Identifying Riding Profiles Parameters from High Resolution Naturalistic Riding Data

E. I. Vlahogianni, Ph.D.*, George Yannis, Ph.D. John C. Golias, Ph.D.
National Technical University of Athens
5 Iroon Polytechniou Str, Zografou Campus, 15773 Athens Greece
elenivl;geyannis;igolias@central.ntua.gr

N. Eliou, Ph.D. and P. Lemonakis
University of Thessaly,
Pedion Areos 38334 Volos, Greece
neliou;plemonak@uth.gr

**corresponding author*

ABSTRACT

A simple and flexible methodology is proposed in order to define Power-Two Wheelers (PTWs) riding profiles by distinguishing between regular and irregular PTW behaviors, by using high resolution naturalistic riding data. "Irregularities" in riding behavior are consistently expressed as outlying values in the multivariate consideration of a set of riding parameters. The detected irregularities are those that diverge from the centroid of the jointly considered riding variables and define critical riding situations that are further associated to typical riding events. Results indicate that the joint consideration of variables that are directly connected to the mechanical characteristics of the PTW, such as breaking, wheel speed, throttle and steering, are adequate to distinguish the regular from irregular riding behavior. Moreover, a regressor is constructed using neural networks and the influential determinants to the deviation from the rider's regular actions are evaluated.

Keywords: Power Two Wheelers, naturalistic riding, rider behavior, microscopic incident analysis, outlier detection, multi-resolution analysis, neural networks.

INTRODUCTION

Risky rider's behavior has been for long a critical consideration in Power Two Wheelers' (PTW) safety. Understanding the riding behavior is critical to the design of efficient accident countermeasures, and is hence essential (Yannis et al., 2005). The best way for understanding riding behavior given improvements in modern technology is through monitoring, recording and analyzing it. Until recently, PTW accident risk has been largely studied through macroscopic and in-depth data analyses (Thomas et al. 2005, Dupont et al. 2009, Yannis et al. 2010), as well as through behavior analyses such as questionnaire based surveys, guided discussions, video-based methods or simulators (Engstrom et 2005, Savolainen and Mannering 2007, Haque et al. 2010, Huang and Abdel-Aty 2010); these analyses are inherently destined to qualitatively assess on the factors that increase accident risk or causalities involved mainly from a social representation theoretic point of view, without being able to extract accurate and detailed information on the manner riders behave on the road and especially before, during and after critical situations.

A new and efficient way to understand riding behavior is by creating a least intrusive - restrictive (naturalistic) - environment to monitor and record riders on the road by employing advanced sensor technologies. A number of such attempts have been conducted in Europe, the US and Australia to understand driver behavior. Some prevailing efforts are described in large scale projects such as the 100-Car study (NHTSA 2006), SHRP 2 Naturalistic Driving Study (SHRP2 2011) and Euro-FOT (Csepinszky Benmimoun 2010); several findings on commercial vehicle driver's behavior and inattention are summarized in Olson et al. (2009) and Klauer et al. (2010). However, so far no results are publicly available concerning rider behavior (Reagan et al. 2006, NHTSA 2006, FESTA 2008, Baldanzini et al. 2009).

A key problem in naturalistic riding studies is to define which riding situation may be considered as critical or risky. Interestingly, in all relevant approaches documented so far, typical driving parameters' thresholds are established, and, based on those values criticalities are extracted and further analyzed (Baldanzini et al. 2009). This technique lacks consistency with the fact that each driver/rider has its personal stock of values, ideas, beliefs and practices, reflecting rigorously on its behavior on the road, such as the braking, overtaking and so on, that may not converge to a typical rider's behavior. In this context, the question that emerges is whether the high resolution naturalistic riding data which is in nature multivariate and noisy can be used to define a self-contained personalized rider's profile.

The objective of this paper is to propose a methodology for identifying a rider's profile based exclusively on high resolution naturalistic riding data without observing the videos. The data exploited concern information on wheel speed, acceleration, throttle, steering, rear braking and so on. A comprehensive methodological shell is proposed in order to distinguish between regular and irregular riding behavior. "Irregularities" in riding behavior are consistently expressed as outlying values in a multivariate consideration of the available riding parameters. The detected irregularities are those values that diverge from the centroid of the jointly considered riding variables and define critical riding situations that may further be associated to typical riding events. Furthermore, artificial intelligence techniques are used to construct a regressor and uncover the influence of riding profile parameters.

METHODOLOGY FOR IDENTIFYING RIDING PROFILES PARAMETERS

The proposed naturalistic riding data-driven methodology for identifying riding profiles consists of three consecutive steps: is the first step, a robust statistical approach is implemented in order to detect the time a deviation from the mean riding behavior occurs that may signify the onset of a critical riding situation (incident). The outlier detection is based on a multivariate set of riding parameters that are jointly considered. A statistical testing strategy will act complementary to the outlier detection methodology; different models in terms of input data considered will be evaluated with respect to their ability to detect critical situations.

