2021 IEEE Intelligent Vehicles Symposium (IV)
July 11-17, 2021. Nagoya, Japan

Analyzing Real-world Accidents for Test Scenario Generation for Automated Vehicles *

E. Esenturk*, S. Khastgir. A. Wallace, P. Jennings

2021 IEEE Intelligent Vehicles Symposium (IV) | 978-1-7281-5394-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IV48863.2021.9576007

Abstract— **Identification of test scenarios for Automated Driving Systems (ADSs) remains a key challenge for the Verification & Validation of ADSs. Various approaches including data based approaches and knowledge based approaches have been proposed for scenario generation. Identifying the conditions that lead to high severity traffic accidents can help us not only identify test scenarios for ADSs, but also implement measures to save lives and infrastructure resources. Taking a data based approach, in this paper, we introduce a novel accident data analysis method for generating test scenarios where we analyze UK's Stats19 accident data to identify trends in high severity accidents for test scenario generation. This paper first focuses on the severity of the accidents with the goal of relating it to static and time-dependent internal and external factors in a comprehensive way taking into account Operational Design Domain (ODD) properties, e.g. road, environmental conditions, and vehicle properties and driver characteristics. For this purpose, the paper utilizes a data grouping strategy (coarse-graining) and builds a logistic regression approach, derived from conventional regression models, in which emerging features become more pronounced, while uninteresting features and noise weaken. The approach makes the relationship between the factors and outcome variable more visible and hence well suited for the severity analysis. The method shows superior performance as compared to ordinary logistic models measured by goodness of fit and accounting for model variance (R^2=0.05 for the ordinary model, R^2=0.85 for the current model). The model is then used to solve the inverse problem of constructing high-risk pre-crash conditions as test scenarios for simulation based testing of ADSs.**

I. Introduction

Over the past five years, more than a half million traffic accidents have been reported in the UK [1]. Despite the traffic-safety measures taken by the UK government, there has been a steady figure of over 1700 on-road fatalities annually. Globally, 1.35 million people die due to road accidents every year [2]. 90% of these accidents have been attributed to driver error [3]. Automated Driving Systems (ADSs) have a potential of reducing the number of on-road fatalities by assisting or removing the driver from the driving task [4], [5]. Along with the potentially huge safety benefits, ADSs can provide various other benefits like improving traffic throughput, lowering emissions, increased productivity. However, in order to reap these immense benefits, it is essential to develop user trust and their acceptance of ADSs [6]. One of the key factors influencing user trust is the safety or perceived safety of the ADSs. However, due to increased complexity of ADSs with over 100 million lines of code [7], ensuring and evaluating the safety of ADSs remains a challenge [8]. It is suggested that in order to prove ADSs are safer than human drivers, they need to be driven for over 11 billion miles [9]. While this might seem like an unreasonable proposition, it is also important to highlight driving 11 billion miles on a deserted road on in sunny weather is of limited value if we want to deploy the ADSs in central London. Therefore, a gradual shift from vehicle miles travelled (VMT) to scenarios experiences in those miles (quality of miles) is becoming a more widely accepted view for evaluating safety of ADSs [10].

While scenario based testing as a verification and validation (V&V) methodology has gained traction, a key challenge of *"identifying scenarios"* remains for the research community and industry. The unbounded scenarios an ADS may experience in its lifecycle, and with the occurrence of emergent behaviour due to the complexity of the system and its interactions with the system, leads to the occurrence of *"unknown unknowns"* or *"black swan"* scenarios [11]. An approach to uncover the *black swan* scenarios or the *"unknown known"* scenarios for ADSs, an alternate approach of Hazard Based Testing (HBT) has been proposed [12]. HBT focuses on the quality of miles and suggests testing for *"how a system fails"* as compared to *"how a system works"*. Understanding how a system may fail can be either done in a proactive manner (e.g. via safety assessments involving hazard identification) [13], or in a reactive manner (e.g. by analyzing road accident databases). While the former would be intrinsic to the system, the latter would yield extrinsic factors, which may lead to hazards. HBT has three steps: 1) identification of hazards; 2) creating test scenarios for the hazards, and 3) pass criteria for the created scenarios. In order to identify hazards and corresponding scenarios, one could take two types of approaches: 1) knowledge based approach; 2) data based

* Research supported by Centre for Connected & Autonomous Vehicles (CCAV, UK), Innovate UK, UK Research & Innovation and HVM Catapult.

E. Esenturk is with WMG, University of Warwick, Coventry, CV4 7AL, UK (e-mail: E.esenturk.1@warwick.ac.uk).

S. Khastgir is with WMG, University of Warwick, Coventry, CV4 7AL, UK (e-mail: S.Khastgir.1@warwick.ac.uk.

A. Wallace is with WMG, University of Warwick, Coventry, CV4 7AL, UK (e-mail: A.Wallace@warwick.ac.uk).

P. Jennings is with WMG, University of Warwick, Coventry, CV4 7AL, UK (e-mail: Paul.Jennings@warwick.ac.uk).

approach [14]. Knowledge based approach include a system analysis using methods like STPA, FTA etc. where the ADS architecture and its interaction with external elements (vehicles, pedestrians etc.) is analyzed for failures. The data based approach can identify extrinsic factors, but requires a deep understanding of their relationship to the outcomes (e.g. frequency or severity of accidents). In a data based approach, one can use data from real-world accident databases, or from insurance claim records, or from naturalistic driving data to analyze and identify parameters (or contexts) that contribute to accidents or near-miss events. Once these *"interesting parameters/context"* are identified, they can then be used as part of the scenario-based testing approach as scenario parameters that are fuzzed during a test suite.

