Analytic Methods in Accident Research 26 (2020) 100114

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

journal homepage: www.elsevier.com/locate/amar

Driver drowsiness detection using mixed-effect ordered logit model considering time cumulative effect

Xuxin Zhang [a,b], Xuesong Wang [a,b,c,*], Xiaohan Yang [d], Chuan Xu [e], Xiaohui Zhu [f], Jiaohua Wei [f]

[a] *The Key Laboratory of Road and Traffic Engineering, Ministry of Education, China*
[b] *School of Transportation Engineering, Tongji University, Shanghai 201804, China*
[c] *Chinese National Engineering Laboratory for Integrated Optimization of Road Traffic and Safety Analysis Technologies, China*
[d] *School of Mathematical Sciences, Tongji University, Shanghai 200092, China*
[e] *School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*
[f] *China Pacific Property Insurance Co., Ltd, Shanghai 200120, China*

ARTICLE INFO

Article history:
Received 1 April 2019
Received in revised form 16 January 2020
Accepted 16 January 2020
Available online 25 January 2020

Keywords:
Drowsy driving
Karolinska Sleepiness Scale
Individual difference
Time cumulative effect
Mixed-effect ordered logit model
Mixed generalized ordered response model

ABSTRACT

Drowsy driving is one of the main causes of traffic crashes, a serious threat to road traffic safety. The effective early detection of a drowsiness state can help provide a timely warning for drivers, but previous studies have seldom considered the cumulative effect of drowsiness over time. The purpose of this study is therefore to establish a model to detect a driver's drowsiness level by considering individual differences combined with the time cumulative effect (TCE) of drowsiness. Driving behavior and eye movement data from 27 drivers were collected by a driving simulator with an eye-tracking system, and the Karolinska Sleepiness Scale (KSS) was used to record drivers' perceptions of their states of drowsiness. Since the degree of driver drowsiness was shown to increase with time, a mixed-effect ordered logit (MOL) model was established, and a non-decreasing function of time was applied to consider time accumulation. Results showed that with increasing drowsiness, the standard deviation of lateral position and percentage of driver eyelid closure (PERCLOS) increased significantly. Consideration of these variables can thus improve the accuracy of drowsy driving detection. The developed MOL-TCE model was compared with a non-TCE MOL and a TCE mixed generalized ordered response (MGOR) model. The drowsiness detection accuracy of the MOL-TCE model was 62.84%, higher than the 61.04% accuracy of the MGOR-TCE and appreciably higher than the 52.47% of the non-TCE MOL model.

1. Introduction

The decrease in alertness that occurs in drowsy driving has been shown to increase driver reaction time (Young et al., 1997), which can lead to impaired performance at the wheel (Vanlaar et al., 2008), impairment that has, in turn, made drowsy driving a major cause of traffic crashes. Research shows that the probability of crashes during drowsy driving is 4–6 times that of alert driving (Klauer et al., 2006). In the United States, drowsy drivers were involved in 2.1% of crash fatalities in 2016 (NHTSA, 2017), and more than 40% of drivers admitted to having "fallen asleep or nodded off" at the wheel (Tefft, 2010). According to data published by the Traffic Management Bureau of the Ministry of Public Security of China

* Corresponding author at: School of Transportation Engineering, Tongji University, Shanghai, China.
E-mail address: wangxs@tongji.edu.cn (X. Wang).

https://doi.org/10.1016/j.amar.2020.100114

(Traffic Management Bureau of Ministry of Public Security, 2016), drowsy driving was the main cause of highway crashes in 2015; it accounted for 8.41% of total crashes and 6.21% of highway crash deaths.

A warning system could be developed to alert drivers reaching a dangerous state of drowsiness if more effective detection methods were found. Although drowsy driving has been generally well researched, most existing studies have focused on driving behavior and physiological characteristic indicators, leaving two main areas in need of more attention: individual differences and time cumulative effects (TCE). In one of the few studies considering individual differences, Wang and Xu (2016) found that the drowsiness detection accuracy of their multilevel ordered logit model was significantly improved. The TCE of drowsiness has also been neglected, despite drowsy driving being long understood as a gradual process that accumulates as driving time increases (Expert Panel on Driver Fatigue and Sleepiness, 1998; Watson et al., 2015). McDonald et al. (2018) presented an approach that used driver behavior data and considered the time dependencies in transitions between drowsy and awake states. They used the Dynamic Bayesian Network (DBN) algorithm to detect drowsiness, which showed good precision in drowsiness-related lane departures. However, the DBN algorithm did not take drivers' individual differences into consideration, so the result may have bias.

Few models have combined individual differences with TCE, resulting in limited accuracy and reliability. Therefore, the purpose of this study is to improve drowsiness detection by including the impact of these key factors. A mixed-effect ordered logit (MOL) model was used as the basic model. The time cumulative effect (TCE) was represented by a non-decreasing function, and a mixed-effect ordered logit model considering the time cumulative effect (MOL-TCE model) was obtained. The Tongji University driving simulator was used to collect driver behavior and vehicle movement data, while its eye-tracking system simultaneously collected eye-feature changes.

2. Literature review

There are many causes of drowsiness, such as excessive daytime sleepiness, habitual snoring, insufficient sleep, untreated sleep disorders, and even depression (Sunwoo et al., 2017; Stutts et al., 2003; Mccartt et al., 2000). Light and circadian rhythm changes have also been associated with drowsiness (Chipman and Yue, 2009), and driving while drowsy is reported with a higher frequency by people who report having less sleep (Maia et al., 2013). The combination of these sleep related factors with the extended period of concentration required by driving can result in drowsy driving. Over long periods of driving, for example, the environment can begin to lack stimulation for the driver. Thiffault and Bergeron (2003) used a driving simulator to measure the influence of a monotonous driving environment on driver behavior. Their results showed that driving ability significantly decreased in a monotonous environment, as demonstrated by increased frequency of steering wheel movement. Farahmand and Boroujerdian (2018) investigated the influence on driver fatigue of monotonous road geometry. They found that greater geometric variety of the road led to better performance and higher vigilance by the studied drivers.

Drowsy driving research has varied in type of data and collection methods, which include observation of physiological data such as electroencephalograms, electrocardiographs, and the recording of eye movement; assessing driving behavior using experimental means such as driving simulator apparatus (Wang and Xu, 2016; Wang et al., 2017); and subjective evaluation of drowsiness such as self-reported questionnaires. Many researchers use subjective sleepiness scales including the Stanford Sleepiness Scale (De Valck and Cluydts, 2001), Visual Analogue Scales (VAS) (Lenné et al., 1997), Swedish Occupational Fatigue Inventory (SOFI) (Orongilad et al., 2008), Chalder Fatigue Scale (CFS) (Stasi et al., 2012), and the Karolinska Sleepiness Scale (KSS) (Åkerstedt and Gillberg, 1990). The KSS is a one-item nine-level scale that is still commonly used to quickly measure states of sleepiness (Kaida et al., 2006; He et al., 2017; Wu et al., 2019). The KSS score has been demonstrated to be significantly correlated with electro-oculogram (EOG) changes and to reflect the effect of sleepiness on task performance (Hu and Zheng, 2009). Subjects self-assess their sleepiness, or drowsiness, states from 1 to 9: 1 is *extremely alert* and 9 is defined as *very sleepy, great effort to keep alert, fighting sleep*. KSS's advantages include its detailed classification while also being simple; that is, it does not require participants to answer multiple questions while driving, which could easily skew results. Subjective visual assessment (not self-reported) has also been used to determine driver drowsiness level. In the Trained Observer Rating (TOR) method (Wierwille and Ellsworth, 1994), raters assess drivers in 10-s intervals according to five different states from 0 (alert) to 4 (extremely drowsy).

In addition to subjective evaluation, driver drowsiness levels have been assessed through the collection and analysis of drivers' physiological characteristics using electroencephalograms (EEGs) (Wei et al., 2018; Qian et al., 2017) and electrocardiograms (ECGs) (Awais et al., 2017). Both show good detection precision, but are intrusive measures that can affect driver physiology and behavior in the experimental process, again skewing results (Rodrigues et al., 2010). The driving process is less disturbed by measuring drivers' eye features such as blink duration (Ingre et al., 2006) and percentage of eyelid closure (PERCLOS). PERCLOS directly reflects the state of the eyes receiving external information, and previous studies have found that the amount of information acquired from the eyes is related to accurate predictions of drowsiness (Jacobé de Naurois et al., 2017). PERCLOS predicts drowsiness based on the percentage of time an individual's eyes are more than 80% closed over a specific period (generally a 30 or 60-s period), and, since eyelid closure is detected by an eye tracker, the method is non-intrusive.

