Using statistical models to characterize eco-driving style with an aggregated indicator

Cindie Andrieu and Guillaume Saint Pierre
IFSTTAR
Vehicle-Infrastructure-Driver Interactions Research Unit
14, route de la minière
78000 Versailles-Satory, France
Email: cindie.andrieu@ifsttar.fr and guillaume.saintpierre@ifsttar.fr

***Abstract*— This paper presents the construction of an aggregated indicator of a fuel-efficient driving style, in order to construct an efficient Ecological Driving Assistance System (EDAS). Such an eco-index can be used to detect eco-driving behaviour, but also to give to the driver useful advices to help him improving his driving efficiency without deteriorating safety. The logistic regression is used to model our experimental dataset of twenty subjects driving twice the same route: normally or following the golden rules of eco-driving. Depending on some driving indicators, the estimated probability of being an eco-driver is used as an eco-index to characterize that driving pattern. This work show how such a simple aggregated indicator, related to driving dynamics rather than fuel consumption, can be useful for driver monitoring and information. Two models, from the simplest to the most complicated, are compared, and their performances analysed.**

Keywords: Eco-driving, EDAS, Driving behaviour, Logistic regression.

I. Introduction

Speed management is a preoccupation for public authority as speed is considered as the main cause of traffic injury accidents. Since the climate change evidence, many attempts are made to merge traditional speed management methods and greenhouse gas emission reduction techniques, optimizing both safety and fuel consumption. Among the most promising solutions to solve such a challenge, the ecological way to drive usually referred as the eco-driving style appears to be one of the best.

The characteristics of eco-driving are generally well defined and easily characterized (see for example [1] and [2]) even if eco-driving rules are slightly different among countries [3]. The advantages of eco-driving, of course, go beyond CO2 reductions [4]. They include reducing the cost of driving to the individual and producing tangible and well-known safety benefits (with fewer accidents and traffic fatalities) [1]. Disadvantages, are little public understanding of the nature of eco-driving, and seemingly ingrained driving habits. According to [3], some potential sources of driving conflicts exist if the eco-driving rules are misinterpreted.

Helping the driver to choose the best compromise between safety and CO2 reduction driving techniques is the goal of a new type of advanced systems called ecological driving assistance systems (EDAS). Even GPS devices or smartphones applications are sometimes providing a dynamic fuel efficiency indicator, while some software are specifically dedicated to eco-driving and are able to deliver tips to help decrease fuel consumption. Most of these devices (embedded or not) are using miles per gallon (or liters for 100km in Europe) as displayed parameters, while some of them use more sophisticated approach and compute a global indicator. According to expert's knowledge (interview with eco-driving professionals), displaying instant fuel use, or battery gauge, is not sufficient to help the drivers in understanding the dynamic relationship between driving actions and fuel efficiency. As most of the people want to keep ecological driving assistance systems (EDAS) simple (see for example [5]), we believe that a global indicator, merging different driving parameters can be more efficient than fuel consumption.

The aim of this study is to provide a methodology suitable to compute an aggregated eco-driving indicator based on statistical models estimated using naturalistic driving data, and evaluated with an experimental study. Four indicators were chosen, each associated with one of the main rules of eco-driving. After a discussion about various suitable statistical models, we demonstrate the interest of using logistic regression for model based estimation of driving fuel efficiency. A second model has been developed in order to allow its implementation on smartphones not connected to the vehicle.

II. METHODOLOGY

A. Experimental design

The experiment goal was to clearly identify two classes of driving behavior on the same test track: "normal" and fuel efficient way to drive commonly known as "eco-driving". Twenty drivers participated in this experiment that took place in June and July 2009 in Ponchartrain (Yvelines) in France. In order to minimize traffic influence, the chosen route is of inter-urban type and a length of 14km. The trips were all performed under free flow conditions and with dry weather. The vehicle used was a petrol-driven Renault Clio III with manual gearshift. First of all, the journey is discovered by the subjects while seeing the experimenter driving and giving safety and direction instructions. Then, the trip was driven twice by each driver: once while driving normally, and secondly while following the "Golden Rules" of eco-driving extracted from the Ecodrive project [2] and summarized in Table I. These rules were given just before the ecological trip. To eliminate a learning effect of the journey, trip's order has been counter-balanced. An on-board logging device was used to monitor key driving parameters. The device is connected to the controller area network (CAN) of the vehicle, logging most of the relevant parameters related to engine state, vehicle dynamic, and driver actions on pedals. The vehicle has been also equipped with a GPS, a camera in front of the vehicle and a fuel flow meter. We used a fuel flow meter DFL1x-5bar to validate the fuel consumption logged with the CAN. Additional variables were post-processed such speed limits, gear ratio, and many indicators inspired from [6].