In this paper, we discuss the test scenario generation for ADSs using accident analysis of the UK's accident database (STATS19), which helped identify parameters and context contributing to high severity (including fatal) accidents. Subsequently, the test scenarios were formatted in a scenario definition language [15], and store them in the UK's National Scenario Database.

This paper is organized in six sections. Section 2 provides a brief review of scenarios and simulation based testing, along with the mathematical literature of the accident data analysis concentrating on regression methods, its drawbacks with a focus on severity analysis. Section 3 gives an overview of the methodology including the data format and how the data were processed into the form that was used in the analysis. Section 4 introduces our modified logistic models and discusses how they were derived from the earlier models and adapted for the severity analysis. In section 5, we discuss our findings, while section 6 discusses them in the context of scenario generation which are then utilized to systematically develop risky pre-crash scenarios. The paper ends with a conclusion in section 7.

II. Literature Review

A. Scenarios and simulation-based testing

Over the years, various definitions have been proposed for the term *"scenario"*. In an ADS context, Ulbrich et al. defined scenario as [16]:

"A scenario describes the temporal development between several scenes in a sequence of scenes. Every scenario starts with an initial scene. Action & events as well as goals & values may be specified to characterise this temporal development in a scenario. Other than a scene, a scenario spans a certain amount of time."

Based on this scenario definition, three different abstraction levels for scenarios have been proposed: function scenarios, logical scenarios and concrete scenarios [10]. Functional scenarios are abstract representation and include a description of various entities and their relations. Logical scenarios defines a functional scenario using state space variable ranges by using parameters for each entity in the scenario. Concrete scenario defines a concrete value for each parameter in the logical scenario.

The scenario abstraction hierarchy is especially useful when put in the context of the simulation based testing approach and the wider V&V framework. It is widely accepted that simulation will play a key role in V&V of ADSs. However, simulation will need to be completed by other test environments like test track testing and public-road testing in order to have confidence in the safety of the ADS. Thus, an *"evaluation continuum"* is required (ranging from simulation to real-world testing) for an ADS V&V framework. While one can run a large number of iterations of the scenarios in simulation, test track and public road testing offer limited flexibility. As a result, one can run a large number of concrete scenarios for the same logical or functional scenario in a simulation framework, to identify the parameter combination causing a failure [17], and subsequently use that combination for test track testing. As the scenarios generated as a result of the accident analysis are presumed to be high-severity accident scenarios (logical scenarios), it will be prudent to run multiple concrete scenarios (for the identified logical scenario) in simulation to gain confidence in ADS performance and its ability to deal with the scenario.

B. Accident Severity Analysis

Over the last forty years, extensive literature grew on the statistical analysis of traffic accidents investigating various issues pertaining to accident occurrences and their causes. While significant advances have been made, many of the studies, in order to address pressing safety issues, have preferred to concentrate on specific types of vehicular accidents (commercial vehicle accidents [18]; motor vehicles accidents, [19]; pedestrian accidents [20]; bicycle accidents and near misses [21]). Beginning in the 1980s, from a frequency analysis perspective, Poisson regression models began to be used [22]. One reason for this choice was that Poisson type regression models did not have the weakness of the linear models (incorrect distributional properties, inability to describe random, discrete, non-negative accident events etc.). Poisson models also provided clear linkage with the accident numbers and probability of accident occurrence.

Despite out-performing the linear counterparts in predictions, Poisson regression models also had their limitations. In particular, the requirement that variance and mean being equal was not satisfied by count data. This is known as the over-dispersion (or under-dispersion) phenomenon. The research efforts during the 1990s focused more on establishing the relation between road geometries and accident rates [23]. Gradually, to deal with the over-dispersion problem, variants of Poisson regression models, such as Poisson-lognormal regression [24] and Negative Binomial (NB) regression [25], began to be used.

Some of these later studies also aimed at measuring the contribution of other factors (such as exposure, weather and daylight) to the variation in the road accident counts. While analyzing special types of accidents are certainly important, development of comprehensive accident models with a focus on severity which can give detailed information on the accidents and the conditions of the site and other causal factors (e.g. vehicles and drivers [26]) are equally important and is one of the main aims of this paper.

Following the 2000s, severity analysis has gained considerable attention [27]. In many of these studies the severity was treated as a categorical variable [28] and hence the use of logistic regression [29]; multinomial logistic regression [30], [31]; and probit models [32], became very popular. Also in the 2000s and 2010s, alternative methods such as Bayesian regression [26], and tree-based regression [33], appeared as new ways of analyzing the accidents. Alongside these developments unobserved heterogeneity also began to be taken into consideration in analyses [34].

While most of these methods were based on analysis of aggregated data points (frequency) or individual data points (for severity analysis) and are valuable, this paper takes an intermediate route emphasizing on an approach that is capable of extracting trends from (locally) grouped accident data (coarse-graining). This requires one to modify the classical regression methods, in this case the classical logistic regression model.

III. Methodology

A. Data Collection

The raw data used in this paper is taken from the police reports in the UK (STATS19) and is publicly available [1]. There are a reported 389,238 accidents in this database during the years 2016-2018. The data are separated into different classes describing the different viewpoints, of the accidents (AccD) and of the vehicles (VehD). AccD mostly cover scenery elements while VehD provides information on the vehicle-driver characteristics. AccD of 2016 consists of 136,621 registered accidents; AccD of 2017 contains 129,982 accidents and AccD of 2018 reports 122,635 accidents. The number of entries in VehD and AccD differ as there may be more than one vehicle involved in one accident. In this paper, on the modelling perspective, one takes the view that accident severity is predominantly caused by local effects, i.e., only the factors that are immediately present at the time and location of the incident matter. The road properties, environmental conditions, vehicular factors and driver characteristics alone determine the result of the accidents. It was decided at the outset to describe the accidents from vehicles' perspective and some of the categories from the AccD and VehD were not to be included in the analysis (e.g. Local Authority District, Police Officer Attendance etc. from the AccD; Vehicle Type, Towing and Articulation, Vehicle Location etc. from the VehD). We have also discounted the effect of societal culture on accidents [35].

