



RESEARCH ARTICLE

Microscopic traffic models, accidents, and insurance losses

Sojung Kim, Marcel Kleiber  and Stefan Weber 

Institute of Actuarial and Financial Mathematics & House of Insurance, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany

Corresponding author: Marcel Kleiber; Email: marcel.kleiber@insurance.uni-hannover.de

Received: 29 March 2023; **Revised:** 11 November 2023; **Accepted:** 13 November 2023

Keywords: Microscopic traffic models; car-following models; SUMO; digital twins; insurance premiums

Abstract

The paper develops a methodology to enable microscopic models of transportation systems to be accessible for a statistical study of traffic accidents. Our approach is intended to permit an understanding not only of historical losses but also of incidents that may occur in altered, potential future systems. Through such a counterfactual analysis, it is possible, from an insurance, but also from an engineering perspective, to assess the impact of changes in the design of vehicles and transport systems in terms of their impact on road safety and functionality.

Structurally, we characterize the total loss distribution approximatively as a mean-variance mixture. This also yields valuation procedures that can be used instead of Monte Carlo simulation. Specifically, we construct an implementation based on the open-source traffic simulator SUMO and illustrate the potential of the approach in counterfactual case studies.

1. Introduction

Every year, traffic accidents cause substantial damage, both property damage and injuries and deaths. For example, nearly 43,000 people died in road traffic accidents in the USA in 2021 (NHTSA, 2022). The frequency and severity of these accidents depend on the driving behavior of vehicles, on the one hand, and on the characteristics of traffic systems themselves, on the other. Improvements in road safety are achieved, for example, by reducing serious injuries in accidents through the design of vehicles, by car body design, airbags, seatbelts, etc. However, the frequency and type of accidents can also be influenced by modifying the transportation system itself and by changes in driving behavior. From a higher-level perspective, at least two dimensions are central, and we will examine them in this paper:

- (i) *Engineering.* From an engineering perspective, the focus is on the good design of vehicles and traffic systems, combining functionality and safety. Instruments in this respect include traffic rules and their implementation, the layout of streets, and innovation in vehicle technology such as advanced driver assistance systems or autonomous driving software. Improvements of this type may reduce the number of accidents and their severity but cannot completely prevent accidents.
- (ii) *Insurance.* Residual risks remain, and accidents cannot be completely prevented. However, at least in financial terms, the associated losses can be covered by insurance contracts. The role of actuaries is to develop adequate contract structures, calculate correct premiums, and implement quantitative risk management in insurance firms. These tasks require the modeling and analysis of probability distributions of accident frequencies, corresponding damages, and insurance losses.

The objective of this paper is to develop a methodology to enable microscopic models of transportation systems to be accessible for a statistical study of traffic accidents. Our approach is intended to permit an understanding not only of historical losses but also of incidents that may occur in altered, potential future systems. Through this, it is possible, from both an engineering and insurance perspective, to assess changes in the design of vehicles (e.g., the driving behavior of autonomous vehicles) and transport systems in terms of their impact on functionality and road safety. This is in stark contrast to simply considering aggregate data, as is typical in auto insurance. Instead our model can simulate traffic in counterfactual situations that mimic modified or future scenarios, and losses can be generated conditionally on local traffic conditions.

The conventional approach is as follows: to understand current traffic events or structural relationships in the past, historical data are used. Historical data can also be applied to test whether a model framework is appropriate in principle to describe traffic systems realistically. These data also constitute the essential basis for the specific pricing of insurance contracts in practice.

But how can we examine risks associated with new technologies and with novel future strategies for traffic systems? Consider autonomous vehicles, for example: due to their altered driving behavior, these will reshape existing traffic patterns, and in turn, accident occurrences and associated losses. Insurance companies presumably will have to adapt their business models as well; in the future, premiums for auto insurance may depend on the driving configuration of the vehicle rather than the risk profile of the driver. Real-time insurance rates for individual trips could also become important, with prices depending on the current traffic situation.

In order to investigate future developments, we are suggesting to devise simulation tools in analogy to digital twins of real transport systems, which allow counterfactual case studies of possible future transport systems. The digital twin paradigm refers to the triad of a “physical entity, a virtual counterpart, and the data connections in between” (Jones *et al.*, 2020). In our application, the physical entity is the (future) real-world transportation system for which data on losses are not yet available. Its virtual counterpart is the model we are building. Counterfactual case studies can be used to generate data and evaluate future driving technologies and their impact on accident losses. Based on the results, newly developed concepts (e.g., modified traffic rules, novel insurance coverage and their insurance premiums, etc.) can be adapted in the real world. The concept of the digital twin makes it possible to experiment with technologies and policies, and their effects on accident damage without having to implement risky tests in reality.

