EJTIR

Issue 14(3), 2014
pp. 201-213
ISSN: 1567-7141
www.ejtir.tbm.tudelft.nl

Evaluation of ecodriving performances and teaching method: comparing training and simple advice

Cindie Andrieu[1], Guillaume Saint Pierre[2]
IFSTTAR, Versailles-Satory, France

Eco-driving style is widely known to induce up to 20% fuel consumption reduction, but little is known about differences due to different learning methods. In order to evaluate the potential impacts of future ecological driving assistance system (EDAS) in comparison with usual techniques, a statistical approach is proposed. Two kinds of experiments are analysed in this paper: In the first one, simple advice were given to the participants, while in the second one, full courses with eco-driving experts were used. Performance indicators were derived from five commonly referred golden rules of eco-driving and used as model inputs. Different kind of statistical models are discussed, among which we choose to apply the ordinary logistic regression to assess the effects of each driving advice separately Results show that ecodriving advices are better applied after a course than just providing tips. The approach is then extended to build a generic model that can be used both to characterize and evaluate eco-driving style.

Keywords: *Eco-driving, logistic regression, driving behaviour, driving evaluation, ecological driving assistance system.*

1. Introduction

Driving more efficiently is part of the solution to reduce the surface transportation greenhouse gas emissions but it is a highly complex task, comprising over hundreds of separate tasks (Walker et al., 2001). Drivers need to simultaneously control the vehicle, adjust their speed and trajectory according to driving environment, deal with hazards, and make strategic decisions such as navigation to progress toward their goal (Young et al., 2010). Since climate change and humanity responsibility has been widely accepted, many drivers have a new goal in mind: fuel efficiency. Eco-driving style is therefore often referred as smart driving because of the necessary complex trade-off between the multiple goals the driver has to manage with. Studies usually simplify the green way to drive using simple advice easily understood by drivers (CIECA, 2007), but sometimes leading to a misunderstanding of the fuel efficient driving strategy. Other studies use trial experiments before and after a training program to assess the eco-driving impact (Symmons et al., 2009). Effects of eco-driving on fuel consumption are well described in the literature, but results are often optimistic: CO_2 emissions reduction can be up to 30% according to many studies. The key question for policy makers is "how big" of an emission reduction we can get by encouraging an eco-driving style, taking into account the diversity in the way to learn eco-driving: just reading a few driving tips, taking a course with a professional, or doing practical exercises with equipped vehicles? Moreover, there is a need to understand the best way to teach and learn eco-driving style, especially for young drivers. Indeed, as stated by the CIECA report

[1] A: IFSTTAR, COSYS, LIVIC, 77, Rue des Chantiers, 78000 Versailles, France; T: +33 130 844 017 E: cindie.andrieu@ifsttar.fr; cindie.andrieu@gmail.com
[2] A: IFSTTAR, COSYS, LIVIC, 77, Rue des Chantiers, 78000 Versailles, France; T : +33 130 844 017 E : guillaume.saintpierre@ifsttar.fr

(2007), " the earlier 'the seed' is planted in the minds and experience of road users, the greater the potential benefit for reducing the impact of transport on the environment".

Beside this, a new type of Advanced Driving Assistance Systems (ADAS) recently appear with specific features devoted to fuel savings. Few of them even try to induce an efficient or "eco" driving style by incorporating high-level advice which goals are to produce a suitable compromise between safety and CO_2 emissions reduction and avoid hypermiling behaviour (Chapnick, (2007)). As most of the people want to keep ecological driving assistance systems (EDAS) simple (see for example Young et al. (2009)), we believe that a global indicator, merging different driving parameters can be more efficient to reflect ecodriving efficiency than simply looking at the instant fuel consumption. Such a global indicator would therefore be suitable for a feedback signal.