After detecting all critical incidents, the next step is to develop a Multi-layer Perceptron (MLP) classifier in order to develop the interrelation between the riding parameters and the deviation from the mean riding pattern. The model will also implement a sensitivity analysis in order to assist the explanatory power of the MLP classifier. The methodology is briefly depicted in Figure 1.

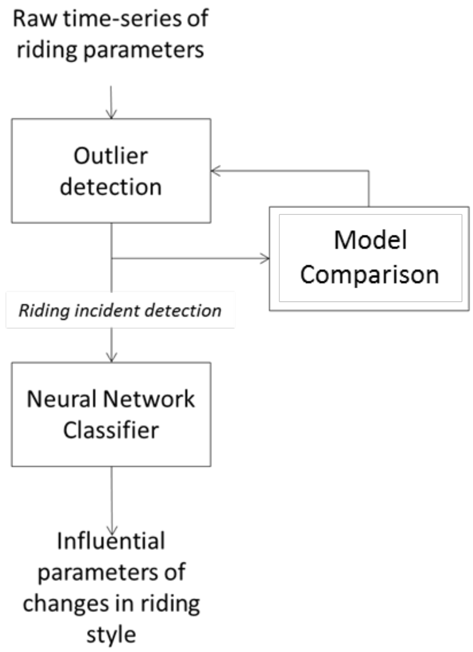

Figure 1: The proposed methodology for identifying riding profiles critical parameters.

Outlier Detection

Although, intuitively, naturalistic data can be used to analyze transportation safety and driving behavior in various ways, their spatial and temporal characteristics is a multiplier of the complexity of any analyses – in both concept and computational intensity - that may be employed. Riding data collected in extremely high resolutions (e.g. 100Hz) often contain outliers. The goal of the proposed methodology is to detect the outliers and then construct a classifier to characterize the parameters that affect the riding style. Outlier detection is a critical step in data mining aiming at describing the abnormal data behavior as reflected in the data that deviates from the natural data variability (Hodge and Austin 2004). According to Barnett and Lewis (1994), an outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data. Outlier detection has been applied in many fields such as chemical engineering and data mining; some specific fields of research include fraud detection, activity monitoring, structural defect detection, fault diagnosis, medical condition monitoring, motion segmentation etc. (Hodge and Austin 2004).

In the specific study, an outlier is defined as an observation for a specific time interval that the rider for some reason drastically alters its riding behavior due to and external or internal stimulus. There exist both univariate and multivariate approaches to outlier detection. In complex and highly volatile phenomena such as the evolution of the riding variables, the multivariate consideration is imperative, as, for a specific time interval, although one or more variables may be considered as outliers, the whole riding behavior defined by the joint consideration of all variables may not be a multivariate outlier. In the specific study a simple and flexible methodology based on the Mahalanobis distance is applied; in contrast to the Euclidean distance, the Mahalanobis distance takes into account the correlation structure of the data as well as the individual scales (Barnett and Lewis 1994). Let $X_i = (x_{i1},...,x_{ip}),\ i=1,...,n$ be the multivariate space of p riding parameters that independently come from a multivariate normal distribution $X = N(\mu, \sigma^2)$, where μ is the mean and σ is the covariance matrix, the Mahalanobis distance d_i is defined as:

$$d_i = \sqrt{(x_i - \hat{\mu}) S^{-1} (x_i - \hat{\mu})} \qquad (1)$$

Where $\hat{\mu}$ and S^{-1} are the sample mean and covariance matrix respectively.

The Mahalanobis distance can be approximated by an F-distribution $\left[p(n-l)(n+l)/n(n-p)\right]F_{p,n-p}$; at a significance level α, a determination as to whether a new observation X_i can be considered as outlier or not can be made based on the following formula:

$$d_i \leq \left[p(n-l)(n+l)/n(n-p)\right]F_{p,n-p} \qquad (2)$$

The described methodology suffers from the masking effect[1] and the swamping effect[2] (Barnett and Lewis 1994). To deal with these effects a robust estimator of the Mahalanobis distance is recommended that leaves out each observation in turn and calculate its scaled distance from the center using the rest of the data; this is known as the jackknifed distance. Jackknifing is a process where the multivariate distance for each observation/object is calculated using means, variances, and covariances that did not consider (i.e., not influenced by) the given observation.