The third common type of data used in assessing drowsiness level is driving behavior collected by driving experiment (for example, by naturalistic driving studies or driving simulators), and includes behaviors such as vehicle lateral position, steering wheel movement, brake/accelerator pedal action, and speed (Jacobé de Naurois et al., 2018; Meng et al., 2019; Sparrow et al., 2018). Lateral drifting demonstrates driver instability in vehicle handling, and Jackson et al. (2016) showed that the standard deviation of a driver's lateral position increases significantly with the increase of drowsiness. Thiffault and Bergeron (2003) found a greater frequency of large steering wheel movements during driving in monotonous road environments, which implied that drivers' fatigue level increased and their vigilance decreased under such circumstances. Hard-braking events appear to show greater frequency with higher levels of fatigue, and are also used as indicators to predict fatigue (Mollicone et al., 2018).

Drowsy driving is typically analyzed by statistical modelling and data mining. Data mining usually refers to searching by algorithm for information hidden in a large quantity of data. Data mining technology such as Random Forest can be used to analyze driver behavior and physiological characteristics data (Atiquzzaman et al., 2018). Hu and Zheng (2009) used the Support Vector Machine (SVM) to predict drowsiness state by drivers' eye movement parameters; results showed that the SVM has a high prediction accuracy when drivers are very sleepy. Using an Artificial Neural Network to analyze drivers' heart rate variability data, Patel et al. (2011) accurately detected drivers' states of early drowsiness. Yang et al. (2009) explored driving performance variation at different levels of sleep deprivation and used a Bayesian Network to hypothesize the driver's drowsiness state. The results showed that when drivers were sleep-deprived, their reaction behavior to unexpected disturbances degraded.

Statistical models such as multinomial logistic regression (Murata et al., 2015), random effects, mixed generalized ordered response (Balusu et al., 2018), multilevel ordered logit, and structural equation models (SEM) (Useche et al., 2017), have also been used for drowsy driving analysis. Using a logistic regression model, Wilkinson et al. (2013) showed that average duration of eye closure was an important eye movement indicator for judging drowsiness. Meng et al. (2019) used a random effects model to evaluate drivers' fatigue levels and their driving performance. Ordered logit models have commonly been used for ordinal data structure (Wang and Abdel-Aty, 2008); as the Karolinska Sleepiness Scale value is an ordinal categorical variable, Wang and Xu (2016) used a multilevel ordered logit model to consider individual differences of drivers based on KSS, driving behavior, and eye movement data. The model demonstrated better accuracy and reliability than models that did not consider individual differences. A mixed generalized ordered response model has also commonly been used to handle ordered outcome problems (Balusu et al., 2018; Yasmin et al., 2015) such as crash injury severity (Eluru et al., 2008). Eluru et al. found that, for several variables, the mixed generalized ordered response logit model provided better statistical fit and more consistent estimates than the standard ordered response logit model. Naderi et al. (2018) and Useche et al. (2017) used SEMs to analyze the relationship between fatigue and driving behavior, and both found that fatigue did lead to an increase in risky driving behavior.

In general, data mining is more effective than statistical modelling for classification prediction and when a large number of samples are needed, but the statistical model with parameter estimation is most effective for interpretation of variables. As interpretation is important to our purpose of improving drowsiness detection, the statistical model was determined to be more applicable to this study. Because drivers are likely to have individual differences in driving behavior and eye response features, a mixed-effect ordered logit (MOL) model was set up according to the stratified driver definitions. To consider the time cumulative effect of drowsiness that has been shown to enhance detection capability, an MOL-TCE model was established by adding a non-decreasing function of time. In order to compare detection accuracy, a non-TCE MOL and a mixed generalized ordered response model considering time cumulative effect (MGOR-TCE) were also established. The Karolinska Sleepiness Scale (KSS) was used to evaluate driver drowsiness.

3. Data preparation

3.1. Apparatus

Driving behavior data was collected through the Tongji University driving simulator shown in Fig. 1. The simulator consists of five sub-systems: dome, motion, instrumented car and control feel, visual control, and simulation control. The dome

(a) Simulator exterior (b) Instrumented car (c) Driver view

Fig. 1. Tongji Advanced Driving Simulator.

includes dome structure, car mount, and electric equipment. The motion sub-system provides an 8-degree of freedom motion control, and the instrumented car and control feel sub-system includes rearview mirrors, audio feedback, intercom, and video monitor. The visual sub-system contains a curved screen with a visual range of $250^\circ \times 40^\circ$, which increases the realism of the experimental scene. Simulation control includes supervision PC and screens. These simulator data are collected at a frequency of 10 Hz.

Eye movement data were collected through a four-channel SmartEye® eye-tracking system. The eye-tracking system includes four cameras placed in front of the vehicle to capture the driver's eye movements at a 60-Hz sampling rate.

3.2. Participants

Twenty-seven male drivers aged 24–52 years (mean 33.2, SD 8.0) were recruited for this study. To ensure that the subjects' alertness was not influenced differently by the time of their working day, all were night-shift workers who work 11 h a day on average. As a result, their bodies are very likely to be in a state of physiological fatigue after work. All subjects were required to have a valid Chinese driver's license, good health, no sleep-related diseases or issues, and have had no drug or medicine use within the month prior to the experiment, no alcohol within 24 h, and no stimulant beverages within 12 h of the experiment. Just prior to the experiment, participants signed an Experiment Informed Consent form that described the study's requirements, risks, and rights. Each participant was paid the equivalent of $30 U.S. at the end of the experimental session.

3.3. Experiment design

The driving course used in the experimental scenario was a 20-km 2-way 6-lane circular rural highway track with a separation median, as shown in Fig. 2. The road was composed of six straight 2-km sections (sections 1, 3, 5, 9, 11, 13), two circular curves with radii of 700 m (sections 7, 15), and eight transition curves (sections 2, 4, 6, 8, 10, 12, 14, 16). In order to increase environmental realism, green grass, trees and a few buildings were added to both sides of the road. Additional vehicles were also included to enhance the visual driving experience. A small number of vehicles moved in the same direction as the test vehicle and a greater number traveled in the oncoming lanes. None of vehicles interacted with the test vehicle so as to reduce the interference of external factors on the experimental results.

To ensure drivers would experience a level of drowsiness high enough to accurately evaluate its influence on driving performance, the subjects were required to arrive at the simulator site at 8:00 a.m. after their night-shift work. Before the experiment, they completed a basic information questionnaire and practiced driving in the simulator. In order to improve the accuracy of the self-reported Karolinska Sleepiness Scale (KSS), the KSS classification criteria were explained in detail.

Beginning at 8:30 a.m., each participant was asked to drive 6 revolutions (about 1 h in total) of the experimental road. To restrict factors that could limit drowsiness, drivers were advised to maintain lane position and a speed of 120 km/h; use of mobile phones and radios was prohibited. Since the driving scenario was set as daytime sunny with no tunnels, drivers had no need to illuminate the vehicle's headlights. Only occasional and uneventful traffic was present, and was standardized for all drivers.

3.4. Data collection

As actual drowsiness is latent and cannot be directly observed, it consequently lacks objective and accurate evaluation parameters. Drowsiness, however, has been correlated in previous research with the more objective indicators of physiological and behavioral changes (Wang and Xu, 2016; Wang et al., 2017), changes that can potentially be used to trigger drowsiness warning systems. To assess the correlation between these objective measures and subjective driver reports of drowsiness, this study collected three types of raw driver data: physiological eye features, driving behavior, and subjective KSS drowsiness levels. Eye feature and driving behavior data were collected throughout the driving task by the simulator laboratory apparatus. Eye features included eyelid closure, pupil diameter, blink frequency and blink duration; driving behavior indicators included vehicle speed, lateral position, and steering wheel movement.