The aim of this study is to provide statistical models suitable to evaluate both training methods and EDAS abilities to produce a green driving behaviour. The main constraint being to stay compatible with a widely accepted eco-driving definition. The main material consist of two different data sets, one with subjects following simple eco-driving advice (leaflet describing the 5 golden rules of ecodriving, see table 1.), the other with subjects driving the way they learned in a course with professional eco-drivers (1h duration practical training). In order to define clearly which driving behaviour is expected by the professionals, eco-driving style is summarized into four different simple advice given in Table 1, each one of them being associated to a quantitative indicator build to reflect the associated driving behaviour. Different kinds of statistical models are discussed, among which we choose to apply the ordinary logistic regression to assess the effects of each driving advice separately. The significance of the differences for each indicator between normal and eco-driving trips allow us to evaluate which advice is practiced by the drivers, according to the way they were trained to eco-driving. The same analysis is done for each different speed limit zone to take into account the effects of the driving environment.

Then, the previous results are extended to build an aggregated model incorporating all parameters effects into a single eco-driving indicator. We prove usefulness of this approach as a tool for building innovative EDAS able to provide efficient feedback to drivers.

2. The experiments

2.1 Experiment 1: simple advice on eco-driving

The experiment goal was to clearly identify two classes of driving behaviour 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. Four of these drivers were eco-driving instructors while others were recruited among one thousand persons working in two different research institutes. 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 (Ecodrive, 2009) and summarized in Table 1. These rules were given in a written form to be read by the drivers 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 some driving pattern parameters extracted from Ericsson (2001) such as distributions of speed, acceleration/deceleration and engine speed, and more complex parameters like RPA (Relative Positive Acceleration) and PKE (Positive Kinetic Energy).

2.2 Experiment 2: eco-driving training

Nineteen drivers (who have not participated in the experiment 1) participated in this experiment that took place near Toulouse in 2004. The trials goal was to evaluate the effect of an embedded EDAS produced by the GERICO project funded by the French program of research, experimentation and innovation in land transport (Barbé et al., 2008). The original design was to compare a control group, a group applying eco-driving, and another group using the system without any advice. For the purpose of this study, the data collected for the group using the ecodriving assistance system have not been used. Analysis is done by comparing results between the control group and the group trained with professionals. The chosen route contains various network categories (urban, rural, motorway) and has a length of 70km. The vehicle used was a Renault Megane Scenic with a four-speed sequential gearbox. The trip was driven twice by each driver: once while driving normally and secondly after an eco-driving training with professional eco-drivers. The training included a theoretical course about fuel savings followed by a practical trip under the supervision of the teachers. In this case, trips are not counter balanced and effects of the eco-driving teaching may be overestimated because of a learning effect.

2.3 Data reduction

Previous datasets provide information related to ecodriving in a different way: experimental designs are different, cars used are different, and trips and road context are also different. It would be hazardous to merge them in a single ready to use data set, and so we present separated analyses for both experiments. In order to build a global model to predict ecodriving likelihood, only the first experiment with data collected under supervision of the authors is used for parameters estimation. Basic data reduction techniques were applied following Dozza et al. (2012) in order to split the collected trips in homogeneous driving context. We choose to search homogeneity according to the speed limit variable, this information being much more reliable than road type, or any other subjective classification. Final data sets consists of several trip sections with a single speed limit value, each section being characterized by a series of performance indicators.

3. Methodology

3.1 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 1. Due to this link, each of these instructions was associated with an indicator. The proposed indicators are summarized in Table 1. 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 *IndexGearRPM*, summarizing these two variables and calculated as follows:

$$IndexGearRPM = \frac{TimeNeut \cdot AvgRPMNeut + Gear1 \cdot AvgRPMGear1 + ... + Gear5 \cdot AvgRPMGear5}{3500} \quad (1)$$

where *TimeNeut* is the percentage time in neutral gear, *AvgRPMNeut* is the average engine speed in neutral gear, *Gear1* 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:

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

where v_f and v_i are respectively the final and the initial speed (in m/s) at each time interval for which $dv/dt > 0$, and x is the total distance travelled (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 on the brake and accelerator pedal.