The proposed approach requires defining which input data should be considered in the process of incident detection. In this paper, different models with different input spaces will be considered and the optimum input space will be identified using as criterion to produce the less data intensive model that will be able to detect all *significant* shifts from regular to irregular riding patterns. Statistical difference between models will be judged based on classical statistics and conditional entropy, an information theoretic criterion that quantifies

[1] An outlier is said to mask a second one that is close by if the latter can be considered an outlier by itself, , but not if it is considered along with the first one. Equivalently after the deletion of one outlier, the other instance may emerge as an outlier.
[2] An outlier swamps another instance if the latter can be considered outlier only under the presence of the first one

the uncertainty of a variable Y conditional on the variable X taking a certain value x provided by the following equation (Shcreiber 1999):

$$H_r(\mathbf{Y} \mid \mathbf{X}) = -\sum_{x \in X, y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)} \tag{3}$$

where *i*, *j* are the clusters και C_X, C_Y the total number of clusters in data clustering ***X*** και ***Y*** equally.

Neural Network Regressors

Neural networks belong to the class of computational intelligence models and have been proven efficient in various transportation problems, mainly due to their powerful data analysis capabilities induced by the process of learning. In transportation applications, one of the most popular neural networks structures is the Multilayer Perceptron (MLP), largely implemented in classification and regression problems (for a review see Karlaftis and Vlahogianni 2010). The Multilayer Perceptron (MLP) can be considered as a generalization of a single-layer Perceptron because it encompass a hidden layer consisting of a set of processing units; this hidden layer introduces non-linearity to the network (Principe et al. 1999). In a MLP, each layer consists of a set of processing units (neurons) that receive inputs and transfer them to the next layer through a set of connections (synapses). Each connection possesses certain strength, the synaptic weight. There are no connections between the units of the same layer. Furthermore, in the input layer no kind of processing is performed. MLPs are usually trained using the error back-propagation learning rule. The output of an MLP is provided by:

$$y_k = f_k\left(a_k + \sum_{jk} w_{jk} f_i\left(a_k + \sum_{ij} w_{ij} x_i\right)\right) \tag{4}$$

where y_k is the output of the MLP, a_k is the bias, w_{jk} and w_{ij} the weights between the hidden and output layer and the input and hidden layer equally, f_i and f_k the activation functions of the hidden and output layer units and x_i the input pattern. Neural networks are valuable in cases where the input-output relationships is complex and of unknown form. Previous research has found that an MLP with one hidden layer is capable of approximating a large class of functions to any desired degree of accuracy (Hornik 1991).

IMPLEMENTATION AND FINDINGS

Data Description and Preliminary Analysis

The data come from a naturalistic riding experiment conducted in and around the city of Volos, a medium size Greek city. 'Naturalistic' refers to a method of observation that captures driver behavior in a way that is most representative of typical driving and not influenced by the artificial features of controlled studies (Baldanzini et al. 2009). This method allows researchers to study drivers in their own vehicle and environment. Such data may provide insight into factors that can influence safe driving.

In the specific naturalistic riding study during the focused research collaborative project “2-BE-SAFE – 2-Wheeler Behavior and Safety”, the most important signals that are being monitored are given in Table 1. Moreover, video installation was used to capture the frontal environment (required a minimum 90° field of view) and the rider’s face. The motorcycle used is a BMW F650 Funduro. The data acquisition of all signals was recorded at an accuracy of 100Hz except for video signals which are sampled at 10Hz and GPS position which can be sampled at 1Hz (values indicate minimal requirements). The selection of the specific signals will ensure that the riders’ evasive maneuvers (necessity for: accelerations, speed, brake activation, throttle position and steering angle), as well as the interaction of riders with the environment will be continuously monitored.

Table 1: The list of monitored variables and their description.

Variable	Description
Longitudinal acceleration [g] Lateral acceleration [g] Vertical acceleration [g]	linear acceleration
speed [kph]	longitudinal speed
yaw rate [deg/s] pitch rate [deg/s] roll rate [deg/s]	roll, yaw and pitch rates
Throttle [%]	throttle position
Brake Rear [%]	brake pressure rear
Steering [%]	steering angle
Brake activity Front [0/100]	brake activation front
Brake activity Rear [0/100]	brake activation rear
Wheel Speed [km/h]	Speed of the rear wheel

Data selected for further study encompass approximately 20 minutes long trips made during a period of three months by one rider in rural two-way roads. The specific route has shoulder width less than 1,2m, rolling terrain and high curvatures, mixed traffic conditions, not equally distributed traffic across the two directions, a significant number of uncontrolled access points and several zones where passing is permitted; this setting seems to be challenging for PTW riders in terms of riding patterns’ complexity and difficulty in maneuvering. All trips were made in daylight with good visibility and fine weather conditions. The final dataset consists of 56 trips of 20 minutes duration, meaning a set of time series of riding parameters with $6.72 \cdot 10^6$ data.