Considering that driving behavior and self-assessments of drowsiness may be impacted by the frequency of KSS questioning, we tested different schedules (queries every 5, 10, and 15 min) in a pilot experiment to ensure that the questioning

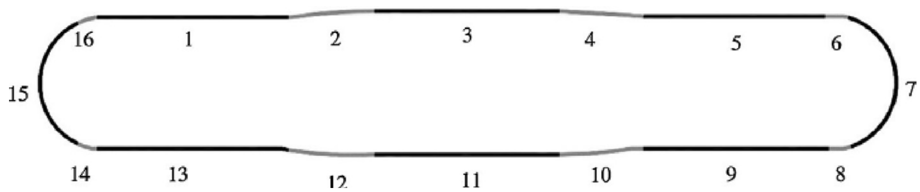

Fig. 2. Driving Course and Segment ID Numbers.

Fig. 3. Driver's Facial Expression and Hand Operation Monitoring.

would elicit a sufficient sample of drowsiness reports, yet have minimal impact on driver behavior. When queried every 5 min, subjects reported that they felt forced to stay awake, but when they were queried every 15 min, the sample was insufficient. The 10-minute interval best met the criteria, so participants were asked to report their self-perceived KSS level every 10 min during the 1-hour driving task. To check the reliability of the drivers' self-reporting, two researchers at a monitor in the simulator control room observed each driver's facial state and body movement in real time (see Fig. 3). Their criteria for judging driver drowsiness were based on the Trained Observer Rating (TOR) method developed by Wierwille and Ellsworth (1994), in which raters assess drivers in 10-s intervals according to five different states from 0 to 4.

The KSS classification criteria are shown in the first column of Table 1, and the TOR criteria are in the two columns to the right. At the end of the experiment, the KSS of each driver was checked against the observers' assessment. When the difference between the corresponding driver's self-reported score and the experimenters' score ≥ 2, the video data were checked to confirm or correct, as needed, the driver's assessment.

Because one driver's drowsiness led to his falling asleep during his fifth circle around the track, only data from his first four rounds were analyzed. Table 2 shows this study's descriptive statistics of several of the empirical explanatory variables that were found significant in previous literature (Merat and Jamson, 2013; Jackson et al., 2016).

Fig. 4 shows the change in the average LP_Stdev, PERCLOS, and KSS over the time cumulation during the six rounds of driving. With the increase in driving time, drivers' average reported drowsiness level grew, and, for most of the test drive, the standard deviation of lane position and eyelid closure also grew. The decrease of LP_stdev beginning in the fourth round is likely due to the increasingly drowsy driver's reduced operation of the vehicle. By the fifth round, the driver may be attempting to open his eyes as behavioral compensation to reduce his drowsiness, thereby decreasing his PERCLOS.

Although the nine detailed classification levels of the KSS are often very useful, for example, in a study of change in behavior or physiological characteristics as drowsiness increases (Vadeby et al., 2010), the drowsiness levels in this study were merged into three before modeling, as shown in Table 3. The nine levels can be difficult for drivers to distinguish, and would make our sample size at each drowsiness level too small to be conducive to model estimation. Additionally, the three-level classification has been shown to be more suitable for the analysis of significant factors related to drowsiness (Hu and Zheng, 2009; Wang and Xu, 2016). Other studies have classified driver drowsiness into just two categories, 0-no drowsiness and 1-drowsiness (Loon et al., 2015; Martensson et al., 2019), but as the purpose of this paper is to improve detection of driver drowsiness in order to support in-vehicle warning systems similar to the DD850 Driver Fatigue Monitor, three categories provide more nuance and are more readily transferable to drowsiness warning systems. As can be seen in Table 3, none of the KSS Levels 1–5 indicate drowsiness, so an early warning system would not yet be triggered. KSS Levels 6–7, which correspond to our Level 2, both indicate some sleepiness, but no problem maintaining wakefulness. An early warning system would give an initial alert to the driver at this point, but a second alert of increased intensity would be given when drowsiness became problematic for the driver, that is, at KSS Levels 8–9, which correspond to our Level 3.

In order to prevent the models from overfitting the data, the original data were divided into training and test sets. A random sampling method was used to extract 2/3 of the data from the three drowsiness levels for the training set (1,637 samples), and the remaining 1/3 (818 samples) were analyzed as the test set.

4. Methodology

4.1. Mixed-effect ordered logit model

An ordered logit model is commonly used to handle order-classified variables, that is, when all explanatory variables have the same parameter estimates at different segmentation points.

Use Y_{ij} to indicate the drowsiness level of the jth driver on the ith road section. It may be assumed that there is a potential continuous variable y^*_{ij} corresponding to Y_{ij}, and therefore that the discrete Y_{ij} is determined by the range of a series of

Table 1
Subjective evaluation criteria of driver drowsiness level.

KSS level (Driver's subjective assessment)	TOR level (Researchers' subjective assessment)	TOR Indicators
1 Extremely alert 2 Very alert 3 Alert 4 Rather alert 5 Neither alert nor sleepy	0 Not drowsy	Normal fast eye blinks, often reasonably regular; Apparent focus on driving with occasional fast sideways glances; Normal facial tone; Occasional head, arm and body movements.
6 Some signs of sleepiness	1 Slightly drowsy	Increase in duration of eye blinks; Possible increase in rate of eye blinks; Increase in duration and frequency of sideway glances; Appearance of "glazed eye" look; Appearance of abrupt irregular movements – rubbing face/eyes, moving restlessly on the chair; Abnormally large body movements following drowsiness episodes; Occasional yawning.
7 Sleepy, but no effort to keep alert	2 Moderately drowsy	Occasional disruption of eye focus; Significant increase in eye blink duration; Disappearance of eye blink patterns observed during alert state; Reduction on degree of eye opening; Occasional disappearance of facial tone; Episodes without any body movements.
8 Sleepy, some effort to keep alert	3 Very drowsy	Discernable episodes of almost complete eye closure, eyes never fully open; Significant disruption of eye focus; Periods without body movements (longer than for level 2) and facial tone followed by abrupt large body movements.
9 Very sleepy, great effort to keep alert, fighting sleep	4 Extremely drowsy	Significant increase in duration of eye closure; Longer duration of episodes of no body movement followed by large isolated "correction" movements.

Table 2
Summary of descriptive statistics of main explanatory variables.

Metrics	Description of the Variables	Mean	Median	S.D.
LP_stdev	Standard deviation of lateral position (m)	0.282	0.258	0.130
LP_avg	Average lateral position (m)	0.097	0.127	0.309
Speed	Average speed ($km{\cdot}h^{-1}$)	136.826	121.056	28.299
Speed_stdev	Standard deviation of speed ($km{\cdot}h^{-1}$)	2.516	1.390	3.261
PERCLOS	Percentage of eyelid closure	0.105	0.074	0.116

thresholds η_{mj} ($m = 1, 2$) and that y^*_{ij} is established by the model (see Eq. (1)). The ordered logit model can be expressed as Eq. (2) and Eq. (3).

$$Y_{ij} = \begin{cases} 1 & y^*_{ij} < \eta_{1j} \\ 2 & \eta_{1j} < y^*_{ij} \leqslant \eta_{2j} \\ 3 & \eta_{2j} < y^*_{ij} \end{cases},\ i = 1, 2, \ldots, 96;\ j = 1,\ 2,\ \ldots, 27 \tag{1}$$

$$p_{ij(m)} = P(Y_{ij} \leqslant m) = P\left(y^*_{ij} \leqslant \eta_{mj}\right) = \frac{\exp\left(\alpha_m + \beta^{Fixed} X^{Fixed}_{ij}\right)}{1 + \exp(\alpha_m + \beta^{Fixed} X^{Fixed}_{ij})},\ m = 1, 2 \tag{2}$$

$$Logit\left(p_{ij(m)}\right) = \log\left(\frac{P(Y_{ij} \leqslant m)}{1 - P(Y_{ij} \leqslant m)}\right) = \alpha_m + \beta^{Fixed} X^{Fixed}_{ij} + \varepsilon_{ij},\ \varepsilon_{ij} \overset{i.i.d.}{\sim} N\left(0, \sigma^2_\varepsilon\right),\ m = 1, 2 \tag{3}$$

where $X^{Fixed}_{ij} = \left(x^{Fixed}_{ij1}, x^{Fixed}_{ij2}, \cdots, x^{Fixed}_{ijp}\right)'$ is a $p \times 1$ vector containing the values of p explanatory variables, in which proportional odds is assumed. $\beta^{Fixed} = \left(\beta^{Fixed}_1, \cdots, \beta^{Fixed}_p\right)'$ is a $p \times 1$ vector of regression coefficients. The value of β^{Fixed} is fixed across equations, α_m represents the cut-off point for the mth cumulative logit, and ε_{ij} is the random error term.