Table 1. Main rules of eco-driving and indicators associated

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.	Avg_RPM_Shift_Up
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.	Index_Gear_RPM
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.	Time_Engine_Brake

3.2 Statistical models

The objective of this study is to build a tool to estimate EDAS efficiency. Such an approach should be able to distinguish between eco-driving after reading simple advices (experiment 1) and eco-driving after being trained by professionals (experiment 2). In order to reach that goal, we choose to develop a predictive model of environmental friendly driving behaviour based on easily interpretable variables. Assuming trips are clustered according to the two real driving conditions, our data can be exploited to train statistically based models. 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:

$$Y_{ij} = \begin{cases} 1 \, if \;\; ecodriving \\ 0 \quad if \;\; not \end{cases} \forall i = 1, \ldots, I; \, j = 1, \ldots, T_i \tag{3}$$

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 modelling that is often appropriate for binary outcome variables. Assume Y_{ij} follows a Bernoulli distribution with parameter $p_{ij} = P(Y_{ij} = 1)$ where p_{ij} represents 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:

$$logit = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = X_{ij}^{'}\beta \tag{4}$$

where X_{ij} is the vector of explanatory variables and β is the vector of regression parameters (Agresti, 2002). 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 (Guo and Hankey, 2010; Benminoun et al., 2011) 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 (1986), 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 4.2 for more details). So we used only ordinary logistic regression models.

4. Results

4.1 Overall effects of eco-driving rules and eco-driving training

Numeric results are summarized in Table 2. A paired *t*-test was performed to assess whether the mean of each parameter differ significantly according to the driving style. Table 2 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 for the experiment 1 and decreased by 11.3% for the experiment 2. These similar results between the two experiment 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% for the experiment 1 and 10.1% for the experiment 2, and the percentage of time beyond the legal speed limit decreased by 30.1% for the experiment 1 and 36.1% for the experiment 2.

Table 2. Effects of eco-driving rules on different parameters

Parameter	Description	Experiment 1			Experiment 2		
		Mean "Normal"	Mean "Eco"	Variation (%)	Mean "Normal"	Mean "Eco"	Variation (%)
AvgFuelConsum	Average fuel consumption (l/100km).	6.86	6.00	-12.5^{***}	9.01	7.99	-11.3^{***}
AvgRPMShiftUp	Average engine speed at the shift into a higher gear (associated with rule 1).	2737.5	2232.8	-18.4^{***}	3177.3	2465.6	-22.4^{***}
IndexGearRPM	Index of gear ratio distribution and engine speed associated (associated with rule 2).	61.0	52.9	-13.3^{***}	70.8	60	-15.3^{***}
PKE	Positive Kinetic Energy (associated with rule 3).	0.343	0.243	-29.2^{***}	0.293	0.197	-32.8^{***}
TimeEngineBrake	Percentage of time in engine brake (associated with rule 4).	20.3	26.3	$+29.6^{**}$	16.2	16.8	+ 0.04
AvgSpeed	Average speed (km/h)	50.85	47.89	-5.8^{**}	61.45	55.22	-10.1^{***}
AvgAccel	Average acceleration (ms^{-2})	0.498	0.387	-22.3^{***}	0.596	0.473	-20.6^{***}
AvgDecel	Average deceleration (ms^{-2})	-0.619	-0.523	-15.5^{***}	-0.672	-0.599	-10.9^{***}
AvgRPM	Average engine speed (rpm)	2097.4	1835.5	-12.5^{***}	2379.6	2009.6	-15.5^{***}
TimeNonLegalSpeed	Percentage of time beyond the legal speed limit	37.9	26.5	-30.1^{***}	28.5	18.2	-36.1^{***}

$^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$.

These reductions reflect a better compliance with speed limits with economical driving regardless of the learning mode. 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 with the two learning mode. However, in the experiment 2, the engine brake (associated with the fourth rule of eco-driving) does not seem to have been used correctly. Furthermore, the average acceleration and deceleration both decrease significantly in the two experiments which is in agreement with the second and the third rules of eco-driving.