Methodologically, this paper combines existing microscopic traffic models with probabilistic tools from actuarial science and quantitative risk management to study accident damage and insurance losses in the context of simulations. In contrast to standard insurance practice, we model losses conditionally on specific traffic situations. These traffic situations are generated within a flexible micro-simulation model. As a specific example, we use the well-established traffic simulator SUMO (Lopez *et al.*, 2018) to illustrate how traffic systems could be realistically modeled.

We extend microscopic traffic models to include random accidents and corresponding losses. The losses are modeled as random variables whose distributions depend on microscopic data. Since insurance contracts typically cover annual periods, we set up a model for aggregate losses over a 1-year time horizon. We also show that aggregate losses can be approximated by a mean-variance mixture of Gaussian distributions. This provides an alternative perspective on the distribution of the aggregate loss and a second method of evaluation besides crude Monte Carlo sampling. For certain insurance contracts, we improve the accuracy of the approximation-based valuation by using a correction term. This was originally developed by El Karoui and Jiao (2009) for the efficient pricing of complex financial instruments, there in the context of a classical Gaussian approximation.

Our digital twin approach enables a comprehensive analysis of risk in transportation systems: we study the impact of fleet sizes and their driving configurations on system efficiency and insurance prices. System efficiency is measured using traditional traffic statistics based on local traffic counts such as traffic flow, average speed, and density. Insurance claims are examined in terms of their probability distributions and selected statistical functionals. The innovation of our approach lies in the fact that counterfactual

traffic scenarios can be considered which allow to test insurance solutions beyond historical data. This could also include online insurance products that depend on specific traffic situations.

The main contributions of this paper are as follows:

- (i) We develop a powerful methodological framework to generate accident data based on microscopic traffic models in analogy to the concept of digital twins.
- (ii) Specifically, we construct an implementation based on the state-of-the-art open-source traffic simulator SUMO and illustrate the potential of the approach in comprehensive case studies.
- (iii) Structurally, we characterize the total loss distribution approximatively as a mean-variance mixture. This also yields alternative valuation procedures. These results hold for general microscopic traffic models; SUMO is used in case studies to illustrate the findings.
- (iv) Based on Stein's method, we obtain a correction term in the valuation, derived from the results of El Karoui and Jiao (2009), which enables surprisingly accurate pricing of insurance contracts.

1.1. Outline

The paper is organized as follows. Section 1.2 discusses related contributions in the literature. Section 2 presents the microscopic traffic model that captures also accidents. Section 3 discusses the evaluation of the losses. Case studies are presented in Section 4. Section 5 concludes and discusses further research challenges. The Supplementary Material contains in Appendix E details on the implemented sampling procedure; further simulation results, not presented in Section 4, are documented in Appendix F.

1.2. Literature

Our paper combines microscopic traffic models with probabilistic tools from actuarial science and quantitative risk management to study risks in traffic systems. The literature can be classified along two dimensions: the engineering perspective and the actuarial perspective.

The Engineering Perspective. An important field of operations research is the analysis and optimization of road traffic systems (see, e.g., Gazis, 2002) with respect to their efficiency. Traffic models are indispensable tools for this purpose: Macroscopic models are based on the functional relationships between macroscopic features such as traffic flow, traffic density, and average speed. These models allow the study of issues such as the efficient routing of vehicles under different constraints (see, e.g., Acemoglu *et al.*, 2018; Colini-Baldeschi *et al.*, 2020). Stochastics can be used to extend such risk considerations in terms of uncertain travel times (e.g., Nikolova and Stier-Moses, 2014). In the context of various applications of transportation systems, tailored stochastic models provide suitable analytical tools; the extensive literature includes, for example, the efficient routing of ambulances (Maxwell *et al.*, 2010) or the allocation of capacity in bike-sharing systems (Freund *et al.*, 2022).