Secondly, for the comparability of the data, one needed to combine the AccD and VehD in a consistent way, i.e., match the cardinalities of the AccD data (>100,000 entries each year) and VehD (>200,000 entries each year). The difference arose because, in general, there can be multiple vehicles involved in one accident. Hence, common variables from the AccD (e.g. weather conditions, light conditions) were duplicated for each of the vehicles involved matching the sizes of the two dataset. Furthermore, of all accidents, only those involving one vehicle or two vehicles were selected. It was assumed that each vehicle was an independent actor (sometimes called as the vehicular chaos assumption [36]). Such an assumption is valid for dilute enough traffic and is widely used in most traffic flow models.

B. Dependent and Independent Variables

Overall, 19 traffic variables were identified. One of the 19 was regarded as dependent, and 18 variables were treated as independent. The dependent variable was accident severity. Furthermore, as an additional explanatory variable, annual average daily flow (AADF) was included in the analysis. The AADF values were extracted from the UK Department of Transport website (DfT). The data obtained were based on the averaged traffic load over the local district where the accidents happened. Therefore, the AADF values used in the analysis do not represent the ADDF of the particular road segments. Some of the dependent and independent variables that were included in the models are illustrated in Table I.

As noted earlier, in STATS 19, severity level is regarded as a local (in time) outcome of the accident. It has three degrees of severity as *slight* = 1, *serious* = 2 and *fatal* = 3. Slight severity refers to those accidents in which at least one person was slightly injured as judged by the officer. Serious accidents are those where at least one person was detained in hospital as in-patient (or equivalent level of injury). Fatal accidents refer to those in which at least one person was killed. As our analysis considers only the accidents involving one or two vehicles in which both are impacted, it is assumed that severity levels are shared by both vehicles. Also, for the purposes of this analysis, we are interested in conditions which leads to severe (serious and fatal) accidents. Thus, the two categories, *"serious"* and *"fatal"* were merged.

C. Form of the Data and Preprocessing

All of the data recorded (except AADF) was essentially categorical in form and were originally labelled somewhat subjectively (though according to the same rule throughout) with many superfluous categories. Therefore, one had to re-categorize all variables in a more ordered way as necessary. This involved merging, within the same attribute, of the categorical values which are alike or removing those that have not been attained or not interesting (e.g. non-impact accidents).

Furthermore, there were rarely instances where some attribute values were missing or unrecorded. In the raw accident data, these unknown categorical values were randomly distributed among the existing values (of the corresponding category). The glossary for the data, i.e., the categorical values corresponding to each of the aforementioned variables, are presented in Table I. The original glossary for the raw data can be found in STATS19 accident database [37].

IV. Modified Logistic Model

As the dependent variable has discrete level outcomes, a logistic regression method is well suited for the analysis. As the independent variables are almost entirely non-numerical, for the analysis one defines dummy variables corresponding to the categorical options of the original variables. This increases the number of explanatory variables from 19 to 60. To predict the odds of severe accidents (against non-severe accidents) we propose and use a modified version of the standard multiple logistic regression.

A. Ordinary Logistic Regression Models and Challenges

In this subsection, we provide the elements of bare logistic regression analysis of the accident data. The goal is to identify the factors that increase the odds of having *severe* accidents (defined as serious or fatal accidents) against slight accidents.

TABLE I.

List of Model Variables and their Values	
Severity	Slight, Severe
Time	12am-3am, 3am-6am, 6am-9am, 9am-12pm, 12pm-3pm, 3pm-6pm, 6pm-9pm, 9pm-12am
1st Road Class	Motorway, A road, B road, C/Unclassified
Carriageway Hazards	Nothing on the road, Object on the road/Prev. accident, Animal/Pedestrian in carriageway
2nd Road Class:	None, Motorway, A road, B road, C/Unclassified
Speed Limit	20 mph, 30 mph, 40 mph, 50 mph, 60 mph, 70 mph
Junction Detail	Not a junction, Private dr./Slip road, T or staggered junction, Crossroads Roundabout, More than 5 arms
Junction Location	Not a junction, Approaching junction Cleared junction, Mid junction
Light Conditions	Daylight, Darkness-lights lit Darkness - no light/lights unlit
Weather Conditions	Fine, Wind, Rain (w/o winds), Snow, Fog/Mist
Rural or Urban Area?	Urban area, Rural area
Was Vehicle Left Hand Drive?	Not left hand drive Left hand drive vehicle
Vehicle Type	Cars/Taxis, Bikes (including motorcycles) Buses/Minibuses/Trams, Horses/Agricultural vehicles, Goods
Vehicle Maneuver	Reversing, Parked/Waiting, Moving off/Going ahead, Slowing, Turning left, Turning right or U Changing lane left, Changing lane right Overtaking (nearside), Overtaking (offside)
Point of Impact	Back, Nearside, Front, Offside
Did Vehicle Leave the Carriageway	Did not leave the carriageway, Nearside, Offside
Sex of the Driver	Male, Female
Age Band of the Driver	Very young (0-20 years old) Young (21-35 years old) Mid aged (36-65 years old), Old (Over 66 years old)

Let $q(\boldsymbol{x_i})$ represent the conditional probability of occurrence of i^{th} accident given the predictor variables $\boldsymbol{x_i}$. Then, in the logistic regression model the log-odds of the outcome variables is given by

$$\log(\frac{q(x_i)}{1-q(x_i)}) = \boldsymbol{x_i} \cdot \boldsymbol{b}$$

(1),

where $\boldsymbol{b}$ is the vector of regression coefficients. To calculate $\boldsymbol{b}$ one uses the maximum likelihood estimation.