4.2 Separated effects of the main eco-driving rules

The aim of this section is to assess the effects of each driving advice after two learning mode: one with subjects following simple eco-driving advice (experiment 1), and the other with eco-driving training (experiment 2). Our approach is to construct, for each experiment, 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 1 as explanatory variables. This means it is possible to predict the binary variable named ″Trip″ which takes the value 0 in normal driving (noted ″normal″) and 1 in eco-driving (noted ″eco″). Here, the goal is not to detect the learning method, but to evaluate afterwards how effective is the training itself, in order to produce knowledge about which training method provide the best results. Effectiveness of the trainings can be evaluated through the ability of the fitted models to discriminate the two driving styles. For each data set, and for each rule, the significance of the associated indicator allow us to evaluate which advice is practically used by the drivers. The more significant is the indicator, the more differences exist between the control and the eco-driving group in applying the associated rule. We claim that this reflects the training performances: An efficient training should lead to a clear modification of the driving behaviour.

However, in our two experiments, both the number of clusters (20 in the experiment 1 and 19 in the experiment 2) and the cluster size (2 in the two experiments) are small, which implies various constraints to choose a suitable statistical model. In a first part, the smallness of our sample size limits the number of predictors for which effects can be estimated precisely. Peduzzi et al. (1996) suggests there should ideally be at least ten outcomes of each type for every predictor. This result constrains us to assess the effects of each driving advice separately and consequently to construct one logistic regression model with each of the four indicators as predictor. In a second part, the smallness of our sample size does not allow us to use the appropriate statistical models taking into account driver specific correlations. Indeed, we tested the GEE method using the PROC GENMOD of the SAS software, but the parameters estimates were closed to zero. Ziegler et al. (1998) 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. We also tested the generalized linear mixed models using the PROC GLIMMIX of SAS but a statement indicates that one of the estimated variance parameters was negative. This result is an underestimate of the true variance component that occurs when the number of observations per random effect category is small or when the ratio of the true variance component to the residual is small. Moreover, several studies (Moineddin (2007), Theall (2011)) have shown that parameters estimates are unbiased with either 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 logistic model can be written as:

$$logit\left[P\left(Trip = Eco\right)\right] = \alpha + \beta X \quad (5)$$

where $\propto$ is the intercept, X is one of the four indicators associated with the main rules of eco-driving (Table 1) and β is the parameter estimate of the predictor X. The results from each logistic model are listed in Table 3 for the experiment 1 and Table 4 for the experiment 2. For each logistic model, we indicate the explanatory variable X, the estimated parameter β, its standard error SE and the p-value of the Wald test. We also indicate the odds ratios (OR) and their 95% Wald confidence limits. The usefulness of each model is measure by the Nagelkerke R^2, denoted R_N^2, 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 variable is in predicting the response variable. Finally, the predictive power of each model is measure by the area under the ROC curve (AUC). This parameter, ranges from zero to one and identical to the concordance index, assess the discrimination power of the model. In our study, it measures the model's ability to discriminate between eco-driving trips versus normal trips. More details on these various parameters are given in Agresti (2002) or Hosmer and Lemeshow (2000).

In Table 3 and Table 4, the four logistic models, assessing the implementation of each rules of eco-driving, are ranked in descending order of both parameters R_N^2 and AUC and thus represents the order of implementation of each driving advice. Table 3 shows that all the indicators are significant (p-value lower than 0.01 and 95% confidence interval including one) in the experiment 1 but the indicators associated with the first three rules are most significant: relatively high R_N^2 reflecting that the three indicators *AvgRPMShiftUp*, *IndexGearRPM* and *PKE* are useful in predicting eco-driving trip, and AUC greater than 0.8 reflecting a high discriminatory power of this three models. On the contrary, the indicator *TimeEngineBrake* is not very useful in predicting eco-driving trip (R_N^2=0.289) even if the discriminatory power of this model is acceptable (0.7 ≤ AUC ≤ 0.8). Table 4 shows the results obtained in the experiment 2. The results are globally similar to those obtained in the experiment 1 except that the indicator *TimeEngineBrake* is no longer significant (one is excluded of the 95% confidence interval) and the model associated is not very useful in predicting eco-driving behaviour (R_N^2 close to zero and AUC close to 0.5 indicating poor discrimination of the model).

Table 3. Experiment 1: logistic regression models with each of the four indicators associated with the main rules of eco-driving and ranked in descending order of implementation of each driving advice.