Detecting Irregular Behavior

The proposed methodology is applied to the available riding variables in order to detect the outliers and distinguish between regular riding behavior and changes to irregular riding style. The first step of the analysis is to assess on the number of variables that should be taken into consideration. The second step of the analysis is to associate the observed outlying behavior to specific riding situations that could be characterized as critical. A look at Table 1 reveals two groups of available variables. The first is related to the mechanical characteristics and includes braking (front/rear brake activation and rear brake pressure), yaw rate, pitch rate and roll rate, throttle, steering and wheel speed. The second group refers to the traffic related variables such as linear acceleration components and speed. This distinction is made in order to differentiate those parameters that are directly related to rider's reactions from those that may be considered as outcome of others, such as acceleration.

On the basis of the above, it was decide that three distinct outlier detection models with respect to the different input space are further evaluated:

- Model 1: steering, throttle, brake activation and wheel speed.
- Model 2: linear acceleration and speed.
- Model 3: All available variables.

The criterion for selecting the optimum input space for the outlier detection methodology is to produce the less data intensive model that will be able to detect all *significant* shifts from regular to irregular riding patterns. Significance is judged by an F-distribution as described in Equation 2.

The three models are first compared on the basis of the distance metric (Equation 1) and its temporal structure based on the Wilcoxon signed rank test (Washington et al. 2010) that revealed different temporal evolution of the distance metric between the three different models. The difference between Model 1 and the other two Models with respect to the evolution of the riding patterns as depicted on the joint consideration of the related riding parameters is significant, whereas – according to the t-statistic – Model 2 does not significantly differ from Model 3.

Regarding the classification c_i of the riding patterns to regular or extreme/outlying patterns, a comparative study is established based on the conditional entropy. Results show that the conditional entropy for Model 2 and Model 3 knowing Model 1 ($H_r(c_{\text{Model2}} \mid c_{\text{Model1}})$ and $H_r(c_{\text{Model3}} \mid c_{\text{Model1}})$) equals to 0.021 and 0.025 respectively. These small values of conditional entropy indicate that the knowledge for the type of riding pattern coming from Model 2 and 3 when the classification from Model 1 is known does not provide any further information.

The three models are further compared based on the detected outliers that they produce using the F-distribution test. An example of the results of Model 1 on a 15 minute trip is seen on Figure 2 that depicts the distance metric (Equation 1) time series using as multivariate input space the time –series of the mechanical related variables; any distance metric value above the 5% threshold value signifies an outlier or a critical irregular riding behavior. As can be observed, the proposed methodology distinguishes between irregular and regular riding behavior and may reveal the specific time interval where a shift to irregular behavior may occur, as well as its duration. Moreover, the larger the metric distance the more extreme the riding behavior of the PTW user is, the more abrupt is the change to its riding style.

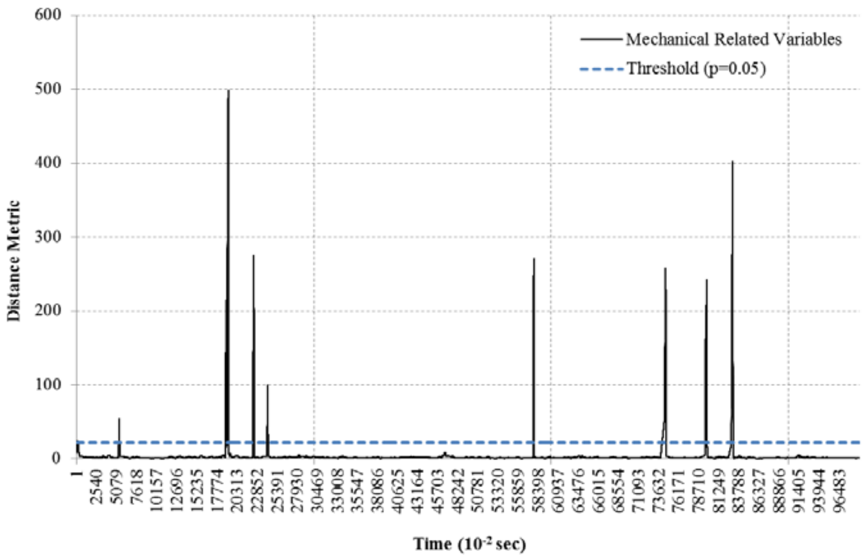

Figure 2: Distance metric time series using as multivariate input space the time –series of the mechanical related variables (Model 1). Any distance metric value above the 5% threshold value signifies an irregular behaviour (outlier).