Next we discuss some of the stumbling blocks that has initiated a shift in the use of the traditional analysis methods in the field (e.g. ordinary logistic or probit models.). One of the issues is the unobserved heterogeneity in the accidents which influence regression coefficients [34]. There is also a strong element of noise, so the factors that may lead one driver to an accident will not necessarily work the same way for other drivers. In addition, the conditions recorded in the data (police reports) are subjective and can only represent the reality crudely (e.g. weather conditions fine or windy?) These unknown or partially known aspects of the accidents are substantial, and contribute to the variability in the coefficients and badly influence the predictions. Nonetheless, intuition and reason cannot deny that patterns do exist in the accidents (e.g. accidents that happen at higher speeds are more likely to cause severe outcomes). To overcome these issues a modified version of the ordinary logistic regression model is proposed below.

B. Coarse-grained Regression Model

The main proposed modification to the ordinary logistic regression of accident data is that instead of trying to fit a microscopic model, which tries to make predictions for single incidents, one zooms out, and try developing an *effective* model which can explain the patterns or trends in accidents. Such approaches are commonly used in statistical physics [36] to account for collective properties of the matter under interest. To this end, we first start by ordering the data with respect to the outcome variable (the severity) in descending order (severe=1, non-severe=0). This is an important step and the reasons for it will become clear later. Next, the data are divided into bundles of chosen size(s). Each bundle is to represent a single point in the "coarse-grained data". Then a *rule* will be needed to assign the new data their respective new values. As the *rule, w*e set, the average value of the variable (in the bundle) as its value, i.e., $\boldsymbol{z_i} = \sum_{k=0}^{l} \boldsymbol{x_{i,k}} / l$ where 'l' is the chosen bundle size and 'i' is the bundle index.

Once the rule for value assignment to the bundles is set, the modified regression equation for the bundles can be written as

$$\log(\frac{Q(z_i)}{1-Q(z_i)}) = \boldsymbol{z_i} \cdot \boldsymbol{a}$$

(2),

where $Q(\boldsymbol{z_i})$ represents the probability of accident occurrence for the bundle. It should be noted that the form (2) is a first order approximation and can be improved by using a more appropriate nonlinear forms on the right hand side.

Given a training set, the process of computing the regression coefficients was done via logistic lasso regression algorithm. By introducing an additional term $\sum_{k=0}^{n} \lambda|\boldsymbol{z_i}|$ to the cost function (λ being a free parameter),

$$C(a) = \sum_{i=1}^{N} y_i^c log(Q(\boldsymbol{z_i})) + \sum_{i=1}^{N}(1-y_i^c)\big(1-log(Q(\boldsymbol{z_i})\big) + \sum_{k=1}^{N} \lambda|\boldsymbol{z_i}|$$

(3),

where y_i^c are the averaged values (over the bundle) for the dependent variable. This algorithm penalizes for the high variance and hence particularly suits models, like ours, with high number of attributes which usually suffer from overfitting. Another reason for this choice is that, due to the large number of explanatory attributes, when coarse graining is applied, certain attributes will have reduced influence while others will be more pronounced (hence an effective model). Lasso regression can serve to systematically remove the less important attributes from the model. In our problem, the free parameter λ was set to the value (λ_{min}) which minimized.

To implement the method (after ordering the data with respect to the severity outcome) one needs to decide on a reasonable bundle size for the coarse-graining procedure. In a way this is an optimization task. If a too small bundle size is chosen the data noise will not be removed sufficiently. Conversely if a very large bundle size is chosen much of the

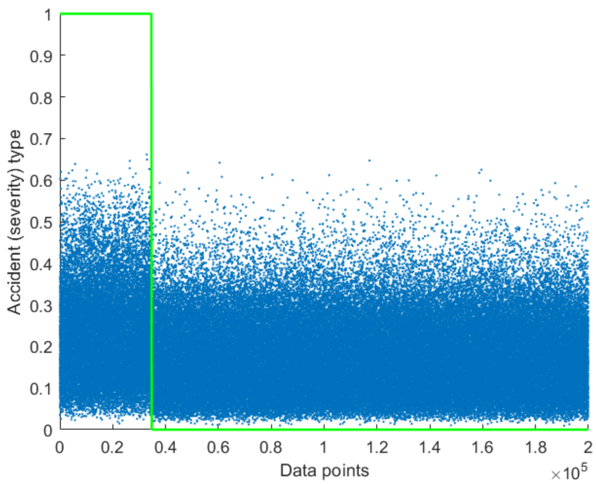

Figure 1. Accident outcome predictions with ordinary logistic regression analysis

Figure 2. Accident outcome predictions with coarse-grained logistic regression analysis

information in the original data will be lost and there will be too few (coarse-grained) "data points" to carry out the statistical analysis. Experiments with very large bundle sizes also exhibit an increased fluctuation for the predicted values, which is a sign of instability. By varying bundle size from 10 to 100, we determined a *sweet region* where model performs well. The appropriate bundles sizes range around 20 to 60 individual data points. While a bundle size of 20-60 might be specific to this dataset, in principle, if the original data size was larger, it would be possible to choose bigger bundle sizes. In the numerical analysis below the chosen bundle size for the modified logistic models were set as 40.

V. Results and Discussion

In this section, we show that the proposed model eliminates the previously noted drawbacks (Section IV) of the ordinary logistic regression model and has a much higher predictive power due to the incorporation of the novel modifications via coarse graining and implementation of lasso regression. For the purposes of this article we are interested in the comparison of slight versus non-slight (or severe) accidents.