Models	β	SE	OR	95% CI	R_N^2	AUC
X= AvgRPMShiftUp (Rule 1)	-0.0068^{**}	0.002	0.993	0.989 - 0.997	0.608	0.908
X=PKE (Rule 3)	-34.0893^{**}	10.622	< 0.001	< 0.001 - < 0.001	0.594	0.898
X= IndexGearRPM (Rule 2)	-0.3068^{**}	0.103	0.736	0.601 - 0.901	0.491	0.866
X= TimeEngineBrake (Rule 4)	0.1849^{**}	0.071	1.203	1.047 - 1.383	0.289	0.780

$^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$.
SE: standard error; OR: odds ratio; CI: confidence interval; R_N^2: Nagelkerke R^2; AUC: area under the ROC curve.

Table 4. Experiment 2: logistic regression models with each of the four indicators associated with the main rules of eco-driving and ranked in descending order of implementation of each driving advice.

Models	β	SE	OR	95% CI	R_N^2	AUC
X= IndexGearRPM (Rule 2)	-1.4262^{*}	0.677	0.240	0.064 - 0.906	0.922	0.989
X= AvgRPMShiftUp (Rule 1)	-0.0126^{**}	0.004	0.987	0.979 - 0.996	0.878	0.976
X=PKE (Rule 3)	-63.7126^{**}	21.715	< 0.001	< 0.001 - < 0.001	0.744	0.952
X= TimeEngineBrake (Rule 4)	0.0254	0.065	1.026	0.902 - 1.166	0.005	0.568

$^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$.
SE: standard error; OR: odds ratio; CI: confidence interval; R_N^2: Nagelkerke R^2; AUC: area under the ROC curve.

4.3 Eco-driving effects for different speed limits

Assuming that eco-driving behaviour depends on the road conditions, previous logistic models were extended to more complex models taking into account the speed limits. The variable "Speed limit" is used as a stratification variable in order to derive specific models. Thus, for each trip of the two experiments, sections corresponding to a specific speed limit were merged for analysis. The calculation of the four indicators defined in Table 1 was then adapted on these new trip to take into account the grouping of sections not necessarily continuous. Table 5, Table 6 and Table 7 contain the estimated parameter, its standard error, the Nagelkerke R² and the AUC for the three main speed limits: 50km/h, 70km/h and 90km/h. Table 5 shows similar results for the two experiments when the speed limit is 50km/h: the three indicators *AvgRPMShiftUp*, *IndexGearRPM* and *PKE* are most significant while the indicator *TimeEngineBrake* is not very useful in predicting eco-driving behaviour. Table 6, corresponding to the speed limit 70km/h, shows that in the experiment 1, the four driving advice have been applied while in the experiment 2, only the first three advice have been applied. Finally, Table 7 shows that when the speed limit is 90km/h, the indicators *AvgRPMShiftUp* and *IndexGearRPM* are most significant in the two experiments whereas the indicator PKE is less significant than with the previous speed limitations. As for areas limited to 50km/h, the indicator *TimeEngineBrake* is not useful in predicting eco-driving behaviour and in the experiment 1, the estimated parameter is negative (but no significant) which means that engine brake seems to have been less used during eco-driving trips than during normal trips.

Table 5. Logistic regression models for 50km/h speed limit.

Models	Experiment 1				Experiment 2			
	β	SE	R_N^2	AUC	β	SE	R_N^2	AUC
X= AvgRPMShiftUp (Rule 1)	-0.007^{***}	0.002	0.62	0.909	-0.012^{**}	0.004	0.83	0.964
X= IndexGearRPM (Rule 2)	-0.371^{**}	0.124	0.56	0.903	-0.908^{*}	0.362	0.85	0.978
X=PKE (Rule 3)	-36.022^{**}	11.969	0.59	0.896	-34.859^{**}	11.301	0.64	0.922
X= TimeEngineBrake (Rule 4)	0.108^{*}	0.045	0.24	0.745	0.074	0.074	0.07	0.676

$^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$.
SE: standard error; R_N^2: Nagelkerke R^2; AUC: area under the ROC curve.