The traditional regression model requires individual independence. However, in practical application, data is acquired through a complex, not simple, random sampling. Thus the data often presents a hierarchical structure. If a traditional regression model is used, the finiteness and statistical characteristics of the explanatory variables' estimated parameters are affected. Therefore, the mixed-effect ordered logit model (MOL) is used to more accurately explain variation of the response variables by random effects with difference, and by fixed effects without difference.

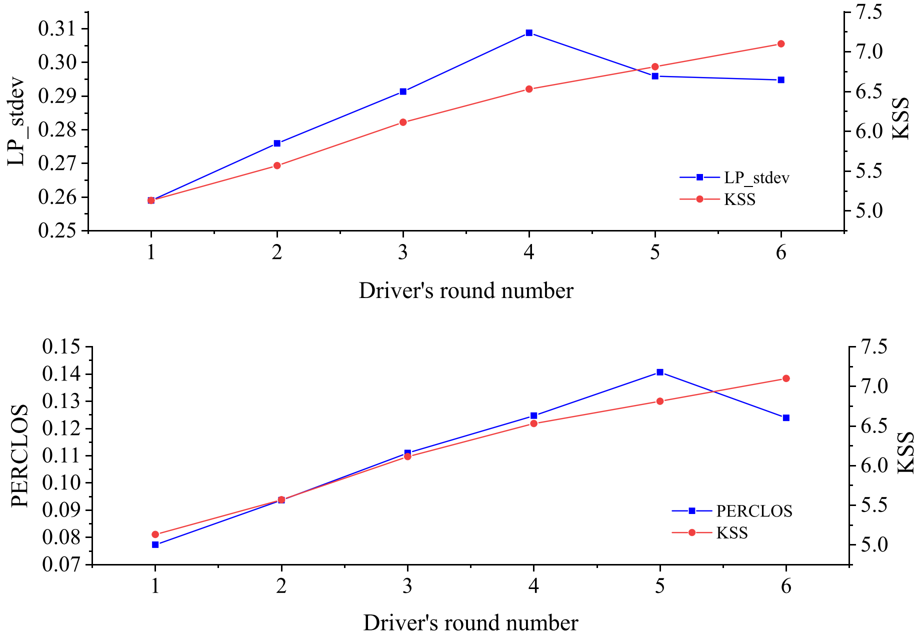

Fig. 4. Changes in average LP_stdev, PERCLOS and KSS over time (indicated by driver's number of rounds).

Table 3
Summary of the merging of KSS drowsiness levels.

Study's Merged Drowsiness Level	Proportion of Study Sample	KSS Level	Researchers' subjective assessment
1- No drowsiness	34.83%	1 Extremely alert 2 Very alert 3 Alert 4 Rather alert 5 Neither alert nor sleepy	0 Not drowsy
2- Mild to moderate drowsiness	48.11%	6 Some signs of sleepiness 7 Sleepy, but no effort to keep awake	1 Slightly drowsy 2 Moderately drowsy
3- Severe drowsiness	17.07%	8 Sleepy, but some effort to keep awake 9 Very sleepy, great effort to keep awake, fighting sleep	3 Very drowsy 4 Extremely drowsy

In a study of drowsy driving, subjects are likely to have individual differences in driving behavior and eye characteristics. Therefore, the data was defined according to driver stratification, and a random effect was introduced into the model. The above ordered logit model was thus extended to a mixed-effect ordered logit model to analyze the relationships between perceived drowsiness, driving behavior, and eye features (see Eqs. (4) and (5)).

$$p_{ij(m)} = \frac{\exp\left(\alpha_{mj}^{Random} + \beta^{Fixed}X_{ij}^{Fixed} + \beta_j^{Random}X_{ij}^{Random}\right)}{1+\exp\left(\alpha_{mj}^{Random} + \beta^{Fixed}X_{ij}^{Fixed} + \beta_j^{Random}X_{ij}^{Random}\right)},\ m = 1,2 \tag{4}$$

$$Logit\left(p_{ij(m)}\right) = \alpha_{mj}^{Random} + \beta^{Fixed}X_{ij}^{Fixed} + \beta_j^{Random}X_{ij}^{Random} + \varepsilon_{ij}, \varepsilon_{ij} \overset{i.i.d.}{\sim} N(0,\sigma_\varepsilon^2),\ m = 1,\ 2 \tag{5}$$

where α_{mj}^{Random} represents the cut-off point for the mth cumulative logit for the jth driver affected by S explanatory variables $(z_{1j},\ z_{2j},\ \cdots,\ z_{Sj})'$, which can be expressed as Eqs. (6) and (7).

$$\alpha_{mj}^{Random} = \alpha_j^{Random} + \alpha_m,\ m = 1,2;\ j = 1,2,\ldots,27 \tag{6}$$

$$\alpha_j^{Random} = \gamma_0 + \gamma_1 z_{1j} + \gamma_2 z_{2j} + \cdots + \gamma_S z_{Sj} + u_j,\ u_j \overset{i.i.d.}{\sim} N(0,\sigma_{Threshold}^2) \tag{7}$$

$X_{ij}^{Random} = \left(x_{ij1}^{Random}, x_{ij2}^{Random}, \cdots, x_{ijr}^{Random}\right)'$ is an $r \times 1$ vector containing the values of r explanatory variables which differ among drivers. $\beta_j^{Random} = \left(\beta_{j1}^{Random}, \cdots, \beta_{jr}^{Random}\right)'$ is an $r \times 1$ vector of random-effect regression coefficients, which can also be expressed as Eq. (8).

$$\beta_{jl}^{Random} = \gamma_{l0} + \gamma_{l1} z_{1j} + \gamma_{l2} z_{2j} + \cdots + \gamma_{lS} z_{Sj} + u_{jl}, \quad l = 1, 2, ..., r \tag{8}$$

The error term vector $u = \begin{bmatrix} u_{j1} \\ \vdots \\ u_{jr} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_{j1}^2 & \cdots & \sigma_{jr1}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{jr1}^2 & \cdots & \sigma_{jr}^2 \end{pmatrix}\right]$, and the covariance matrix $G \hat{=} \begin{pmatrix} \sigma_{j1}^2 & \cdots & \sigma_{jr1}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{jr1}^2 & \cdots & \sigma_{jr}^2 \end{pmatrix}$.

4.2. Mixed-effect ordered logit model considering time cumulative effect

Because this study's preliminary results showed that drowsiness is a one-directional process that accumulates gradually with time (see Fig. 3), drivers' continuous driving time is a significant variable that cannot be ignored. Therefore, the non-linear and non-decreasing function of driving time is considered in the MOL model as an explanatory variable, thereby establishing an MOL model with a time cumulative effect, or an MOL-TCE model (see Eqs. (9) and (10)).

$$p_{ij(m)} = P(Y_{ij} \leqslant m) = P\left(y_{ij}^* \leqslant \eta_{mj}\right) = \frac{\exp(\alpha_{mj}^{Random} + f(t_{[i-1]j}) + \beta^{Fixed} X_{ij}^{Fixed} + \beta_j^{Random} X_{ij}^{Random})}{1 + \exp(\alpha_{mj}^{Random} + f(t_{[i-1]j}) + \beta^{Fixed} X_{ij}^{Fixed} + \beta_j^{Random} X_{ij}^{Random})},$$
$$m = 1, 2;\ i = 1, 2, ..., 96;\ j = 1, 2, ..., 27 \tag{9}$$

$$Logit\left(p_{ij(m)}\right) = \log\left(\frac{P(Y_{ij} \leqslant m)}{1 - P(Y_{ij} \leqslant m)}\right) = \alpha_{mj}^{Random} + f(t_{[i-1]j}) + \beta^{Fixed} X_{ij}^{Fixed} + \beta_j^{Random} X_{ij}^{Random} + \varepsilon_{ij}, \varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2) \tag{10}$$

In the MOL-TCE model, $t_{[i-1]j} = t_{ij} - t_{1j}$, t_{ij} indicates the time when the jth driver passes through the ith section, so $t_{[i-1]j}$ indicates the cumulative time of driving the vehicle until the ith section observation point.