A. Analyses

Of the three years of accident data recorded in the years 2016-2018 randomly chosen ~350,000 points were made as the training set from which the regression coefficients were calculated. To determine the standard errors in the computed coefficients we used bootstrap method and formed samples (100) of training set by random sampling with replacement and then averaged them. The computations were carried out using the MATLAB software (MATLAB, 2019).

For the ordinary logistic model, the log-likelihood was found as $LL = 1.75 \times 10^5$ showing that ordinary model was significantly different from the null model $LL_0 = 1.85\ 10^5$ ($p < 0.01$). Also, the regression coefficients were found to be significant ($p < 0.01$). However, the predictive power of the model is quite poor (as can be seen from Figure 1).

For the goodness of fit of the regression model, the sum of Pearson residual squares, i.e., $\sum_{i=1}^{N} \frac{(y_i - q(x_i))^2}{q(x_i)(1-q(x_i))}$ was computed. Under the null hypothesis that model fits well Pearson test shows that model does not fit well ($p < 0.01$). Furthermore, Hosmer-Lemeshow (HL) goodness of fit tests was also conducted (splitting the outcomes into 10 groups) also affirmed *not* good fit ($p < 0.01$). As an additional measure goodness of fit, the ratio of wrongly classified outcomes over the total number of cases was also calculated

$$w = \frac{\sum_{i=1}^{N}[|q(x_i) - y_i|]}{N} \qquad (2),$$

where [.] represents the nearest integer value. For the ordinary model it was found that $w = 0.27$, a large fraction of misclassified outcomes. Finally, the model is also weak in accounting for data variation, which can be seen from computing (Mc Fadden) pseudo R-squared value ($R^2 = 0.05$).

With the coarse-graining procedure, the model is substantially improved. Of the 60 regression coefficients computed only the variable *3am-6am* was found to be not statistically significant ($p > 0.05$) while all other coefficients were statistically significant ($p < 0.01$). The log-likelihood of the model (after removing *3am-6am* from the model was found to be $LL = 603$ which shows that model is significant compared to the null model $LL_0 = 4618$ ($p < 0.01$). The table showing the full results (the frequencies, coefficients, standard errors and p values) is provided in the supplementary materials. The variables with the most contributing coefficients are listed below in Table II.

For the model's goodness of fit (not including the non-significant variable *(3am-6am)* one finds that both Pearson test and HL test show that there is no evidence of bad fitting ($p = 1.00$). This can also be seen from the substantial reduction in the number of wrongly classified outcomes ($w = 0.04$) and the plot of predicted and observed outcomes (Figure 2). Furthermore, the modified model also accounts for the data variation much better than the ordinary model ($R^2 = 0.85$).

B. Discussion

As we emphasized in the introduction, the main motivations of this paper is the identification of the test scenarios (temporal/spatial conditions) for ADSs that are correlated with high-severity accident outcomes. In this section, we demonstrate systematic generation of test scenarios for ADSs.

Before the generation of the scenarios for ADSs, we first discuss, qualitatively, the findings as summarized in Table II. Concerning road related variables, of all *1st Road Class* categories, the accidents on type *B roads* seem to increase the odds of a severe accident most. This might be because while vehicles generally drive faster on Motorways or A roads, possibly more accidents take place on B roads (due total length of B roads being longer). Secondly, contrary to the expectation the objects on the road seem to have a negative effect on the severity compared to empty road. This might be because an object can serve as a sign and make the drivers more alert. For the speed limit, roads with 60 mph limit seem to cause the higher severity accidents. For junctions, on the other hand, roundabouts seem to be the safest type.

For the environmental effects, one can infer that absence of *light* negatively contributes to the severity likelihood of accidents. This is understandably due to the reduced visibility. On the other hand, among the *Weather Conditions* elements, *snowy* weather seems to cause the accident to be less severe compared to other conditions. This might be because snowy weather leads to drivers' lowering their speeds.

Concerning vehicle characteristics, the results suggest that left hand vehicles are safer. This may be because the drivers of these vehicles likely to pay extra attention to the traffic due to opposite positioning of the driver with respect to traffic flow. This can also be due to considerably low frequency of left hand vehicles in the UK traffic. Secondly, among the vehicle types, bikes (including motored and non-motored) are more likely to experience more severe accidents, which can be attributed to the inherent vulnerability of bikes.

For the driver behavior and characteristics, one can also make several interesting conclusions. Concerning the manoeuvre taken by the drivers, the highest severity probability increasing maneuvers are *right turns* (including *U-turns*) while least concerning manoeuvre is *changing lane left*. As a natural result of such manoeuvres, the most severity causing impact types are seen to be *frontal* and *offside.* Expectedly, if *vehicle leaves the carriageway at offside*, then such accidents, contribute to higher severity. From a vehicle behaviour perspective, this is an interesting finding for test scenario generation suggesting that test scenarios should focus more on *right turn* manouvres. Furthermore, concerning the driver profile, *female* drivers distinct themselves to be more careful drivers as their influence to severity was negative. For the age band of the drivers, old drivers are at increased risk of experiencing more severe accidents.

C. Systematic Test-scenario Generation

A systematic way of generating *"interesting"* test scenarios based on the accident data analysis is as follows. Instead of finding/determining the outcome (i.e., the severity class) of a particular accident, we ask the inverse question: for a given accident which resulted in a severe outcome, what were the possible combination of pre-crash conditions that lead to it?

Additionally, what are the conditions that lead to increased probability of severe accidents? These questions are inverse problems, which, in our context, have many answers. To be more precise, for the logistic model the aforementioned question corresponds to finding solutions of the equation

$$\log\left(\frac{Q}{1-Q}\right) = \boldsymbol{z} \cdot \boldsymbol{a} \qquad (5)$$

where Q is a value of choice above a threshold probability.