B. Selection of indicators associated with each of the main rules of eco-driving

Driving style to reduce fuel consumption is related to the implementation of the four main eco-driving rules set out in Table I. Due to this link, each of these instructions was associated with an indicator. The proposed indicators are summarized in Table I. So the first rule state to shift up early. Therefore, it is natural to associate the indicator *AvgRPMShiftUp* which is the average engine speed (in rpm) at the shift into a higher gear. The second rule is related both to the gear and the engine speed. So we created an indicator, called *IndexGear-RPM*, summarizing these two variables and calculated as follows:

$$\begin{aligned}&IndexGearRPM = \\&\tfrac{1}{3500}(TimeNeut \times AvgRPMNeut \\&\quad + TimeGear1 \times AvgRPMGear1 + \ldots \\&\quad + TimeGear5 \times AvgRPMGear5)\end{aligned}$$

where *TimeNeut* is the percentage time in neutral gear, *AvgRPMNeut* is the average engine speed (in rpm) in neutral gear, *TimeGear1* is the percentage time in gear 1 (with pressing the accelerator pedal), etc. Note that the condition of pressing the accelerator pedal ensure to ignore the time in engine brake which is associated to the fourth rule. Note also that the division by 3500, representing the maximum engine speed, is just a normalization factor. Then the third rule related to the anticipation of traffic is associated to the parameter *PKE* (Positive Kinetic Energy) calculated as follows:

$$PKE = \frac{\sum(v_f^2 - v_i^2)}{x} \text{ when } \frac{dv}{dt} > 0$$

where v_f and v_i are respectively the final and the initial speed (in m/s) at each time interval for which $\frac{dv}{dt} > 0$, and x is the total distance traveled (in m). This indicator represents the ability to keep the vehicle's kinetic energy as low as possible. So a nervous driving will be associated with a high *PKE*, and conversely a smoothly driving will be associated with a *PKE* close to zero. Finally, the fourth rule is naturally associated with the percentage of time in engine brake characterized by the following conditions: non zero speed, no neutral, no pressure the brake pedal and the accelerator pedal.

C. Statistical models

The objective of this study is to construct an aggregated indicator of an eco-driving style. Our approach relies on developing a predictive model of economic driving behavior based on easily interpretable variables. Assuming trips are clustered according to the two driving conditions, it is worth trying a statistically based approach to predict the driving style. Such models are well suited in estimating the relationship between an outcome variable and a set of explanatory variables. In this paper, the outcome variable is from a binary distribution with two possible values:
For $i = 1, \dots, I$ and $j = 1, \dots, T_i$,

$$Y_{ij} = \begin{cases} 1 \; if \; ecodriving \\ 0 \; if \; not \end{cases}$$

where I is the number of drivers and T_i is the number of observations for the driver *i*. Logistic regression is a form of statistical modeling that is often appropriate for independant binary outcome variables. Assume Y_{ij} follows a Bernoulli distribution with parameter $p_{ij} = P(Y_{ij} = 1)$ where p_{ij} represent the probability that the event occurred for the observation Y_{ij}. The relationship between the event probability p_{ij} and the set of factors is modeled through a logit link function with the following form:

$$logit(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = X'_{ij}\beta$$

where X_{ij} is the vector of explanatory variables and β is the vector of regression parameters [7]. The ordinary logistic regression assumes independent observation and the vector β is estimated by the method of maximum likelihood. However, the assumption of data independence does not suit our data very well, as it will contain unavoidable driver-specific correlations (i.e. observations from the same driver are assumed to be correlated) that should be treated as random effects. The standard errors from the ordinary logistic regression are then biased because the independence assumption is violated.