Two types of covariables affect the level of drowsiness. The first type is X_{ij}^{Fixed}, which is the explanatory variable with no driver difference and no difference in drowsiness level. The coefficient of this type of explanatory variable is recorded as β^{Fixed} and is consistent with all drivers and drowsiness levels. The second type is the explanatory variable X_{ij}^{Random} with driver differences but no difference in drowsiness level; its coefficient is denoted by β_j^{Random}.

4.3. Mixed generalized ordered response model considering time cumulative effect

Driver stratification and random effect were used in the above MOL-TCE model to handle subjects' individual differences in driving behavior and eye characteristics. However, this kind of data often presents a hierarchical structure. Consequently, this study also used a mixed generalized ordered response (MGOR) model to make a comparison. As with the MOL-TCE, data was defined according to driver stratification, a random effect was introduced, and the time cumulative effect was considered. The model was established as shown in Eq. (11).

The thresholds were ensured by using a nonlinear threshold specification where each threshold was specified by adding a non-negative term to the preceding threshold. To account for unobserved heterogeneity in the parameter estimates across observations, random parameters were included in the propensity and threshold functions as shown in Eqs. (12) and (13).

$$y_{ij}^* = \beta^{Fixed} X_{ij}^{Fixed} + \beta_j^{Random} X_{ij}^{Random} + f\left(t_{[i-1]j}\right) + \varepsilon_{ij}, \varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2) \tag{11}$$

$$\text{Threshold}: \ \eta_{1ij} = \alpha_1^{Fixed} U_{ij}^{Fixed} + \alpha_{1j}^{Random} U_{ij}^{Random} \tag{12}$$

$$\eta_{2ij} = \eta_{1ij} + \exp\left\{\alpha_2^{Fixed} U_{ij}^{Fixed} + \alpha_{2j}^{Random} U_{ij}^{Random}\right\} \tag{13}$$

where X_{ij}^{Fixed} and X_{ij}^{Random} are vectors of explanatory variables that influence y_{ij}^*, β^{Fixed} is a vector of fixed parameters, and β_j^{Random} is a vector of random parameters in the propensity function. Similarly, U_{ij}^{Fixed} and U_{ij}^{Random} are vectors of explanatory variables, and α_1^{Fixed}, α_2^{Fixed} and α_{1j}^{Random}, α_{2j}^{Random} are vectors of fixed and random parameters, respectively. ε_{ij} is an unobserved random term which is assumed to follow a known probability distribution. For identification reasons, and without loss of generality, all the parameters in η_{1ij} except a constant were set to zero $\left(\eta_{1ij} = \alpha_1^{Fixed}\right)$.

As noted above, in the hierarchical specification of thresholds in Eq. (13), where a higher order threshold η_{2ij} is specified as a sum of its preceding threshold η_{1ij} plus a non-negative random term, $\exp\left\{\alpha_2^{Fixed} U_{ij}^{Fixed} + \alpha_{2j}^{Random} U_{ij}^{Random}\right\}$ ensures that the thresholds are established in an increasing order.

The random parameter vectors β_j^{Random}, α_{1j}^{Random} and α_{2j}^{Random} are realizations from multivariate distributions $f\left(\beta_j^{Random}\right)$, $f\left(\alpha_{1j}^{Random}\right)$ and $f\left(\alpha_{2j}^{Random}\right)$. The log-likelihood LL_{ij1} and LL_{ij2} for observation j are:

$$LL_{ij1} = \int_{\beta_j^{Random}} \int_{\alpha_{1j}^{Random}} F\left[\left(\eta_{1ij}|\alpha_{1j}^{Random}\right) - \left(\beta^{Fixed} X_{ij}^{Fixed}|\beta_j^{Random}\right)\right] \cdot f\left(\beta_j^{Random}\right) f\left(\alpha_{1j}^{Random}\right) d\left(\beta_j^{Random}\right) d\left(\alpha_{1j}^{Random}\right) \tag{14}$$

$$LL_{ij2} = \int_{\beta_j^{Random}} \int_{\alpha_{2j}^{Random}} \left\{F\left[\left(\eta_{2ij}|\alpha_{2j}^{Random}\right) - \left(\beta^{Fixed} X_{ij}^{Fixed}|\beta_j^{Random}\right)\right] - F\left[\left(\eta_{1ij}|\alpha_{1j}^{Random}\right) - \left(\beta^{Fixed} X_{ij}^{Fixed}|\beta_j^{Random}\right)\right]\right\} \cdot f\left(\beta_j^{Random}\right) f\left(\alpha_{2j}^{Random}\right) d\left(\beta_j^{Random}\right) d\left(\alpha_{2j}^{Random}\right) \tag{15}$$

Considering normally distributed random parameters, which are generally employed in empirical research involving MGOR models, the thresholds can be viewed as a sum of multiple log-normally distributed random variables. The expressions for the variance of thresholds η_{1ij} and η_{2ij} with normally distributed random parameters can be written as shown in Eqs. (16) and (17).

$$Var\left(\eta_{1ij}\right) = 0 \tag{16}$$

$$Var\left(\eta_{2ij}\right) = Var\left(\exp\left\{\alpha_2^{Fixed} U_{ij}^{Fixed} + \alpha_{2j}^{Random} U_{ij}^{Random}\right\}\right) \tag{17}$$

5. Results

Correlation analysis was conducted for several of the empirical explanatory variables in the driving behavior and physiological indices. Two significant variables were obtained: standard deviation of lateral position (LP_stdev) and percentage of eyelid closure (PERCLOS), both of which met the proportional odds assumption, so the mixed-effect ordered logit model was adopted.

It is not possible to directly determine which covariate's slope in the model was the random coefficient, and we stood by the assumption that too many random coefficients can make model estimations difficult to converge. By testing the randomness and reciprocity of LP_stdev, PERCLOS, and the other explanatory variables, it was concluded that LP_stdev had a random effect while PERCLOS showed no random effects.

The cumulative effect of time on drowsiness is part of a non-decreasing function, denoted as $f(t_{[i-1]j})$. Comparing several commonly used function types, such as linear, power, exponential and logarithmic functions, the fitting effect is best when $f(t_{[i-1]j})$ takes a logarithmic function as $\lambda \ln t_{[i-1]j}$. Therefore, this study used the MOL-TCE model to judge and analyze the drowsiness levels (see Eq. (18)).

$$Logit\left(p_{ij(m)}\right) = \alpha_{mj} + \lambda \ln t_{[i-1]j} + \beta_1^{Fixed} PERCLOS_{ij} + \beta_{j1}^{Random} LP\_stdev_{ij} + \varepsilon_{ij} \tag{18}$$

where $\alpha_{mj} = \alpha_m + \alpha_j = \alpha_m + \gamma_0 + u_{j1}$, $\beta_{j1}^{Random} = \gamma_{10} + u_{j1}$. Thus Eq. (18) can be rewritten as Eq. (19).

$$\mathrm{Logit}(p_{ij(m)}) = \alpha_m + \lambda \ln t_{[i-1]j} + \beta_1^{Fixed} PERCLOS_{ij} + \gamma_{10} LP\_stdev_{ij} + \left(u_{j1} LP\_stdev_{ij} + \varepsilon_{ij}\right),\ m = 1, 2\ i = 1, 2, \dots, 96 \quad j = 1, 2, \dots, 27 \tag{19}$$

where $u = \begin{bmatrix} u_{j1} \\ \varepsilon_{ij} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_{j1}^2 & \sigma_{j12}^2 \\ \sigma_{j12}^2 & \sigma_{j2}^2 \end{pmatrix}\right]$, $Cov(\varepsilon_{ij}, u_{j1}) = 0$.

In the process of solving the model, the Cholesky decomposition method was used to estimate the variance/covariance matrix of random effects in order to ensure stability. An MOL model without $f(t_{[i-1]j})$ was established for comparison with the MOL-TCE to ascertain the effects of time accumulation on drowsiness level.