Table 6. Logistic regression models for 70km/h speed limit.

Models	Experiment 1				Experiment 2			
	β	SE	R_N^2	AUC	β	SE	R_N^2	AUC
X= AvgRPMShiftUp (Rule 1)	-0.005^{**}	0.002	0.48	0.871	-0.013^{**}	0.005	0.88	0.986
X= IndexGearRPM (Rule 2)	-0.293^{**}	0.105	0.43	0.851	-0.459^{***}	0.139	0.76	0.938
X=PKE (Rule 3)	-25.350^{***}	7.705	0.46	0.863	-31.830^{**}	10.967	0.61	0.922
X= TimeEngineBrake (Rule 4)	0.178^{**}	0.063	0.35	0.795	0.034	0.050	0.02	0.562

$^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$.
SE: standard error; R_N^2: Nagelkerke R^2; AUC: area under the ROC curve.

Table 7. Logistic regression models for 90km/h speed limit.

Models	Experiment 1				Experiment 2			
	β	SE	R_N^2	AUC	β	SE	R_N^2	AUC
X= AvgRPMShiftUp (Rule 1)	-0.005^{**}	0.002	0.47	0.868	-0.019^{*}	0.009	0.90	0.989
X= IndexGearRPM (Rule 2)	-0.225^{**}	0.077	0.43	0.850	-0.494^{**}	0.159	0.78	0.956
X=PKE (Rule 3)	-14.054^{*}	5.463	0.27	0.745	-21.121^{*}	8.280	0.33	0.758
X= TimeEngineBrake (Rule 4)	-0.015	0.080	0.001	0.521	0.038	0.051	0.02	0.651

$^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$.
SE: standard error; R_N^2: Nagelkerke R^2; AUC: area under the ROC curve.

4.4 Characterize eco-driving style with an aggregated indicator

The four rules of eco-driving can be merged together in order to perform a global evaluation of the performance reached by the driver, not based on fuel consumption, but on an academic definition of eco-driving represented by the 4 golden rules (Table 1). The estimated logistic model using data from experiment 1 is the following:

$$logit\left[P\left(Trip = Eco\right)\right] = 8.967 - 0.007 \cdot X_1 + 0.242 \cdot X_2 - 31.684 \cdot X_3 + 0.148 \cdot X_4 \quad (6)$$

where $X_1, \dots, X_4$ are the four indicators associated respectively with the four instructions of eco-driving (Table 1). The model (6) reached a total $R_N^2 = 0.74$, 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 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 $P(Trip = Eco)$ multiplied by one hundred. So we obtain an index of eco-driving which varies between 0 and 100 for easier interpretation. 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 $R_N^2 = 0.70$, which shows that our eco-index is closely related to the average fuel consumption. If we take into account previous results about the way subjects apply eco-driving, the sole parameter PKE can be used as a predictive indicator by using the corresponding estimated value given at Table 3. More details on this simpler model are given in Andrieu and Saint Pierre (2012). 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 estimated by model (6), distinguishing "normal" and "eco" trips, is represented in Figure 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 good eco-driving indicator since "eco" trips are associated with high eco-index whereas "normal" trips are associated with lower eco-index. This procedure can be extended to 30 seconds time driving data, for which an instantaneous eco-index can be computed using model (6). It is therefore possible to provide instant feedback to the driver, by using the eco-index information available at each seconds and computed from the last 30 seconds of driving. These principles are now applied in the recently launched ecoDriver European project (http://www.ecodriver-project.eu/) which goal is to build an EDAS to support the driver in conserving energy and reducing emissions. A modified version of the eco-index presented in this work will be implemented as a tool to provide simple feedback to the driver via an android application.

