach decreases the precision of the measure, but reducing significantly the cost of equipment and vehicle instrumentation, too.

It is estimated a ten of one’s cost for vehicle equipping in this study when compared to a high level instrumentation used in other naturalistic studies.

3 Data treatment and evaluation

3.1 Vehicle trajectory

The horizontal path is described by a set of points taken by the GPS during experiment where the default coordinates (N, E) were converted to a local grid system (X, Y). This set of points have some errors with respect to the actual location which may introduce unexpected results on the curvature and direction measures. To solve this problem, a method of curve fitting based on cubic smoothing spline (Castro et al., 2006; Cafiso and Di Graziano, 2008) was used to describe the PTW’s trajectory through the GPS locations. Unlike interpolation function which creates a curve that passes exactly through all points, a smoothing spline $S(x)$ comes only reasonably close to the data.

Given $x_i$ and $y_i$ the points coordinates in a local grid system and $\sigma$ the standard deviation of data $\chi^2$ is a measure to quantify how close the spline comes to the data:

$$\chi^2 = \sum_{i=0}^{n} \frac{(s(x_i) - y_i)^2}{\sigma_i^2}$$  \hspace{1cm} (1)

If the spline is a good representation of the data the average value of each term in ‘$\chi^2$’ is expected to be about one. To quantify how much the curve is approximating our data the integral of the square of the second derivative of the spline must as much small as possible:

$$\int |S''(x)|^2 \, dx$$  \hspace{1cm} (2)

But this is in contrast with equation (1) because the smaller is the (2) the bigger is the (1). From these assumptions, to optimise the process it is necessary that the smoothing cubic spline minimises:

$$W = \rho \chi^2 + \int |S''(x)|^2 \, dx$$  \hspace{1cm} (3)

The constant ‘$\rho$’ determines the tradeoff between (1) and (2). Putting $\rho = 0$ the spline becomes a straight line, instead $\rho = 1$ puts high emphasis on minimising $\chi^2$, forcing the spline to interpolate (de Boor, 1978).

The fitting curve was found by using MATLAB’s spline toolbox which creates a cubic smoothing spline in its piecewise polynomial form (PP-form). The PP-form is a numerical implementation that, as the name suggests, divides the spline curve into pieces. Then for each piece it was possible to define different values of $\rho$ in equation (3).

Higher the precision in GPS data, closer to 1 would be $\rho$. Therefore $\rho$ was posed in function of HDOP. HDOP is a measure in the precision of GPS horizontal location. Neglecting ionospheric and tropospheric effects, the relative satellite-receiver geometry plays a role in determining the precision of estimated positions. Similarly, the greater the number of satellites, the better the value of HDOP. Particularly, as different HDOP values can arise in the case of kinematic applications and rapid recording processes (Figure 2), differentiating $\rho$ parameter was useful to increase smoothing (i.e., trajectory not exactly through GPS points) only where GPS precision decreases.
Using GPS data to detect critical events in motorcycle rider behaviour

Figure 2  Different HDOP in (a) good (HDOP = 0.88) and (b) bad (HDOP = 3.32) satellite-receiver geometry and number of satellites (see online version for colours)

Figure 3  Equation for fitting a circular curve

Given the spline computed as described above, a new set of coordinates \((x_i', y_i')\) was assigned to the PTW’s path along the smoothing spline. Replacing the original GPS coordinates \((x_i, y_i)\) with the new set \((x_i', y_i')\) which smoothes the GPS error, the curvature path was calculated for every three points using simply the equation of depicting a circle’s radius from three points on the circumference (Said and Halim, 2007; Imran et al., 2006) (Figure 3).

\[
R = \sqrt{(x_i' - x_0)^2 + (y_i' - y_0)^2}
\]  

(4)

Analogously, the angular change \(\Delta \theta_i\) in the trajectory at \(P_i(x_i, y_i)\) can be computed with the formula:

\[
\Delta \theta_i = \arctan\left(\frac{y_i' - y_{i-1}'}{x_i' - x_{i-1}'}\right) - \arctan\left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right)
\]  

(5)