The MGOR model considering TCE was also tested to fit the driver drowsiness data, as shown in Eqs. (20)–(22).

$$y_{ij}^* = \lambda \ln t_{[i-1]j} + \beta_1^{Fixed} PERCLOS_{ij} + \beta_{j1}^{Random} LP\_stdev_{ij} + \varepsilon_{ij} \tag{20}$$

$$\text{Threshold}: \ \eta_{1ij} = \alpha_1 \tag{21}$$

$$\eta_{2ij} = \alpha_1 + \exp(\alpha_2) \tag{22}$$

When using the exhaustive method to process the MGOR-TCE, we found that a random effect did exist in the data. However, the results showed that when a random effect or random parameter was added to both the latent propensity function y_i^* and threshold function, the model could not converge. Therefore, we developed three scenarios to test, adding a random effect or parameter to only the latent propensity function y_i^* *or* threshold function, which allowed the model to converge and predict (Table 4). Among the three scenarios, S_1, S_2, and S_3, S_1, in which a random effect was added to the propensity function y_i^*, showed the best fit. It is important to note that in S_1, the MGOR-TCE was identical to the MOL-TCE model due to its structure.

The results of the parameter estimation of the MOL-TCE, MOL and the three convergent scenarios in the MGOR-TCE model are shown in Table 5.

In the MOL-TCE model, the parameter estimation values of fixed effect covariates and random effect covariates are all significant, as are the threshold values of η_1 and η_2 ($p < 0.01$). Driving time was demonstrated to be a significant predictor ($p < 0.01$) of drowsiness level. The results showed that, ceteris paribus, when ln(time) increased one unit, the probability of a driver perceiving his or her drowsiness level as high (i.e. level = 2 or 3) was 2.78 (exp(1.328)-1) times higher than the probability of perception of low drowsiness (i.e. level = 1), indicating that drowsiness level increased with increased driving time. Results, consequently, showed that LP_stdev and PERCLOS can significantly indicate drowsiness level. In the same way, ceteris paribus, when either LP_stdev or PERCLOS increased one unit, the probability of a driver perceiving a higher drowsiness level was about 7.00 (exp(2.080)-1) and 47.57 (exp(3.883)-1) times higher, respectively, than an evaluation of no drowsiness, indicating that a driver's increase in drowsiness level also showed a greater fluctuation in both eye movement and lane departure. The fitness measures of -2log likelihood, Bayesian information criterion (BIC), and Akaike information criterion (AIC) had much smaller values in the MOL-TCE model than in the MOL, which demonstrated that the MOL-TCE model was more accurate.

In the MOL model, the thresholds η_1 and η_2 were both significant ($p < 0.01$). LP_stdev and PERCLOS were shown to be significant indicators of drowsiness level, as they are in the MOL-TCE model. The covariance σ_{j12}^2 between the random intercept and random slope was not statistically significant in either the MOL-TCE or MOL model, as shown by their estimated values of less than or close to the standard error. Both the MOL-TCE and the MGOR-TCE models show three significant variables for predicting drowsiness: ln(time), LP_stdev and PERCLOS. However, in the MGOR-TCE model, the -2log likelihood, BIC, and AIC of S_1 (which is the same as the MOL-TCE model) are much smaller than that of S_2 and S_3, which indicates the MOL-TCE model performs better.

The drowsiness detection accuracy results of the MOL-TCE, MOL and MGOR-TCE models are shown in Table 6. The Subjective Drowsiness Level column represents the modified-KSS self-reported level, and the Objectively Detected Drowsiness Level columns represent the levels as detected by the physiological and behavioral variables together. As shown by the highest percentages at every level (highlighted), generally accurate correspondence is evident between the two types of evaluation. For example, in the MOL-TCE test set, when 58.07% of subjects perceived Level 3 (severe) drowsiness, the overall effects on their physiology and behavior also indicated Level 3 drowsiness, though 41.93% of self-assessments exceeded the objective evidence, which detected Level 2 (mild to moderate) and Level 1 (no) drowsiness for 35.48% and 6.45% of subjects, respectively. Among all models, the detection accuracy was highest when the drowsiness level = 1 (e.g., 71.79% accuracy in the MOL-TCE), which suggests that the models were better at detecting no drowsiness than they were at detecting either mild to moderate or severe drowsiness. Results show, however, that overall detection is more accurate in the MOL-TCE model. Comparing the MOL-TCE with the MOL model, for example, the detection of Level 1 drowsiness increased in accuracy from 56.00% in the MOL to 71.79% in the MOL-TCE model, and the detection accuracy of Level 3 drowsiness increased from 44.12% to 58.07%.

As can be seen in the accuracy percentages of the MGOR-TCE S_1, S_2 and S_3 models, prediction accuracy improved substantially after adding the random parameter to the threshold function in S_2 and S_3 (refer back to Table 5). However, because the randomness of the structure of the MGOR-TCE model itself is greater than that of the MOL-TCE, the stability of the MGOR-TCE model is reduced and the prediction accuracy is consequently lower than that of MOL-TCE. On the other hand, the detection accuracy of all three MGOR-TCE scenarios are higher than that of the MOL model, thus confirming the conclusion that models considering the cumulative effect of drowsiness with time should improve detection accuracy.

6. Discussion and conclusion

To improve early detection of driver drowsiness, this study used the Tongji University driving simulator combined with an eye-tracking system to determine which individual physiological and behavioral characteristics might be used as drowsiness detectors. Three drowsiness levels were defined based on the Karolinska Sleepiness Scale (KSS), and a mixed-effect ordered logit model was developed to consider the time cumulative effect of drowsiness (MOL-TCE). To take into account the rationality and solvability of the model, a ln $t_{[i-1]j}$ function was adopted to transform the cumulative drowsiness into the nonlinear function of driving time $f(t_{[i-1]j})$, which was then incorporated into the model. When the MOL-TCE model

Table 4
Three different scenarios to solve the MGOR-TCE model.

		S_1 (same as MOL-TCE)	S_2	S_3
Propensity function	β_{j1}	Random effect	Fixed	Fixed
Threshold function	α_1	Fixed	Random parameter	Fixed
	α_2	Fixed	Fixed	Random parameter
Convergence		Converge	Converge	Converge

Table 5
Estimation results of MOL-TCE, MOL and MGOR-TCE models.

Model	Parameters		Estimate	S.D.	t	p-value
MOL-TCE	Threshold	η_1	7.904	0.893	8.852	0.000**
		η_2	10.828	0.792	13.671	0.000**
	ln(time)		1.328	0.059	22.526	0.000**
	LP_stdev		2.080	0.255	8.144	0.000**
	PERCLOS		3.883	0.742	5.231	0.000**
	σ^2_{j1}		2.015	0.476		
	σ^2_{j2}		3.817	1.026		
	σ^2_{j12}		0.103	0.081		
	−2Log Likelihood		2707.458			
	AIC		2717.458			
	BIC		2754.461			
MOL	Threshold	η_1	0.585	0.141	4.143	0.006*
		η_2	2.991	0.345	8.664	0.000**
	LP_stdev		3.015	0.479	6.300	0.000**
	PERCLOS		4.332	0.745	5.813	0.000**
	σ^2_{j1}		4.588	1.126		
	σ^2_{j2}		6.013	2.354		
	σ^2_{j12}		0.525	0.242		
	−2Log Likelihood		4750.896			
	AIC		4758.896			
	BIC		4788.498			
		Parameters	Estimate (S.D.)			
			S_1 (same as MOL-TCE)	S_2	S_3	
MGOR-TCE	Propensity Function	ln(time)	1.328 (0.059)	1.484 (0.186)	1.106 (0.204)	
		LP_stdev	2.080 (0.255)	2.371 (0.336)	2.524 (0.452)	
		PERCLOS	3.883 (0.742)	3.620 (0.902)	3.783 (0.599)	
		$\sigma^2_{LP\_stdev}$	2.015 (0.476)	–	–	
	First threshold	Constant α_1	7.904 (0.893)	6.7105 (0.571)	7.125 (0.900)	
		$\sigma^2_{\alpha_1}$	–	5.447 (4.522)	–	
	Second threshold	Constant α_2	1.073 (0.893)	1.511 (0.713)	1.354 (0.278)	
		$\sigma^2_{\alpha_2}$	–	–	1.217 (0.587)	
	−2Log Likelihood		2707.458	2741.769	2730.147	
	AIC		2717.458	2751.769	2740.147	
	BIC		2754.461	2788.772	2777.150	

Note: $\sigma^2_{LP\_stdev}$ is the variance estimate of the random effect of LP_stdev in the propensity function; $\sigma^2_{\alpha_1}$ is the variance estimate of the random parameter α_1 in the first threshold function, and $\sigma^2_{\alpha_2}$ is the variance estimate of the random parameter α_2 in the second threshold function.
** $p < 0.01$, * $p < 0.05$.

was compared with a non-TCE MOL model, results showed that the time factor had a significant effect on the model's detection capability. The −2loglikelihood, AIC, and BIC of the MOL-TCE had lower values, which demonstrated a greater prediction accuracy at each level of drowsiness. The overall drowsiness detection accuracy of the MOL-TCE model was 62.84%, appreciably higher than the 52.47% accuracy of the non-TCE model, and also higher than the 61.04% accuracy of the MGOR-TCE model, due to the randomness of the MGOR-TCE model's structure.