Comparing the detected outliers from all three models, it is found that the larger the number of variables considered in the input space of the algorithm, the lesser the ability of the methodology to discern irregular from regular riding behavior and detect critical incidents degrades with the increase of the input space. This is probably due to the fact that certain variables are correlated (e.g. rear brake activation and break pressure, acceleration and throttle position). Moreover, it seems that correlations between variables mask (smooth out) the dynamics of riding behavior. This is clearly observed in Figure 3 where the distance metric of the time series of all available variables is depicted; a critical incident is canceled out when using the entire set of riding variables instead of the ones that are directly connected to the mechanical characteristics of the PTW.

The finding that the variables that are related to direct mechanical characteristics of the PTW are more influential to the process of detection can be attributed to the fact that speeding or accelerating/decelerating is the direct outcome of changes in throttle and brake activation. Furthermore, speeding or accelerating/decelerated may be the effect of more than one combination of the mechanical-related characteristics, such as steering, throttle and/or braking.

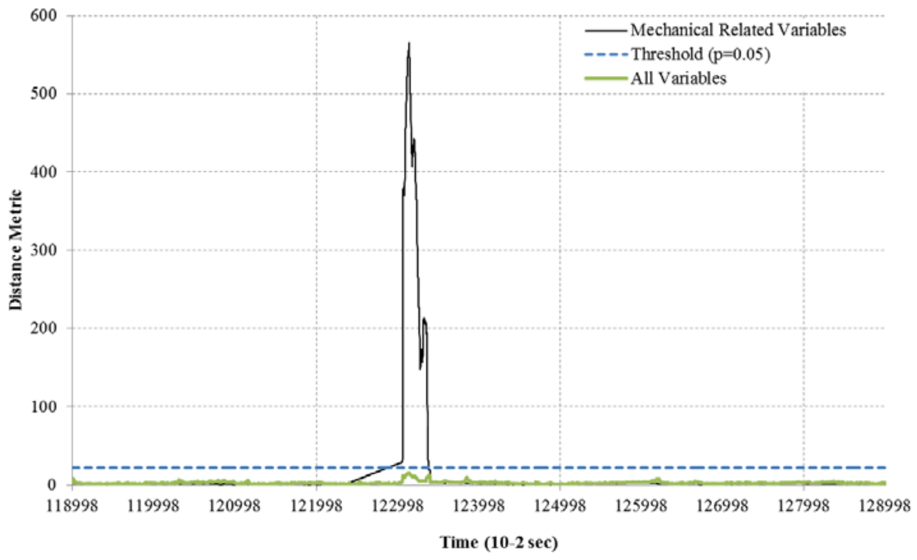

Figure 3: A single riding situation as detected by the distance metric time series using as multivariate input space the time –series of the mechanical related variables and all available variables. Any distance metric value above the 5% threshold value signifies an irregular behaviour.

Outliers detected using the mechanical parameters are further evaluated by observing the respective videos. All outliers detected point to a change – regardless of being smooth or abrupt - in the riding style.

Among the situations detected are:

- Moped braking and moving on the right to avoid opposing vehicle,
- Braking due to pedestrians in high grade, waiting at an intersection to enter main traffic,
- Moped moving on the left to avoid fixed object,
- Entering sharp turn when interacting with opposite lane's traffic,
- Braking and moving on the right after having overtaken vehicles (vehicles are in front of the moped as well),
- Overtaking more than one vehicle,
- Moped moving to the left to avoid stationary object.

These situations may be considered as incidents where a rider is engaged to an unusual - far from the mean - riding behavior and, thus, may be candidate riding situations with high accident risk. The detected set of incidents may vary with regards to the accident risk they encompass. Regarding the associated deviations from the mean riding styles, a meaningful result on which of the observed situations is associated to more extreme/ irregular riding behaviors cannot be extracted; further research is needed for uncovering the determinants of each riding behavior with respect to the manner a rider conduct a maneuver in the beginning or during a critical riding situation. Moreover, as the available dataset did not encompass any crash, the deviations from the mean riding style may not characterize direct accident risk. The situations revealed may be considered as potential situations with increased probability of having a "near-miss".

Uncovering Actions Influencing the Riding Profiles

A predictor is developed using Neural Network in order to model the relationship of the mechanical related parameters (independent variables) with the deviation from the mean riding behavior as this is expressed from the distance metric of Equation 1 (dependent parameter). The Neural Network implemented is a simple and flexible Multi-Layer Perceptron (MLP) that acts as a non-linear regressor; the proposed model has been systematically implemented with success in various transportation problems (Karlaftis and Vlahogianni 2010).