To account for these driver specific correlations as random effects, more sophisticated statistical models need to be applied. These models are particularly useful for naturalistic driving study [8] and specially event based approach (EBA) which basic principle is to identify time segments that can be predictive of an event (e.g. crash, near-crash, ...). Indeed, these models include additional parameters to deal with correlations, and confounding factors are viewed as explicative variables that can be used to predict event probability. One such model is the "Generalized Estimated Equations" (GEE) model or marginal models, originally developed to model longitudinal data by Liang and Zeger [9], which assumes that observations are marginally correlated. Another approach for modeling correlated data is "Generalized Linear Mixed Models" (GLMM). The GLMM model introduces a random effect specific to each subject whereas the GEE approach models the marginal distributions by treating correlation as a nuisance parameter. Therefore the inference is individual (subject-specific approach) in contrast to marginal models that model the average population (population-averaged approach). However, in our study, we didn't use these two sophisticated statistical models because of the small sample size (see Section III-B for more details). So we used only ordinary logistic regression models.

III. Results

A. Overall effects of eco-driving rules

Numeric results are summarized in Table II. A paired *t*-test was performed to assess whether the mean of each parameter differ significantly according to the driving style. Table II indicates the p-values of these tests. Among the most interesting ones, the average fuel consumption across drivers decreased by 12.5% between normal driving and eco-driving; this fall being of 26% for some drivers. These results show that it seems quite simple to reduce fuel consumption by applying some basic rules of eco-driving. The average speed decreased by 5.8% and the percentage of time beyond the legal speed limit decreased by 30.1%. These reductions reflect a better

Instruction	Indicator	Abbreviation
1. Shift up as soon as possible: Shift up between 2.000 and 2.500 revolutions per minute.	Average engine speed at the shift into a higher gear.	AvgRPMShiftUp
2. Maintain a steady speed: Use the highest gear possible and drive with low engine RPM.	Index of gear ratio distribution and engine speed associated.	IndexGearRPM
3. Anticipate traffic flow: Look ahead as far as possible and anticipate the surrounding traffic.	Positive Kinetic Energy.	PKE
4. Decelerate Smoothly: When you have to slow down or to stop, decelerate smoothly by releasing the accelerator in time, leaving the car in gear.	Percentage of time in engine brake.	TimeEngineBrake

TABLE I

MAIN RULES OF ECO-DRIVING AND INDICATORS ASSOCIATED

compliance with speed limits with economical driving. As regards the application of eco-driving rules, the four associated indicators are significantly different among the two driving conditions, indicating that the instructions were applied. Furthermore, the average acceleration and deceleration both decrease significantly which is in agreement with the second and the third rules of eco-driving.

B. Construction of an eco-index based on the main rules of eco-driving

The aim of this work is the development of a predictive model of economic driving behavior based on easily interpretable variables excluding the variable on fuel consumption. Indeed, fuel consumption is closely related to road type (urban, interurban, motorway) and traffic, but it cannot be considered itself as an eco-driving indicator. It is obvious as even a very efficient driver will have a high fuel consumption when driving under congestion or on hilly roads. An efficient indicator of eco-driving should not depend too much on such external conditions and rely more on driver actions. Thus, we constructed a predictive model of the probability of being in an eco-driving situation using a binomial logistic regression model with the four indicators in Table II as explanatory variables. According to our experiment, we predict the binary variable named "Trip" which takes the value 0 in normal driving (noted "normal") and 1 in eco-driving (noted "eco").

In our experiment, both the number of clusters (20) and the cluster size (2) are small. These constraints do not allow us to use the appropriate statistical models taking into account driver specific correlations. Thus, Ziegler et al. [10] recommend an application of the GEE only, if the number of clusters is at least 30 for a cluster size of about 4 for a low to moderate correlation. Moreover, several studies (e.g. [11]) have shown that parameters estimates are biased with both fixed or random effects logistic models when the number of clusters and the cluster size are small. However these studies show that the estimates of the random intercept and random slope have larger biases compared to the fixed effect parameters. Thus, later in this paper, we use an ordinary logistic regression. The estimated logistic model is the following:

$$\begin{aligned} &logit\left[P\left(Trip = Eco\right)\right] = 8.967 \\ &-0.007 \times X_1 + 0.242 \times X_2 \\ &-31.684 \times X_3 + 0.148 \times X_4 \end{aligned} \tag{1}$$