Other driving behavior and physiological factors, however, corresponded well with self-reported drowsiness results, signifying a useful degree of reliability for those indicators. In both the MOL-TCE and MOL models, the percentage of driver eyelid closure (PERCLOS) was shown to be a significant indicator of drowsiness, confirming the observations of Jacobé de Naurois et al. (2017) and De Rosario et al. (2010). PERCLOS directly reflects the state of the eyes receiving external information. When a driver is drowsy, eyelid closure and blink duration are prolonged, leading to inefficiency in obtaining and processing external information, which can increase collision risk. Our results showed the standard deviation of lateral position, or drifting, was also demonstrated to be a significant indicator, results in line with other studies of drowsy driving (Vadeby et al., 2010; Wang and Xu, 2016). The standard deviation in our study increased with higher levels of drowsiness, which further confirms that the driver's stability in vehicle handling is reduced during the drowsy state.

For drowsy studies, the individual differences between drivers should be considered (Ingre et al., 2006; Wang and Xu, 2016). In this study, random effect is used to address the individual differences among drivers. However, in our MOL-TCE model, the thresholds of random effect and the proportional odds cannot be randomized at the same time due to the limited sample size. Future studies may enlarge the sample size to increase the efficiency of convergence of the random slope. In addition, the difficulty in finding female night-shift workers in China resulted in recruitment of only male drivers, which may limit generalization of the results. Future research should analyze drowsy driving among female drivers, as well as day-shift workers at the end of their work days.

Nonetheless, the findings of this study contribute to the experimental basis for the design of drowsiness detection warning systems. A driver's eye movement and vehicle data can be collected by an in-vehicle device and calculated in real time to determine the driver's drowsiness level. When the driver shows no signs of drowsiness, the device would not activate, but if

Table 6
Accuracy analysis for MOL-TCE, MOL and MGOR-TCE models.

Model	Dataset	Subjective Drowsiness Level [a]	Objectively Detected Drowsiness Level			Level Accuracy	Overall Accuracy
			1	2	3		
MOL-TCE (same as MGOR-TCE S_1)	training set	1	72.48%	26.68%	0.84%	72.48%	63.89%
		2	21.60%	58.08%	20.32%	58.08%	
		3	8.47%	28.81%	62.72%	62.72%	
	test set	1	71.79%	27.78%	0.43%	71.79%	62.84%
		2	17.54%	58.05%	24.41%	58.05%	
		3	6.45%	35.48%	58.07%	58.07%	
MOL	training set	1	53.19%	40.99%	5.82%	53.19%	50.93%
		2	30.81%	51.58%	17.61%	51.58%	
		3	34.72%	20.83%	44.45%	44.45%	
	test set	1	56.00%	38.86%	5.14%	56.00%	52.47%
		2	25.78%	52.87%	21.35%	52.87%	
		3	35.29%	20.59%	44.12%	44.12%	
MGOR-TCE (S_2)	training set	1	68.36%	28.90%	2.74%	68.36%	63.16%
		2	20.63%	59.82%	19.55%	59.82%	
		3	9.52%	28.57%	61.91%	61.91%	
	test set	1	67.80%	29.66%	2.54%	67.80%	60.47%
		2	17.48%	58.20%	24.32%	58.20%	
		3	7.41%	40.74%	51.85%	51.85%	
MGOR-TCE (S_3)	training set	1	69.06%	28.69%	2.25%	69.06%	62.05%
		2	20.84%	57.68%	21.48%	57.68%	
		3	9.09%	30.91%	60.00%	60.00%	
	test set	1	68.47%	28.83%	2.70%	68.47%	61.04%
		2	17.11%	57.04%	25.85%	57.04%	
		3	8.57%	34.29%	57.14%	57.14%	

[a] Note: Level 1 - KSS $\leq$ 5; Level 2 - KSS = 6 or 7; Level 3 - KSS $\geq$ 8.

the driver becomes mild to moderately drowsy, a Level 2 visual alert could be given, such as flashing light. If and when the driver's drowsiness reaches the severe level of 3, a combined audio and visual warning could be triggered to give the driver a stronger warning. The results of this study also confirm the increase of drowsiness over time, underscoring the need for an effective system to continue its warning until the driver responds.

Conflict of interest

None.

Acknowledgements

This study was sponsored by the National Key R&D Program of China (2018YFB0105202), the Open Project of Key laboratory of Ministry of Public Security (2019ZDSYSKFKT02-2).

References

Åkerstedt, T., Gillberg, M., 1990. Subjective and objective sleepiness in the active individual. International Journal of Neuroscience 52, 29–37.
Atiquzzaman, M., Qi, Y., Fries, R., 2018. Real-time detection of drivers' texting and eating behavior based on vehicle dynamics. Transportation Research part F 58, 594–604.
Awais, M., Badruddin, N., Drieberg, M., 2017. A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. Sensors 17 (9), 1991.
Balusu, S.K., Pinjari, A.R., Mannering, F.L., Eluru, N., 2018. Non-decreasing threshold variances in mixed generalized ordered response models: a negative correlations approach to variance reduction. Analytic Methods in Accident Research 20, 46–67.
Chipman, M., Yue, L.J., 2009. Drowsy drivers: the effect of light and circadian rhythm on crash occurrence. Safety Science 47 (10), 1364–1370.
De Rosario, H., Solaz, J.S., Rodriguez, N., Bergasa, L.M., 2010. Controlled inducement and measurement of drowsiness in a driving simulator. IET Intelligent Transport Systems 4 (4), 280–288.
De Valck, E., Cluydts, R., 2001. Slow-release caffeine as a countermeasure to driver sleepiness induced by partial sleep deprivation. Journal of Sleep Research 10 (3), 203–209.
Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention 40 (3), 1033–1054.
Expert Panel on Driver Fatigue and Sleepiness., 1998. Drowsy driving and automobile crashes. Publication DOT-HS-808-707. NHTSA, U.S. Department of Transportation.
Farahmand, B., Boroujerdian, A.M., 2018. Effect of road geometry on driver fatigue in monotonous environments: a simulator study. Transportation Research Part F: Traffic Psychology and Behaviour 58, 640–651.
He, J., Choi, W., Yang, Y., Lu, J., Wu, X., Peng, K., 2017. Detection of driver drowsiness using wearable devices: a feasibility study of the proximity sensor. Applied Ergonomics 65, 473–480.
Hu, S., Zheng, G., 2009. Driver drowsiness detection with eyelid related parameters by support vector machine. Expert Systems with Applications 36 (4), 7651–7658.