The MLP consists of three layers (input layer, hidden layer and output layer) and is trained using the back-propagation algorithm. To deal with several optimization issues, such as those referring to structure and learning rate (Vlahogianni et al. 2005) genetic algorithms (GAs) are implemented. GAs are stochastic search algorithms that search over a population of possible solutions but differ from the most commonly encountered optimization methods in that they work with a coding of the parameter set, not the parameters themselves, they use payoff (objective function) information and are based on use probabilistic transition rules (Goldberg 1989).

In the specific paper, genetic optimization is applied to the structure (number of hidden units) and the learning process (values of learning rate and momentum in each layer) of the developed MLP. Data is divided into three sets: the training set, the cross-validation set – to validate the training - and the test set – to validate the models' ability to generalize; all results reported source from the test set. The specifications of the model are seen on Table 2.

Table 2: MLP input, structural, learning and optimization specifications.

Parameter		**Value**
Input space	Training (60%) Cross-Validation (15%) Testing (25%)	Braking activation (Front, Rear), Steering, Throttle, Wheel Speed
Architecture	Hidden Layer	GA optimized
	Activation	*tanh*
	Output space	1
Learning rule		Back-propagation
Genetic Algorithm	Population	25
	Generation	50
	Selection	roulette
	Crossover probability (p_c)	uniform:0.9
	Mutation probability (p_m)	0.02
	Fitness Function	Mean square error (cross-validation set)

As seen in Figure 4, although the model seems to underestimate the very extreme deviations from the regular riding behavior, the MLP model fits well the data (r=0.97). The proposed regressor can be, thus, used to further study the manner each input influence the deviation (Mahalanobis distance d_i) from the regular riding style.

A sensitivity analysis is conducted in order to estimate the rate of change in the output of a model with respect to changes in the model inputs. Such information will foster the proper evaluation of the model, while help determining parameters for which it is important to have more accurate values; overall a better understanding of the behavior of the system under consideration may be achieved.

The central idea is to apply a mechanism of random perturbation around its mean value in one input variable at a time and, then, while keeping the other in their mean values, measure the change of the output (Principe *et al.* 1999). The change in the input is normally done by adding a random value of a known variance to each sample and computing the output. The sensitivity for input *k* is expressed as (Principe *et al.* 1999):

$$s_k = \frac{\sum_{p=1}^{P}\sum_{i=1}^{o}(y_{ip} - \overline{y}_{ip})^2}{\sigma_\kappa^2} \tag{5}$$

Where $\overline{y}_{ip}$ is the i^{th} output obtained with the fixed weights for the p^{th} riding input pattern – a vector of riding parameters, o is the number of outputs, P is the number of patterns, and σ_κ^2 is the variance of the input perturbation. In general, inputs that have large sensitivity are more crucial in the mapping.

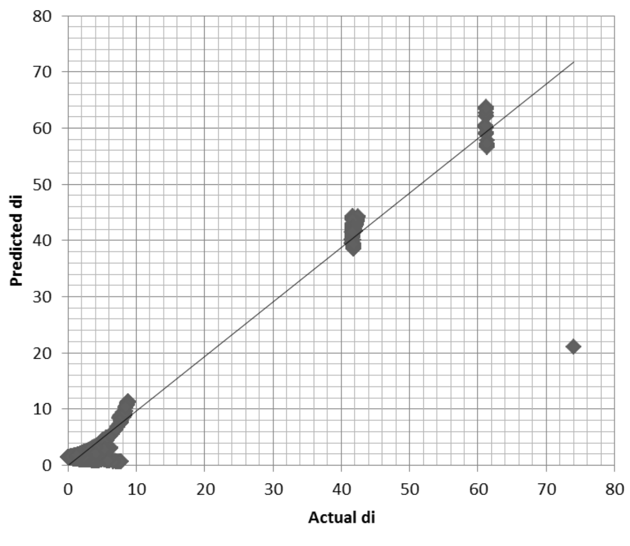

Figure 4: Scatter plot of the actual versus the predicted Mahalanobis distance di.

Results are seen in Figure 5 where the mean values of the standard deviation of the Mahalanobis distance d_i for each varied input are plotted.

As can be observed, the riding profile of the specific rider that conducts trips of 20 minute duration in rural two-lane highways is dominated by braking activity. Wheel speed is the less critical parameter for describing the deviation from the mean riding style of the specific rider. Regarding the rest of the parameters, it is worth noting that throttle seems to be more influential than steering.

Evidently, the level of influence of each riding parameter to the knowledge of the deviation from the mean riding behavior is closely related to the manner a rider reacts to specific stimuli and, thus, it may not claim transferability to either different riders, or different types of trips (e.g. urban areas, adverse weather conditions and so on).