For consistency with the model, we need to restrict the values of $\boldsymbol{z}$ to lie in the hyper-cube $[0,1] \times \ldots \times [0,1]$ keeping in mind that numerical closeness each component z_k to 0 implies the absence of the particular pre-crash condition that it represents. Similarly, numerical closeness of each z_k ($k \in 1, \ldots, \mathrm{p}$) to 1 is interpreted as that particular condition is present at the time of the accident. Then the remaining question is to find the distinct solutions that Equation 5 has.

TABLE II.

Key Regression Coefficients, Errors and Significance Levels				
Attributes	**Most contributing variables**	**Freq**	**Coeff. (a/40), (with SE)**	**P value**
Intercept			-0.63 (0.02)	< 0.001
Time	9am-12pm	0.14	-0.42 (0.01)	< 0.001
1st Road Class	B road	0.12	0.52 (0.01)	< 0.001
Carriageway Hazards	Animal/Ped. in carriageway	0.01	-0.42 (0.02)	< 0.001
Speed Limit	60 mph	0.12	0.64 (0.01)	< 0.001
Junction Detail	Roundabout	0.11	-0.49 (0.02)	< 0.001
Junction Location	Cleared junction	0.10	0.35 (0.01)	< 0.001
Light Conditions	Darkness - no light/lights unlit	0.06	0.20 (0.01)	< 0.001
Weather Conditions	Snow	0.02	-0.41 (0.01)	< 0.001
Was Vehicle Left Hand Drive?	Left hand drive vehicle	0.02	-0.61 (0.01)	< 0.001
Vehicle Type	Bikes	0.18	0.81	< 0.001
Vehicle Maneuver	Turning right / U	0.12	0.09 (0.01)	< 0.001
	Change lane left	0.01	-0.59 (0.01)	< 0.001
Point of Impact	Front	0.54	0.54 (0.01)	< 0.001
Did Vehicle Leave the Carriageway	Offside	0.05	0.33 (0.01)	< 0.001
Sex of the Driver	Female	0.31	-0.19	< 0.001
Age Band of the Driver	Old (Over 66 years old)	0.10	0.45 (0.01)	< 0.001
Log-likelihood = 603, Mc Fadden $R^2 = 0.85$				
$\lambda_{min} = 3.28 \times 10^{-6}$				

a. The table summarizes the analysis output for the more important variables.

For practical purposes, we restrict ourselves to integer solutions which are finitely many and more indicative of the pre-crash conditions (existing condition = 1, non-existing conditions = 0), which directly correspond to real distinct world scenarios for a given accident. Below are two such examples scenarios.

<u>More likely severe outcome test-scenario 1 ($Q > 0.9$)</u>:

On a morning around *9:00am-12:00pm* in *daylight,* a *young male* had an accident while he was driving a *goods* vehicle on a *B road* in an *urban* area with *speed limit 40*. He was *at a T-junction* and was *turning right.* While *entering junction,* he was hit from *offside.* The weather at the time was *fine with no winds.*

<u>More likely severe outcome test-scenario 2 ($Q > 0.9$)</u>:

On an evening around *6:00pm-9:00pm,* a *mid aged male* driver had an accident while he was driving a *bus* on a *Motorway in rural area* with *speed limit 60*. He was not *near a junction and there was an object on the road.* It was dark but road lights *were lit.* He was *overtaking on the nearside and* hit another vehicle by its *front.* As a result of the collision *the vehicle left the carriageway from nearside.* The weather at the time was *snowy.*

At first sight, the above scenarios might seem common and not give enough insights to detail the specific dangerous situations that are to be avoided. However, as discussed earlier in this section, a scenario, which is categorized as severe, does not imply that every accident under those conditions will be severe. Rather, it tells us that on average such conditions generate severe outcomes. This is especially powerful in a simulation-based testing framework where multiple instances of a scenario are executed with different parameter values for the scenario parameters. Furthermore, it is important to note that our analysis is limited by the coverage and details of the recorded information. Since the police records do not contain all the fine details of the accidents (adding to the possibility of incomplete records) [38], the conclusions that can be drawn from them are also limited. However, future studies will involve applying the method in a more general setting with more finely recorded data (e.g. telematics data recorded on vehicles or insurance claim records).

D. Scenario based testing and safety evaluation

Scenario generation based on accident data is an extrinsic method which focuses on the influence of external factors on the severity of the accidents. This is especially useful when put in the context of an Operational Design Domain (ODD) [39]. An ODD of an ADS describes the operating environment of an ADS and include scenery (geostationary attributes), environment (weather, connectivity etc.) and dynamic elements (pedestrians, other vehicles etc.). In general, explanatory variables in the scenario generated as result of the accident data analysis involving scenery, environment, dynamic and other elements can be more than the ones presented in the earlier section. Therefore, it is desirable to extract the *relevant* scenarios from the rest which can be considered as a focussed set of scenarios for ADS simulations. This can be done by further exploiting the lasso regression to yield the most important features of the scenarios. Additionally, one can map the ODD defined for an ADS on to the attributes of the scenarios generated from the analysis to identify an overlap, which will further filter the set of relevant scenarios. From a safety validation perspective, it is essential to that the ADS is tested in the scenarios it will encounter either inside its ODD or that the boundaries of the ODD. Therefore, using an extrinsic method like the method proposed in this paper enables for an efficient selection of scenarios, even post the analysis of the accident data.

From a simulation-based testing perspective, suppose one desires to incorporate into a real time simulation package a predetermined number of attributes, say p_{max}, from the severity models developed in this paper. By increasing the value of the free parameter λ in the developed coarse-grained lasso regression models (logistic) one can select the most important features. For instance, of the nearly 60 attributes considered in this paper, p_{max} most important features are to be used in a simulation run. Then, by letting λ increase in the model, one reaches to a point where only p_{max} attributes remain positive. This means the remaining attributes represent the conditions that can be used in the simulation run. Additionally, in order to identify concrete scenarios, i.e., values for the scenarios parameters which cause the ADS to fail, one could use algorithms like Bayesian Optimization for concretization [17].