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, G., 2006. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. Journal of Sleep Research 15 (1), 47–53.
Jackson, M.L., Raj, S., Croft, R.J., Hayley, A.C., Downey, L.A., Kennedy, G.A., et al, 2016. Slow eyelid closure as a measure of driver drowsiness and its relationship to performance. Journal of Crash Prevention and Injury Control 17 (3), 251–257.
Jacobé de Naurois, C., Bourdin, C., Bougard, C., Vercher, J.L., 2018. Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness. Accident Analysis and Prevention 121, 118–128.
Jacobé de Naurois, C., Bourdin, C., Stratulat, A., Diaz, E., Vercher, J.L., 2017. Detection and Prediction of Driver Drowsiness Using Artificial Neural Network Models. Accident Analysis and Prevention. In press.
Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., et al, 2006. Validation of the Karolinska sleepiness scale against performance and EEG variables. Clinical Neurophysiology 117 (7), 1574–1581.
Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The Impact of Driver Inattention on Near-crash/Crash Risk: An Analysis using the 100-car Naturalistic Driving Study Data. Publication DOT-HS-810-594. NHTSA, U.S. Department of Transportation.
Lenné, M.G., Triggs, T.J., Redman, J.R., 1997. Time of day variations in driving performance. Accident Analysis and Prevention 29, 431–437.
Loon, R.J.V., Brouwer, R.F.T., Martens, M.H., 2015. Drowsy drivers' under-performance in lateral control: how much is too much? Using an integrated measure of lateral control to quantify safe lateral driving. Accident Analysis and Prevention 84, 134–143.
Maia, Q., Grandner, M.A., Findley, J., Gurubhagavatula, I., 2013. Short and long sleep duration and risk of drowsy driving and the role of subjective sleep insufficiency. Accident Analysis and Prevention 59, 618–622.
Martensson, H., Keelan, O., Ahlstrom, C., 2019. Driver sleepiness classification based on physiological data and driving performance from real road driving. IEEE Transactions on Intelligent Transportation Systems 20 (2), 421–430.
Mccartt, A.T., Rohrbaugh, J.W., Hammer, M.C., Fuller, S.Z., 2000. Factors associated with falling asleep at the wheel among long-distance truck drivers. Accident Analysis and Prevention 32, 493–504.
Mcdonald, A.D., Lee, J.D., Schwarz, C., Brown, T.L., 2018. A contextual and temporal algorithm for driver drowsiness detection. Accident Analysis and Prevention 113, 25–37.
Meng, F., Wong, S.C., Yan, W., Li, Y.C., Yang, L., 2019. Temporal patterns of driving fatigue and driving performance among male taxi drivers in Hong Kong: a driving simulator approach. Accident Analysis and Prevention 125, 7–13.
Merat, N., Jamson, A.H., 2013. The effect of three low-cost engineering treatments on driver fatigue: a driving simulator study. Accident Analysis and Prevention 50, 8–15.
Mollicone, D., Kan, K., Mott, C., Bartels, R., Bruneau, S., Van Wollen, M., et al., 2018. Predicting performance and safety based on driver fatigue. Accident Analysis and Prevention. In press.
Murata, A., Fujii, Y., Naitoh, K., 2015. Multinomial logistic regression model for predicting driver's drowsiness using behavioral measures. Procedia Manufacturing 3 (2), 2426–2433.
Naderi, H., Nassiri, H., Sahebi, S., 2018. Assessing the relationship between heavy vehicle driver sleep problems and confirmed driver behavior measurement tools in Iran. Transportation Research Part F-Traffic Psychology and Behaviour 59, 57–66.
NHTSA, 2017. 2016 Fatal Motor Vehicle Crashes: Overview. Publication DOT-HS-812-456. NHTSA, U.S. Department of Transportation.
Orongilad, T., Ronen, A., Shinar, D., 2008. Alertness maintaining tasks (AMTs) while driving. Accident Analysis and Prevention 40, 851–860.
Patel, M., Lal, S.K.L., Rossiter, P., Rossiter, P., 2011. Applying neural network analysis on heart rate variability data to assess driver fatigue. Expert Systems with Applications 38 (6), 7235–7242.
Qian, D., Wang, B., Qing, X., Zhang, T., Zhang, Y., Wang, X., et al, 2017. Drowsiness detection by bayesian-copula discriminant classifier based on EGG signals during daytime short nap. IEEE Transactions on Bio-medical Engineering 64 (4), 743–754.
Rodrigues, J., Vieira, F., Vinhoza, T., Barros, J., Cunha, J., 2010. A Non-Intrusive Multi-Sensor System for Characterizing Driver Behavior. International IEEE Conference on Intelligent Transportation Systems, 1620–1624.
Sparrow, A.R., LaJambe, C.M., Van Dongen, H.P.A. 2018. Drowsiness measures for commercial motor vehicle operations. Accident Analysis and Prevention. In press.
Stasi, L.L.D., Renner, R., Catena, A., Cañas, J.J., Velichkovsky, B.M., Pannasch, S., 2012. Towards a driver fatigue test based on the saccadic main sequence: a partial validation by subjective report data. Transportation Research part C 21 (1), 122–133.
Stutts, J.C., Wilkins, J.W., Scott, O.J., Vaughn, B.V., 2003. Driver risk factors for sleep-related crashes. Accident Analysis and Prevention 35 (3), 321–331.
Sunwoo, J.S., Hwangbo, Y., Kim, W.J., Chu, M.K., Yun, C.H., Yang, K.I., 2017. Sleep characteristics associated with drowsy driving. Sleep Medicine 40, 4–10.
Tefft, B.C., 2010. The Prevalence and Impact of Drowsy Driving. AAA Foundation for Traffic Safety.
Thiffault, P., Bergeron, J., 2003. Monotony of road environment and driver fatigue: A simulator study. Accident Analysis and Prevention 35 (3), 381–391.
Traffic Management Bureau of Ministry of Public Security., 2016. Statistics of Road Traffic Accidents in PRC (2015). Traffic Management Bureau of Ministry of Public Security, China.
Useche, S.A., Ortiz, V.G., Cendales, B.E., 2017. Stress-related psychosocial factors at work, fatigue, and risky driving behavior in bus rapid transport (BRT) drivers. Accident Analysis and Prevention 104, 106–114.
Vadeby, A., Forsman, Å., Kecklund, G., Åkerstedt, T., Sandberg, D., Anund, A., 2010. Sleepiness and prediction of driver impairment in simulator studies using a cox proportional hazard approach. Accident Analysis and Prevention 42 (3), 835–841.
Vanlaar, W., Simpson, H., Mayhew, D., Robertson, R., 2008. Fatigued and drowsy driving: a survey of attitudes, opinions and behaviors. Journal of Safety Research 39 (3), 303–309.
Wang, J.H., Sun, S.Y., Fang, S.E., Fu, T., Stipancic, J., 2017. Predicting drowsy driving in real-time situations: using an advanced driving simulator, accelerated failure time model, and virtual location-based services. Accident Analysis and Prevention 99, 321–329.
Wang, X.S., Xu, C., 2016. Driver drowsiness detection based on non-intrusive metrics considering individual specifics. Accident Analysis and Prevention 95, 350–357.
Wang, X.S., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis and Prevention 40, 1674–1682.
Watson, N.F., Morgenthaler, T., Chervin, R., Carden, K., Kirsch, D., Kristo, D., et al, 2015. Confronting Drowsy Driving: The American Academy of Sleep Medicine Perspective. Journal of Clinical Sleep Medicine 11 (11), 1335–1336.
Wei, C.S., Wang, Y.T., Lin, C.T., Jung, T.P., 2018. Toward drowsiness detection using non-hair-bearing EEG-based brain-computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering 26 (2), 400–406.
Wierwille, W.W., Ellsworth, L.A., 1994. Evaluation of driver drowsiness by trained raters. Accident Analysis and Prevention 26 (5), 571–581.
Wilkinson, V.E., Jackson, M.L., Westlake, J., Stevens, B., Barnes, M., Swann, P., et al, 2013. The accuracy of eyelid movement parameters for drowsiness detection. Journal of Clinical Sleep Medicine 9 (12), 1315–1324.
Wu, Y., Kihara, K., Takeda, Y., Sato, T., Akamatsu, M., Kitazaki, S., 2019. Effects of scheduled manual driving on drowsiness and response to take over request: A simulator study towards understanding drivers in automated driving. Accident Analysis & Prevention 124, 202–209.
Yang, J.H., Mao, Z.H., Tijerina, L., Pilutti, T., Coughlin, J.F., Feron, E., 2009. Analysis and detection of driver fatigue caused by sleep deprivation. IEEE Transactions on Systems, Man, and Cybernetics – Part A 39 (4), 694–705.
Yasmin, S., Eluru, N., Pinjari, A., 2015. Analyzing the continuum of fatal crashes: a generalized ordered approach. Analytic Methods in Accident Research 7, 1–15.
Young, T., Blustein, J., Finn, L., Palta, M., 1997. Sleep-disordered breathing and motor vehicle accidents in a population-based sample of employed adults. Sleep 20 (8), 608–613.