Also speed, acceleration and jerk showed some noise due to the GPS accuracy. With an analogy to signal processing, to eliminate noise (i.e., artefacts) maintaining the information, an appropriate filter was applied. The purpose of a filter was the elimination of part of the harmonic content of a signal (unwanted frequencies) while leaving unchanged the remaining portion (wanted frequencies). An ideal filter “should not only completely reject the unwanted frequencies but should also have uniform sensitivity for the wanted frequencies” (Butterworth, 1930). Such an ideal filter cannot be achieved, but Butterworth showed that successively closer approximations is obtained with increasing numbers of filter elements of the right values. Butterworth filter design and specification are not of interest in the present paper, therefore only the main practical application and results are reported in the following. The reader can find more information in the referenced source (Butterworth, 1930).
In the present study, the filtering parameters were set to:

- cut-off frequency: $W_n = 0.20$
- filter order: $N = 6$.

This solution was able to cut single events with an abrupt change in motorcycle kinematic and a very short duration (i.e., less than 0.5 sec) which can be considered artefacts. This is consistent with an average duration for a traffic conflict of more than one second as estimated both during riders’ interviews and data analysis.

### 3.2 Outlier detection

After smoothing and filtering the data, the measures were used to identify the occurrence of a critical event in PTW riding. A critical event was defined as an outlying values in the riding parameters.

In complex and highly volatile phenomena such as the evolution of the riding variables, one or more variables may be considered as outliers while the whole riding behaviour defined by the joint consideration of more variables may not be a multivariate outlier and vice versa. Therefore, the use of multivariate methods to identify outlier is to be preferred (Vlahogianni et al., 2012). We used the Mahalanobis’ distance that is a well-known outlier detection technique in multivariate data. The Mahalanobis’ distance identifies observations that lie far away from the centre of the data cloud, giving less weight to variables with large variances or to groups of highly correlated variables (Jolliffe, 1986). This distance is often preferred to the Euclidean distance which ignores the covariance structure and treats all variables equally.

Given $n$ observations from a $p$-dimensional dataset, the Mahalanobis’ distance ($M_i$) for each multivariate data point $i$, $i = 1, \ldots, n$, is given by

$$ M_i = \left[ \sum_{i=1}^{n} (x_i - \bar{x}_n)^T V_n^{-1} (x_i - \bar{x}_n) \right]^{1/2} $$

(6)

where $\bar{x}_n$ denotes the sample mean vector ($p \times 1$) and $V_n$ the sample covariance matrix ($p \times p$):

$$ V_n = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)(x_i - \bar{x}_n)^T $$

(7)

It is well known that both mean and standard deviation are extremely sensitive to outliers. Therefore, the presence of outliers affects the Mahalanobis distance technique. Specifically, datasets with multiple outliers are subject to masking and swamping effects (Ben-Gal, 2005):

- Masking effect: it is said that one outlier masks a second outlier, if the second outlier could be considered as an outlier only by itself, but not in the presence of the first outlier. Thus, after the deletion of the first outlier the second is emerged as an outlier;
- Swamping effect: it is said that one outlier swamps a second observation, if the latter could be considered as an outlier only under the presence of the first one. In other words, after the deletion of the first outlier the second observation become a non-outlying observation.

Masking and swamping effects play an important role in the effectiveness of the Mahalanobis distance as a criterion for outlier detection. Masking effects may decrease the Mahalanobis distance of an outlier, swamping effects may increase the Mahalanobis distance of non-outlying observations.

To deal with these effects, a robust estimator of Mahalanobis distance has to be used. The purpose of robust estimation is to produce an optimal estimator in the presence of outliers, while minimising bias. This is done by reducing the influence of the outliers on the estimator. There are different tools which can be applied to make it, in the present work an iterative process was applied to remove all outliers from the dataset. After the data reduction, the new set of data was used to carry out mean and standard deviation to be used in Mahalanobis distance computation.

Data reduction and filtering, improved the computation of Mahalanobis distance reducing the large number of outliers produced by the high variability in the raw data (Figure 4).