Using the proposed scenario generation approach together with simulation based testing, one can help establish the true capabilities and limitations of an ADS, which can then be imparted to the user and help create a state of *"informed safety"* [6]. Informed safety is key for ensuring safe use and appropriate trust in ADSs, and help prevent any misuse or disuse of the system, ultimately leading to the society reaping the benefits of automation.

VI. Conclusion

Scenario based testing has been widely accepted as a key enabler in the safety assurance of ADSs. However, generation relevant scenarios remains a key challenge for the industry and the research community. Various approaches to scenario generation include: 1) knowledge based approach (identifying hazards and failure modes of a system); 2) data based approach (analyzing accident data, insurance claim records or naturalistic driving data. In this paper, we analyzed the police reports in the UK accident database (STATS19) for the years 2017, 2018 to identify the causal relationship between the pre-crash accident conditions and accident severity in a way that takes into account the major factors leading to the severe and risky outcomes. Our main motivation was to develop an effective accident model which is able to capture time dependent effects on ODD parameters such as (weather conditions, light conditions, traffic exposure etc.) thereby complementing the existing studies which generally focus on trends over a long time. We developed a modified logistic regression model which was capable of analysing the categorically recorded data. This required, after appropriately pre-processing (cleaning, reformatting), developing and interpreting coarse-grained versions of the standard regression models in order to extract the useful information from the noise. The calculated severity values and conditional probabilities of the major accidents were found to be in good agreement with the test data.

The second main goal of this paper was the development of a systematic way to generate pre-crash scenarios that pose high risk. For this purpose, the logistic regression equations were used in a similar spirit to solving an inverse problem, to produce the pre-crash conditions that lead to fatal (or serious) accidents. Application of the procedure demonstrated that many scenarios can be generated in an automated fashion which can potentially provide useful test cases for ADSs and can be compared easily with a defined ODD to identify relevant scenarios for any ADS with a defined ODD. In summary, the method can serve as a useful tool for development of realistic test scenarios and in the wider simulation based V&V framework. While the data used in this paper (police records) had intrinsic limitations due to the lack of precision measurement, the proposed methods are robust and can be applied to more finely recorded data to extract more detailed conclusions about accidents and pre-crash scenarios.

Acknowledgment

The work presented in this paper has been carried under the Innovate UK and Centre for Connected and Autonomous Vehicles (CCAV) funded OmniCAV project (Grant No. 104529). This work is also supported by UKRI Future Leaders

Fellowship (Grant MR/S035176/1). The authors would like to thank the WMG centre of HVM Catapult and WMG, University of Warwick, UK, for providing the necessary infrastructure for conducting this study.

REFERENCES

[1] "Road Safety Data - STATS19," *UK Department for Transport*, 2020. https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data (accessed Apr. 02, 2020).

[2] WHO, "Global status report on road safety 2015," 2015. http://bit.ly/2Q7F3SQ (accessed Oct. 28, 2018).

[3] S. Singh, "Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. (Traffic Safety Facts Crash Stats. Report No. DOT HS 812 115)," Washington, DC, 2015.

[4] M. Guériau, R. Billot, N. E. El Faouzi, J. Monteil, F. Armetta, and S. Hassas, "How to assess the benefits of connected vehicles? A simulation framework for the design of cooperative traffic management strategies," *Transp. Res. Part C Emerg. Technol.*, vol. 67, pp. 266–279, 2016, doi: 10.1016/j.trc.2016.01.020.

[5] C. Tingvall, "The Zero Vision: A Road Transport System Free from Serious Health Losses," *Transp. Traffic Saf. Heal. New Mobil.*, pp. 37–57, 1997.

[6] S. Khastgir, S. Birrell, G. Dhadyalla, and P. Jennings, "Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles," *Transp. Res. Part C Emerg. Technol.*, vol. 96, pp. 290–303, 2018, doi: 10.1016/j.trc.2018.07.001.

[7] R. N. Charette, "This Car Runs on Code," *IEEE Spectrum*, Feb. 2009.

[8] S. Khastgir, S. Birrell, G. Dhadyalla, and P. Jennings, "Identifying a gap in existing validation methodologies for intelligent automotive systems: Introducing the 3xD simulator," in *Proc. of the IEEE Intelligent Vehicles Symposium 2015*, 2015, pp. 648–653.

[9] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?," *Transp. Res. Part A Policy Pract.*, vol. 94, no. December, pp. 182–193, 2016, doi: 10.1016/j.tra.2016.09.010.

[10] T. Menzel, G. Bagschik, and A. M. Maurer, "Scenarios for Development, Test and Validation of Automated Vehicles," *IEEE Intell. Veh. Symp. Proc.*, vol. 2018-June, no. Iv, pp. 1821–1827, 2018, doi: 10.1109/IVS.2018.8500406.

[11] T. Aven, "Implications of black swans to the foundations and practice of risk assessment and management," *Reliab. Eng. Syst. Saf.*, vol. 134, pp. 83–91, 2015, doi: 10.1016/j.ress.2014.10.004.

[12] S. Khastgir, S. Birrell, G. Dhadyalla, and P. Jennings, "The Science of Testing: An Automotive Perspective," 2018, doi: 10.4271/2018-01-1070.

[13] S. Khastgir, S. Brewerton, J. Thomas, and P. Jennings, "Systems Approach to Creating Test Scenarios for Automated Driving Systems," *Reliab. Eng. Syst. Saf.*, doi: doi.org/10.1016/j.ress.2021.107610.