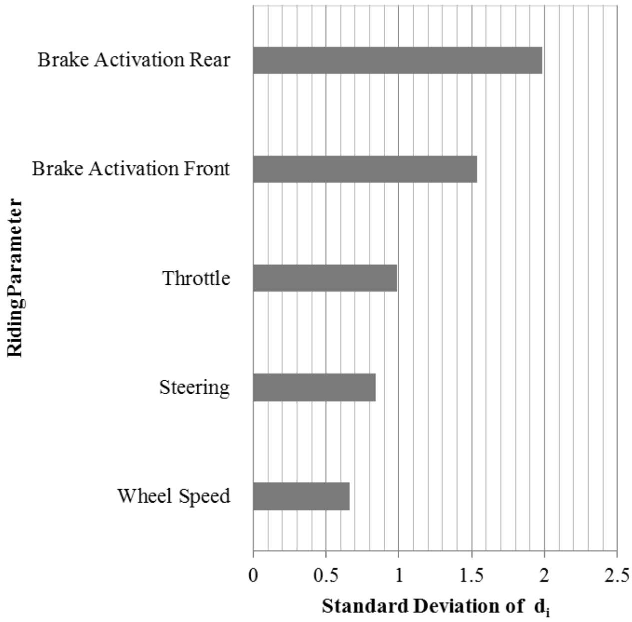

Figure 5: The mean standard deviation of the Mahalanobis distance d_i for each varied input.

Moreover, the specific approach, that does not take into consideration the riding behavior diversities hindering in the different types of incidents, describes solely the influence of each parameter on average. The influence of each parameter may significantly differ between two different riding incidents e.g. a critical overtaking and a situation of maneuvering to avoid a fixed object. Nevertheless, the transferability of the methodology results are beyond the scope of the present paper that is to provide a flexible and simple statistical approach to assist the detection of critical riding situations that may encompass increase accident risk.

CONCLUSIONS

PTW riders are a critical group of road users frequently engaging to high risk riding situations. During risky conditions, the PTW riders often alter their behavior from a regular riding pattern to an irregular chain of riding actions by braking, changing speed, maneuvering and so on, or combinations of the above. However, both the actual and perceived thresholds of regular and irregular riding behavior differ among riders. To address this issue, the present paper presents a methodology to provide custom-made riding profiles by automatically detecting from collected riding parameters deviations from the regular riding behavior by using high resolution naturalistic riding data, without need for observations from the videos recorded. The detected irregularities are those that diverge from the centroid of the jointly considered riding variables and define critical riding situations that may further be associated to typical riding events.

The proposed modeling approach is applied to naturalistic riding trips of approximately 20 minutes duration made during a period of three months by one rider in rural two-way roads. A preliminary multi-resolution analysis showed that the riding patterns resulting from the consideration of different riding parameters may significantly differ in both the frequency and the intensity of the observed deviations from the regular behavior. Results from the application of the proposed methodology indicate that the joint consideration of variables that are directly connected to the mechanical characteristics of the PTW, such as breaking, wheel speed, throttle and steering, are adequate to distinguish the regular from irregular riding behavior.

The further evaluation on the observed irregular riding patterns led to the identification of a set of riding situations, in which riders actions may be considered to deviate from the mean riding style. The detected set of incidents may vary with regards to the accident risk they may encompass. Moreover, an example on the manner custom-made riding profiles may be uncovered using the specific approach is provided. A neural network regressor was developed to approximate the relationships between the mechanical related riding parameters and the deviation from the regular riding actions. For the specific rider and trips of 20 minute duration in rural two-lane highways, this regression revealed the influence of each input parameter to the manner the rider reacts to external or internal stimuli.

From a conceptual perspective, the proposed modeling although being capable to define custom-made riding profiles, may rank the observed critical riding incidents with regards to the risk they encompass - as potential near-misses - only on the basis of the magnitude of the deviation from the mean riding style. The specific approach is purely data-driven and, thus, cannot claim transferability between different riders, road types, trip characteristics, or even the underlying riding behavior diversities that may vary with respect to the different types of incidents. Further research therefore is needed for uncovering the determinants of each riding behavior with respect to the manner a rider reacts in the beginning or during a critical riding situation.

ACKNOWLEDGMENTS

Data used in this paper is a part of the work conducted during the focused research collaborative project “2-BE-SAFE – 2-Wheeler Behavior and Safety” co-funded by the European Commission under the Seventh Framework R&TD Programme (www.2besafe.eu).

REFERENCES

Baldanzini, N., Hurth, V., Regan, M., Spyropoulou, I., Eliou, N., Lemonakis, P and Galanis, P. (2009). Literature review of data analysis for naturalistic driving study, deliverable 4, December, 2BESAFE project, European Commission, Belgium.