Finally, to identify outliers with the Mahalanobis’ distance ($M_i$), it can be considered that for multivariate normally distributed data, the variable $\frac{(n-p)\alpha}{(n-1)\alpha(n-p)} \times M_i$ has an approximate F distribution with $p$ and $n-p$ degrees of freedom ($F_{p,n-p}$) (Afifi and Azen, 1972). Therefore at a confidence level $\alpha$, a determination as to whether an observation can be considered as outlier or not, can be made based on the comparison between the Mahalanobis distance $M_i$ and the critical value of the F-distribution at the $\alpha$ level of confidence ($F_{p,n-p}^{\alpha}$). Accordingly, those observations with a Mahalanobis distance ($M_i$) greater than the critical value are indicate as outliers (Hawkins, 1980; Penny and Jolliffe, 2001):

$$ M_i \geq \frac{(n^2-1)}{(n-p)n} \times F_{p,n-p}^{\alpha} $$

(8)

Finally, for each of the 22 combinations of the raw and derived measures, an event characterised by a Mahalanobis distance above the defined percentile (i.e., $\alpha = 0.95$) was classified as outlier (Mahalanobis positive, MP) (Figure 5).
4 Experimental results

In a safety study based on large scale naturalistic experiments the first interest is to detect in the huge database, crashes (if they occur) and other traffic conflict situations (near crashes, traffic conflicts) of interest for the analysis. The capability of the proposed experimental approach to give accurate results was checked comparing outliers identified by Mahalanobis distance with actual ‘critical events’ marked by riders during the tests. Riders were instructed to identify a traffic conflict as a situation in which the PTW approaches another road user in space and time to such an extent that there is risk of collision if their movements remain unchanged (Hyden, 1987).

The events identified by the Mahalanobis distance as outliers were marked ‘MP’, the events identified by the rider as actual traffic conflicts were marked ‘rider positive’ (RP).
From the comparison there were four possible outcomes (Figure 6):

- **true positive**: if the event is classified both as MP and RP
- **false negative**: if the event is not classified as MP, but it is classified as RP
- **true negative**: if the event is not classified as MP and RP
- **false positive**: if the event is classified as MP, but it is not classified as RP.

The performance of the method for each combination of riding parameters was evaluated from the count of the total number of TP, FP and FN in the experiment, using the following indicators:

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100
\]

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100
\]

Recall indicates the proportion of true positives (TP) identified by the method in the total amount of actual positive marked by the rider (RP = TP + FN). The higher the ‘recall’ value is, the higher the probability of correctly detecting all the events classified as critical by the rider (RP).

Precision indicates the proportion of true positives (TP) identified by the method in the total amount of outliers (MP = TP+FP). The higher the ‘precision’ value is, the higher the probability of correctly addressing an event as true positive.

Overall experimental results in terms of recall and precision are reported in Figure 7.

Results showed that using more than one riding parameter, can improve both precision and recall. For example the combination of acceleration (A) and jerk (J) gave higher recall (30%) and precision (6.25%) than each parameter considered alone (recall = 19%, precision = 4.8% and recall = 14%, precision = 5.8%, respectively). The best performances were reached combining three and four parameters.

**Figure 6** Confusion matrix

<table>
<thead>
<tr>
<th>Rider</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mahalanobis</td>
<td>True positive</td>
<td>False positive</td>
</tr>
<tr>
<td>Positive</td>
<td>False negative</td>
<td>True negative</td>
</tr>
</tbody>
</table>

**Figure 7** Recall and precision (see online version for colours)

Note: A = acceleration, J = jerk, AC = centripetal acceleration, VA = angular change in the direction of trajectory with respect to time, DS angular change in the direction of trajectory with respect to space.
5 Conclusions and discussion

Relevant research issues in naturalistic studies are related at beginning of the experiment to equipping vehicles with expensive instrumentations and at the end to analysing the huge data recorded during the experiment. A limitation in the spreading of these studies can be found in the use of high resolution but also expensive systems used to equip and monitor a limited number of vehicles. On the other hand, nowadays a large and increasing percentage of the vehicle fleet is equipped with GPS for driving assistance, safety and navigation.

PTWs, recently introduced as monitored vehicle in a limited number of naturalistic study, presents further issues related to limitation in the space to install instrumentation and complexity in vehicle dynamics if compared to other vehicle types (e.g., passenger car, truck, bus)

Starting from these considerations, this study evaluated the capability to identify the occurrence of a critical event in PTW riding using only high frequency GPS data and video recording.