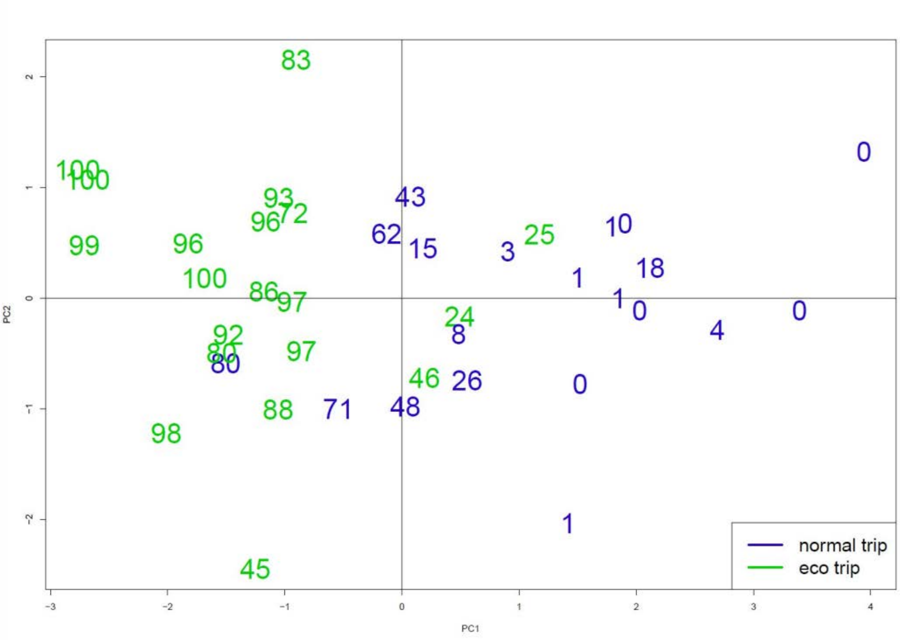

Figure 1. Principal component analysis and the estimated eco-index ("normal" trips are in blue, "eco" trips are in green).

5. Conclusion

This study provides a statistical analyses of two learning mode of eco-driving: one with simple eco-driving advice and the other with eco-driving training. The study of different parameters like average fuel consumption, average speed, or average acceleration shows a real positive impact of eco-driving style regardless of the learning mode. The association of each of the main eco-driving rules with a quantitative indicator allows us to assess the effect of each driving advice separately using logistic regression models.

It is shown that drivers succeed efficiently in applying advice related to constant speed or gearshift strategy regardless of the learning mode of eco-driving, while they are less efficient in using engine brake (small parameter influence for experiment 1 and insignificant for experiment 2). The same analysis is done for each different speed limit zone in order to take into account the effects of the driving environment. Results are all together in line although significant differences are found for the engine brake related rule. On 70km/h limited areas the engine brake was not correctly used in experiment 2 (with eco-driving training) while all the four driving advice were correctly implemented in experiment 1 (with simple advice). On the contrary, on 90km/h limited areas, the 4th rule effect is insignificant for both experiments although the engine brake seems to have been less used during eco-driving trips than during normal trips.

Golden rules indicators show that fuel efficient driving is better implemented after a course than just applying eco-driving tips (greater R_N^2 and AUC). Differences are small due to the bias introduced by the presence of an experimenter in the car in both experiment. Suitable experimental designs and specific studies are needed to quantify precisely the size of the differences between the two learning modes.

Data sets used in this paper are small and lack of consistency between controlled factors for each experiment (different drivers, cars, driving conditions, etc.) but it is worth trying a meta-analysis to improve veracity of the results. Effects sizes are in line all together showing the ability of our indicators to represent eco-driving capacities. Our work show that just reading simple eco-driving advice allows drivers to reduce significantly their fuel consumption and to adopt an eco-driving behaviour although performances are better after a course. The important question now is to find how long a fuel efficient driving behaviour last depending on the way drivers learned it.

This study also provides a methodology suitable to compute a global eco-driving indicator based on statistical models, taking into account various behaviour related parameters. A logistic regression model has been developed, which explanatory variables are the four performance indicators associated with each of the main rules of eco-driving. This model provides an 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. Such a work is under progress as a part of the ecoDriver project EDAS developments.

Assuming that eco-driving behaviour depends on the road conditions, it is possible to extend the full model (6) to a more complex model taking into account the speed limit as a stratification variable. This approach could improves the properties of model (6) by producing more accurate information, and could help informing the driver on the network categories (urban, rural, ...) on which he can improve his efficiently driving.