where $X_1, \ldots, X_4$ are the four indicators associated respectively with the four instructions of eco-driving (Table I). The usefulness of the model is measured by the Nagelkerke R^2, denoted R^2_N, which is an adjusted version of the Cox & Snell R^2 and which is similar to the coefficient of determination R^2 in linear regression. This parameter does not measure the goodness of fit of the model but indicate how useful the explanatory variables are in predicting the response variable. The model 1 reached a total $R^2_N = 0.74$, with a strong influence of the variable *PKE*, leading to increase the probability of being in a situation of eco-driving. Using a decision rule's cutoff value of 0.5, the model correctly classified 85% of true positives ("normal") and 80% of true negatives ("eco") even though this

Parameter	Description	Mean "Normal"	Mean "Eco"	Variation (%)
Avg_Fuel_Consum	Average fuel consumption (l/100km).	6.86	6.00	-12.5^{***}
Avg_RPM_Shift_Up	Average engine speed at the shift into a higher gear (associated with rule 1).	2737.5	2232.8	-18.4^{***}
Index_Gear_RPM	Index of gear ratio distribution and engine speed associated (associated with rule 2).	61.0	52.9	-13.3^{***}
PKE	Positive Kinetic Energy (associated with rule 3).	0.343	0.243	-29.2^{***}
Time_Engine_Brake	Percentage of time in engine brake (associated with rule 4).	20.3	26.3	$+29.6^{**}$
Avg_Speed	Average speed	50.85	47.89	-5.8^{**}
Avg_Accel	Average acceleration	0.498	0.387	-22.3^{***}
Avg_Decel	Average deceleration	-0.619	-0.523	-15.5^{***}
Avg_RPM	Average engine speed	2097.4	1835.5	-12.5^{***}
Time_NonLegal_Speed	Percentage of time beyond the legal speed limit	37.9	26.5	-30.1^{***}

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE II

EFFECTS OF ECO-DRIVING RULES ON DIFFERENT PARAMETERS

classification results from using all observations to fit the model, which can bias the results. For pedagogical purposes, we call ”eco-index” of the observed trip, the model output probability $P\left(Trip = Eco\right)$ multiplied by one hundred. So we obtain an index of eco-driving which varies between 0 and 100 for easier interpretation. One of the main objectives of eco-driving is to reduce fuel consumption; evaluating the performance of such an eco-index can be done by studying the relationship strength between our eco-index and the average fuel consumption. We conducted a linear regression between these two parameters for all 40 trips from our experiment. This model reached a total coefficient of determination $R^2 = 0.70$, which shows that our eco-index is closely related to the average fuel consumption.

C. Construction of an eco-index based on a simple indicator: eco-index for smartphone

In this section, we use the same method as in the previous section to build a new model of eco-index based only on the parameter *PKE*. Indeed, on the one hand, we observed that this parameter had a strong influence with the probability of being in an eco-driving situation. On the other hand, the advantage of this parameter is its easiness to be calculated since it depends only on speed. Thus, a model of eco-index based solely on *PKE* can be implemented easily on a smartphone. The logistic model is the following:

$$\begin{aligned} &logit\left[P\left(Trip = Eco\right)\right] = 9.773 \\ &- 34.089 \times PKE \end{aligned} \tag{2}$$

This model reached a total Nagelkerke $R^2_N = 0.59$ and correctly classified 80% of true positives ("normal") and 85% of true negatives ("eco"). The linear regression between the eco-index derived from the model 2 and the average fuel consumption has a coefficient of determination $R^2 = 0.69$. This simple model has good features, similar to the complete model.

D. Factorial analysis

A principal component analysis (PCA) was performed with the forty original trips using the four indicators based on the main rules of eco-driving. The first factorial plan with the value of the eco-index related to the model 1, distinguishing "normal" and "eco" trips, is represented in Fig. 1.