The set of riding parameters which can be directly measured or derived by GPS, the filtering process to improve the quality of the measures and the statistical method to identify critical events in the dataset are presented in the paper to provide the rider with information useful for similar studies.

Results, even if based on a limited experimental test covering about two hours of PTW’s riding, were useful to test the procedure and to evaluate the capability to correctly identify critical traffic conflicts. Beside, smoothing cubic spline and filtering the measures to reduce noise improved the quality of results and Mahalanobis distance was able to identify multivariable outliers (MP), recall and precision of the proposed method reached values not higher than 60% and 10% respectively. These results are not in line with the targets of 90% recall and 70% precision fixed in the 100-car study (Klauer et al., 2006). Even if improvement in our results can be expected with a sensitivity analysis of varying frequency acquisition (i.e., 10, 20 Hz) is necessary only for high vehicle speed (rural roads, motorways). In urban area a lower frequency could be appropriate, but usually should not less than 5 Hz.

For this specific application, an advantage in the outlier detection using Mahalanobis is given by the relative evaluation of the distance with respect to the average.

In the evaluation of the driving data, this implies that for a given rider, the above algorithm may distinguish between typical mean-riding patterns and irregular-far from the mean-riding behaviour. The larger the deviation the greater the change of the rider’s riding style. With respect to the more traditional use of threshold values (e.g., maximum acceleration of 0.5 g), Mahalanobis distance was able to combine multiple parameters and to compare results with a percentile of the ‘normal’ behaviour of the rider.

At the present stage of the research, authors rely that the proposed system and process can be used to analyse driver behaviour in traffic conflict situations and results could be used, also, to design advanced warning system, but these system have to be based on equipment with higher reliability and precision.

5.1 Lessons learned

Regarding the use of GPS as primary source of data, the study confirmed as GPS is simple and cheap to install. The system used has a cost of 2,000 Euros, with capability to store GPS data at 10 Hz synchronised with two cameras video recording. The system provides location and time information in all weather, anywhere on the earth where there is a line of sight to four or more GPS satellites. High frequency acquisition (i.e., 10, 20 Hz) is necessary only for high vehicle speed (rural roads, motorways). In urban area a lower frequency could be appropriate, but usually should not less than 5 Hz.

For this specific application, an advantage in the outlier detection using Mahalanobis is given by the relative evaluation of the distance with respect to the average.

In the evaluation of the driving data, this implies that for a given rider, the above algorithm may distinguish between typical mean-riding patterns and irregular-far from the mean-riding behaviour. The larger the deviation the greater the change of the rider’s riding style. With respect to the more traditional use of threshold values (e.g., maximum acceleration of 0.5 g), Mahalanobis distance was able to combine multiple parameters and to compare results with a percentile of the ‘normal’ behaviour of the rider.

At the present stage of the research, authors rely that the proposed system and process can be used to analyse driver behaviour in traffic conflict situations and results could be used, also, to design advanced warning system, but these system have to be based on equipment with higher reliability and precision.

5.2 Further research

To make comparable the evaluation of recall and precision with other naturalistic studies (i.e., 100-car, 2-BE-SAFE), the research is in progress using both video recording and rider judgment to identify the total number of incidents and near crashes occurred during the test.

In the experiment, riders were instructed to classify conflicts, but the identification remain subjective both in terms of identification and classification. This topic has to be better investigated in future researches. For these reasons, in the present paper no distinction was made in terms of conflict severity.

Table 2 Classification of events in 100-car naturalistic studies

<table>
<thead>
<tr>
<th>Event category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crashes</td>
<td>Any contact between the subject vehicle and another vehicle, fixed object, pedestrian, pedacyclistic, animal</td>
</tr>
<tr>
<td>Near crashes</td>
<td>Defined as a conflict situation requiring a rapid, severe evasive manoeuvre to avoid a crash</td>
</tr>
<tr>
<td>Incidents</td>
<td>Conflict requiring an evasive manoeuvre, but of less magnitude than a near crash</td>
</tr>
</tbody>
</table>
A sensitivity analysis will be performed varying trigger values of smoothing, filter and Mahalanobis percentile distance to optimise the recall and precision, using receiver operating characteristics (ROC) curve as alternative to plot the proportion of successful identifications vs. proportion of false identifications of the phenomenon.

References