We have noted that different statistical models able to take into account driver specific correlations, namely GEE and mixed models, could have been used. Unfortunately 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 Moulton(1989). New experiments are scheduled to extend our datasets in order to use a more appropriate statistical model.

Future works will focus on the validation of the 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. A practical implementation will be made for the ecoDriver project with the help of a nomadic ecological driving assistance application.

Acknowledgements

Authors would like to acknowledge the French Ministry of Ecology, Sustainable Development and Energy for its financial support.

References

Agresti, A. (2002). *Categorical data analysis.* 2nd edition. Wiley Series in Probability and Mathematical Statistics, Chichester.

Andrieu, C. and Saint Pierre, G. (2012). Using statistical models to characterize eco-driving style with an aggregated indicator, *IEEE Intelligent Vehicle Symposium (IV 2012)*, Alcala de Henares, Spain.

Barbé, J., Boy, G.A. & Sans, M. (2008). GERICO: A human centered eco-driving system. *IFAC analysis, Design, and Evaluation of Human-Machine Systems*, Volume 10 / Part 1 (http://www.ifac-papersonline.net/Detailed/39408.html).

Benminoun, M., et al. (2011). Safety analysis method for assessing the impacts of advanced driver assistance systems within the European large scale filed test "EuroFOT". *8th ITS European Congress* 2011, Lyon.

Chapnick, N. (2007). Hypermiling: quest for ultimate fuel economy, Edmunds.com, Retrieved from http://www.edmunds.com/fuel-economy/hypermiling-quest-for-ultimate-fuel-economy.html.

CIECA (International commission for driver testing authorities) (2007). Final report of the internal project on 'Eco-driving' in category B driver training & the driving test. [Online]. Available: http://www.cieca.be/.

Dozza, M., Bärgman, J., and Lee, J. D. (2012). Chunking: A procedure to improve naturalistic data analysis *Accident Analysis and Prevention, 58*, 309-317.

Ecodrive project (2009). found at: http://www.ecodrive.org.

EcoDriver project (2012), found at http://www.ecodriver-project.eu.

Ericsson, E. (2001). Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D: Transport and Environment*, 6, 325-345.

Guo, F. and Hankey, J. (2010). Modeling 100-Car Safety Events: A Case-Based Approach for Analyzing Naturalistic Driving Data. *The National Surface Transportation Safety Center for Excellence.*

Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression.* Second edition. John Wiley and Sons, New York, NY.

Liang, K.-Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22.

Moineddin R, Matheson FI, Glazier RH. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Med Res Methodol 2007*;7:34.

Moulton, L. H., and Zeger, S. L. (1989). Analyzing repeated measures on generalized linear models via the bootstrap, *Biometrics*, 45, 381–394.

Peduzzi, P., et al. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49: 1372-1379.

Symmons, M., Rose, G., VanDoorn, G., Ecodrive as a road safety tool for Australian conditions, *Monash University*, 2009.

Theall, K.P., et al. (2011). Impact of small group size on neighbourhood influences in multilevel models. *J Epidemiol Community Health*, 65(8), 688-695.

Walker, G.H., Stanton, N.A. and Young, M.S. (2001). Hierarchical Task Analysis of Driving: A New Research Tool, In M.A.Hanson (Ed), *Contemporary Ergonomics*, Taylor & Francis Ltd., London, 435-440.

Young, M. S., Birrell, S. A., and Stanton, N. A. (2009). Design for Smart Driving: A Tale of Two Interfaces, In *Lecture Notes in Computer Science: Engineering Psychology and Cognitive Ergonomics*, Springer Berlin / Heidelberg (ed), *477–485.*

Young, M. S., Birrell, S. A., and Stanton, N. A. (2010). Safe driving in a green world: A review of driver performance benchmarks and technologies to support "smart" driving, *Applied Ergonomics*.

Ziegler, A., Kastner, C. and Blettner, M. (1998). The generalised estimating equations: an annotated bibliography, *Biometrical Journal*, 40(2), 115-139.