[14] M. Brackstone *et al.*, "OmniCAV: A Simulation and Modelling System that enables 'CAVs for All,'" 2020.

[15] X. Zhang, S. Khastgir, and P. Jennings, "Scenario Description Language for Automated Driving Systems: A Two Level Abstraction Approach," 2020.

[16] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, "Defining and Substantiating the Terms Scene , Situation , and Scenario for Automated Driving," 2015, doi: 10.1109/ITSC.2015.164.

[17] B. Gangopadhyay, S. Khastgir, S. Dey, P. Dasgupta, G. Montana, and P. Jennings, "Identification of Test Cases for Automated Driving Systems Using Bayesian Optimization," pp. 1961–1967, 2019.

[18] M. Osman, S. Mishra, and R. Paleti, "Injury severity analysis of commercially-licensed drivers in single-vehicle crashes: Accounting for unobserved heterogeneity and age group differences," *Accid. Anal. Prev.*, vol. 118, no. May, pp. 289–300, 2018, doi: 10.1016/j.aap.2018.05.004.

[19] S. P. Miaou and H. Lum, "Modeling vehicle accidents and highway geometric design relationships," *Accid. Anal. Prev.*, vol. 25, no. 6, pp. 689–709, 1993, doi: 10.1016/0001-4575(93)90034-T.

[20] C. Lee and M. Abdel-Aty, "Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida," *Accid. Anal. Prev.*, vol. 37, no. 4, pp. 775–786, 2005, doi: 10.1016/j.aap.2005.03.019.

[21] L. B. Meuleners, M. Fraser, M. Johnson, M. Stevenson, G. Rose, and J. Oxley, "Characteristics of the road infrastructure and injurious cyclist crashes resulting in a hospitalisation," *Accid. Anal. Prev.*, vol. 136, no. August 2019, 2020, doi: 10.1016/j.aap.2019.105407.

[22] P. P. Jovanis and H. L. Chang, "Modeling the Relationship of Accidents To Miles Traveled.," *Transportation Research Record*. pp. 42–51, 1986.

[23] D. Lord, A. Manar, and A. Vizioli, "Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments," *Accid. Anal. Prev.*, vol. 37, no. 1, pp. 185–199, 2005, doi: 10.1016/j.aap.2004.07.003.

[24] J. Ma, K. M. Kockelman, and P. Damien, "A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods," *Accid. Anal. Prev.*, vol. 40, no. 3, pp. 964–975, 2008, doi: 10.1016/j.aap.2007.11.002.

[25] S. R. Geedipally and D. Lord, "Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models," *Accid. Anal. Prev.*, vol. 42, no. 4, pp. 1273–1282, 2010, doi: 10.1016/j.aap.2010.02.004.

[26] C. Chen *et al.*, "Driver injury severity outcome analysis in rural interstate highway crashes: a two-level Bayesian logistic regression interpretation," *Accid. Anal. Prev.*, vol. 97, pp. 69–78, 2016, doi: 10.1016/j.aap.2016.07.031.

[27] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives," *Accid. Anal. Prev.*, vol. 43, no. 5, pp. 1666–1676, 2011, doi: 10.1016/j.aap.2011.03.025.

[28] A. Khorashadi, D. Niemeier, V. Shankar, and F. Mannering, "Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory analysis," *Accid. Anal. Prev.*, vol. 37, no. 5, pp. 910–921, 2005, doi: 10.1016/j.aap.2005.04.009.

[29] K. A. Krull, A. J. Khattak, and F. M. Council, "Injury effects of rollovers and events sequence in single-vehicle crashes," *Transp. Res. Rec.*, no. 1717, pp. 46–54, 2000, doi: 10.3141/1717-07.

[30] V. Shankar and F. Mannering, "An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity," *J. Safety Res.*, vol. 27, no. 3, pp. 183–194, 1996, doi: 10.1016/0022-4375(96)00010-2.

[31] G. F. Ulfarsson and F. L. Mannering, "Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents," *Accid. Anal. Prev.*, vol. 36, no. 2, pp. 135–147, 2004, doi: 10.1016/S0001-4575(02)00135-5.

[32] C. Lee and M. Abdel-Aty, "Presence of passengers: Does it increase or reduce driver's crash potential?," *Accid. Anal. Prev.*, vol. 40, no. 5, pp. 1703–1712, 2008, doi: 10.1016/j.aap.2008.06.006.

[33] L. Y. Chang and H. W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques," *Accid. Anal. Prev.*, vol. 38, no. 5, pp. 1019–1027, 2006, doi: 10.1016/j.aap.2006.04.009.

[34] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Anal. Methods Accid. Res.*, vol. 11, pp. 1–16, 2016, doi: 10.1016/j.amar.2016.04.001.

[35] R. C. McIlroy *et al.*, "Who is responsible for global road safety? A cross-cultural comparison of Actor Maps," *Accid. Anal. Prev.*, vol. 122, no. June 2018, pp. 8–18, 2019, doi: 10.1016/j.aap.2018.09.011.

[36] D. Helbing, "Traffic and related self-driven many-particle systems," *Rev. Mod. Phys.*, vol. 73, no. 4, pp. 1067–1141, 2001, doi: 10.1103/RevModPhys.73.1067.

[37] Department for Transport - HM Government, "Reported road casualties in Great Britain: 2018 annual report," 2019. doi: 10.1007/s10549-016-3800-5.

[38] M. Imprialou and M. Quddus, "Crash data quality for road safety research: Current state and future directions," *Accid. Anal. Prev.*, vol. 130, pp. 84–90, 2019, doi: 10.1016/j.aap.2017.02.022.

[39] SAE International, "Surface Vehicle Recommended Practice - J3016," 2018.