The first axis is correlated with the three first indicators defined in Table 1, and the second axis is correlated with the fourth indicator *TimeEngineBrake*. We observe that the two driving styles are well discriminated by the four indicators. Moreover, these results confirm that the eco-index is a well eco-driving indicator since "eco" trips

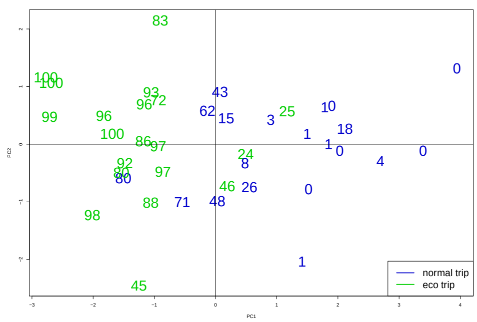

Fig. 1. Principal component analysis and eco-index (blue for "normal" trips and green for "eco" trips).

are associated with high eco-index whereas "normal" trips are associated with lower eco-index.

IV. Conclusion

This study provides a methodology suitable to compute a global eco-driving indicator based on statistical models, taking into account various behavior related parameters. Two logistic regression models of this eco-index, from the simplest to the most complicated, have been developed: the first one is based on the four performance indicators associated with each of the main rules of eco-driving, the second one is based only on the variable PKE. The first model provides the most appropriate information to be displayed in a future ecological driving assistance system (EDAS). Indeed, each performance indicator being associated with a rule of eco-driving, it is possible to display quantitative feedback to the driver, specifically for each one of the four main rules of eco-driving. The second model based on PKE has the advantage of being easily calculated and therefore suitable for nomadic devices implementation.

Assuming that eco-driving behavior depends on the road conditions, we have extended the full model 1 to a more complex model taking into account the speed limit as a stratification variable. This third model, not introduced in this paper, improves the properties of the full model and allows to inform the driver on the network categories (urban, rural, ...) on which he can improve his efficiently driving. However, this model needs the knowledge of speed limits for the traveled route.

Other statistical models taking into account driver specific correlations, namely GEE and mixed models, have been mentioned but we could not implement them because of the small sample size of our experiment. However, it might be interesting to test bootstrap methods suitable if the number of clusters is small, as discussed in [12].
Future works will focus on the validation of the two logistic models presented in this paper, and on the development of a dynamic eco-index providing information to the driver during the trip and allowing self-evaluation throughout the journey.

References

[1] J. Barkenbus, “Eco-driving: An overlooked climate change initiative,” *Energy Policy*, vol. 38, no. 2, pp. 762–769, 2010.
[2] “Ecodrive project,” 2009. [Online]. Available: http://www.ecodrive.org
[3] CIECA (International commission for driver testing authorities), *CIECA internal project on 'Eco-driving' in category B driver training and the driving test*. CIECA, 2007. [Online]. Available: http://www.cieca.be/
[4] R. Vermeulen, “The effects of a range of measures to reduce the tail pipe emissions and/or the fuel consumption of modern passenger cars on petrol and diesel.” TNO science and industry, Tech. Rep., 2006.
[5] M. S. Young, S. A. Birrell, , and N. A. Stanton, *Lecture Notes in Computer Science: Engineering Psychology and Cognitive Ergonomics*. Springer Berlin / Heidelberg, 2009, ch. Design for Smart Driving: A Tale of Two Interfaces, pp. 477–485.
[6] E. Ericsson, “Independent driving pattern factors and their influence on fuel-use and exhaust emission factors,” *Transportation Research Part D: Transport and Environment*, vol. 6, pp. 325–345, 2001.
[7] A. Agresti, *Categorical data analysis. 2nd ed.* Chichester: Wiley Series in Probability and Mathematical Statistics., 2002.
[8] F. Guo and J. Hankey, “Modeling 100-car safety events: A case-based approach for analyzing naturalistic driving data.” Report No. 09-UT-006, Virginia Tech Transportation Institute, Tech. Rep., 2009.
[9] K. Y. Liang and S. L. Zeger, “Longitudinal data analysis using generalized linear models,” *Biometrika*, vol. 73, pp. 13–22, 1986.
[10] A. Ziegler, C. Kastner, and M. Blettner, “The generalised estimating equations : an annotated bibliography,” *Biometrical Journal*, vol. 40 (2), pp. 115–139, 1998.
[11] R. Moineddin, F. I. Matheson, and R. H. Glazier, “A simulation study of sample size for multilevel logistic regression models,” *BMC Medical Research Methodology*, p. 7:34, 2007.
[12] L. H. Moulton and S. L. Zeger, “Analyzing repeated measures on generalized linear models via the bootstrap,” *Biometrics*, vol. 45, pp. 381–394, 1989.