Barnett, V. & Lewis, T. (1994). Outliers in Statistical Data, 3rd edn. John Wiley & Sons.

Billor, N., Hadi, A. S., and Velleman, P. F. (2000). BACON: Blocked Adaptive Computationally-Efficient Outlier Nominators; Computational Statistics and Data Analysis 34, 279–298.

Csepinszky, A. and Benmimoun, M. (2010). An operational perspective on the organisation of large scale field operational tests of intelligent vehicles, András Csepinszky, ERTICO - ITS Europe, ITS World Congress, Busan, 25-29 October .

Dupont E., Martensen H., Stipdonk H., Bijleveld F., Commandeur J., Papadimitriou E., Yannis G., Antoniou C, Bergel R., Brandstaetter C., (2009). Analyses using the European Road Safety Observatory, Proceedings of the 4th IRTAD Conference, Seoul, September 2009.

Engstrom J., Johansson E., Ostlund J., (2005). Effects of visual and cognitive load in real and simulated motorway driving, Transportation Research Part F: Traffic Psychology and Behaviour, 8(2), 97-120.

FESTA (2008), Field opErational teSt supporT Action (FESTA) - D2.4 Data analysis and modelling, FESTA project, European Commission, Belgium.

Goldberg, D.E. (1989) Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley. ISBN 0201157675.

Haque M.D. M., Chin H. C. and Huang H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections, Accident Analysis & Prevention, 42(1), 203-212.

Hodge V. J. and Austin J. (2004). Survey of Outlier Detection Methodologies, Artificial Intelligence Review 22, 85–126.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward nets, Neural Networks, 4, 231–242.

Huang H. and Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety, Accident Analysis & Prevention, 42(6), 1556-1565.

Karlaftis, M. G. and Vlahogianni, E. I. (2011). Statistics versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights, Submitted to Transportation Research Part C: Emerging Technologies, 19(3), 387-399.

Klauer S. G., Guo F., Sudweeks, J. and Dingus T. A. (2010). An Analysis of Driver Inattention Using a Case-Crossover Approach 100-Car Data: Final Report Final Report, May, DTNH22-00-C-07007, National Highway Traffic Safety Admin. (NHTSA) Washington, DC.

NHTSA (2006), The 100-Car Naturalistic Driving Study Phase II – Results of the 100-Car Field Experiment, www.nhtsa.dot.gov, USA.

Olson, R.L., Hanowski, R.J., Hickman, J.S., & Bocanegra, J. (2009). Driver distraction in commercial vehicle operations (Report no. FMCSA-RRR-09-042). www.fmcsa.dot.gov/facts-research/research-technology/report/FMCSARRR- 09-042.pdf Washington, DC: U.S. Department of Transportation, Federal Motor Carrier Safety Administration.

Principe J.C., Euliano N. R., Lefebvre C. W., (1999). “Neural And Adaptive Systems: Fondamentals Through Simulations”. John Wiley and Sons, Inc.

Regan, M., Triggs, T., Young, K., Tomasevic, N., Mitsopoulos, E., Stephan, K. and Tingvall, C. (2006) On-road evaluation of Intelligent Speed Adaptation, Following Distance Warning and Seatbelt Reminder Systems: final results of the TAC SafeCar project, Monash University Accident Research Centre, Report No. 253.

Savolainen, P. and Mannering F. (2007). Probabilistic models of motorcyclists’ injury severities in single- and multi-vehicle crashes, Accident Analysis and Prevention 39, 955–963.

Schreiber, T., (1999). Interdisciplinary application of nonlinear time series methods. Physics Reports 308, 1–64.

SHRP2 (2011). A Foundation for Safer Driving, April, Transportation Research Board, Washington DC US.

Thomas P., Morris A., Yannis G., Lejeune P., Wesemann P., Vallet G., Vanlaar W., (2005). Designing the European Road Safety Observatory, International Journal of Injury Control and Safety Promotion, Vol. 12, No. 4, pp. 251-253.

Vlahogianni, E. I., Karlaftis, M. G, and Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach, Transportation Research Part C: Emerging Technologies, 13(3), 211-234.

Washington S. P., Karlaftis M. G., Mannering F. L. (2010). Statistical and Econometric Methods for Transportation Data Analysis, Second Edition, Chapman & Hall / CRC Press, Boca Raton, FL.

Yannis G., Papadimitriou E., Dupont E., Martensen H., (2010). Estimation of fatality and injury risk by means of in-depth fatal accident investigation data, Traffic Injury Prevention, 11(5), 492-502.

Yannis, G. Golias, J. Papadimitriou, E. (2005). Driver age and vehicle engine size effects on fault and severity in young motorcyclists accidents, Accident Analysis and Prevention, 33, 327-333.