To model transportation systems at a level of higher granularity, microscopic traffic models are used (see, e.g., Helbing, 2001). Their simulation, that is, the computation of trajectories from accelerations, is computationally more demanding. There are established software solutions that facilitate the application of microscopic models. In this work, we use in the context of illustrative case studies the simulation engine SUMO (see the Section 2.3 for an overview). Examples of competing microscopic traffic simulators include VISSIM (Fellendorf and Vortisch, 2010) and Aimsun (Casas *et al.*, 2010). While SUMO is open-source software, these competitors are commercial. Software packages such as SUMO are also important in the context of testing and validating new technological developments such as autonomous vehicles. This is, for example, discussed in Schwarz and Wang (2022), Szalai *et al.* (2020), and Kusari *et al.* (2022). To deploy autonomous vehicles in the real world, lengthy and expensive testing phases are required. Acceleration strategies are being developed to shorten these times (e.g., Zhao *et al.*, 2018). These approaches rely on importance sampling techniques to overcome the rare event

nature of safety-critical situations. Arief *et al.* (2018) develop simulation-based testing methodologies in order to analyze autonomous vehicles in relevant scenarios that are constructed using collected data. Norden *et al.* (2019) create a framework for the black-box assessment of the safety of autonomous vehicles. They apply their framework on a commercial autonomous vehicle system. Our work focuses on aggregate losses over relatively long time horizons. By considering 1-year losses via a conditional loss modeling approach, we bypass the problem of simulating rare events. Further literature on microscopic traffic models, their calibration, and applications for traffic safety is reviewed in the online appendix in Appendix A.

The Actuarial Perspective. The ambitious goal of achieving maximum efficiency and complete safety through engineering design cannot be realized in reality; accidents can never be completely excluded, even if residual risks can be kept very small. Insurance is an instrument to deal with the residual risk of infrequent losses, cf. McNeil *et al.* (2015) and Wüthrich (2013).

The premiums of motor insurance contracts are traditionally based on historical claims data collected by insurance companies. Insurance premiums are calculated based on individual characteristics of the driver (age, driving experience, etc.) and the vehicle (type, location, etc.). These tariffs are often complemented by bonus-malus schemes (see, e.g., Denuit *et al.*, 2007; Lemaire *et al.*, 2015; Afonso *et al.*, 2017) to incentivize more careful driving and prevent insurance fraud.

Novel pricing approaches use telematics technology (see, e.g., Husnjak *et al.*, 2015 for an overview). This involves collecting GPS data from vehicles, which can be analyzed and classified. Machine learning techniques are suitable to process these large amounts of data. We refer to Gao *et al.* (2022) for a methodological overview. Verbelen *et al.* (2018), Corradin *et al.* (2021), So *et al.* (2021), and Henckaerts and Antonio (2022) discuss telematics pricing and usage-based auto insurance products.

Our approach can be understood as complementary to telematics pricing: instead of analyzing driving data to determine the driving behavior of individuals, we model the behavior of vehicles as a *driving configuration* and subsequently generate driving data and insurance claims. Our approach is in particular suitable, if novel technologies are studied in counterfactual situations, for example, autonomous vehicles. To our knowledge, there is no other work that develops a microfounded model of traffic accidents that can be leveraged to study insurance pricing.

2. Model components

Our microfounded simulation model for investigating accident losses is based on two components:

- (i) At its core is a deterministic microscopic traffic model that realistically characterizes the motion of vehicles in a traffic system typically represented by a system of ordinary differential equations. The particular strength of the simulation-based approach is that also counterfactual situations can be modeled – capturing, for example, the consequences of technological innovation.
- (ii) This microscopic traffic model is extended to include the possibility of random accidents. At random accident times, local traffic data are observed which characterize the probability distribution of the occurring losses. In our specific implementation of this general conceptual approach, the SUMO microscopic traffic simulator is used, but our theoretical results hold for any microscopic traffic model that satisfies the properties described in the next section.

2.1. Microscopic traffic networks

We consider a road network that is typically embedded into a two-dimensional area $A \subseteq \mathbb{R}^2$. The network may consist of roads, junctions, roundabouts, intersections, highways, etc. on which vehicles move. The collection of all vehicles in the network is denoted by \mathcal{M} . Each vehicle $i \in \mathcal{M}$ is assigned an origin–destination pair $(O^i, D^i) \in A$.

We consider a fixed time horizon $T > 0$. Vehicles move over time from their origin to their destination on a (potentially changing) path. We denote by $x^i(t)$ the position of vehicle i at time $t \in [0, T]$, by $v^i(t) = \frac{d}{dt}x^i(t)$ their velocity, and by $a^i(t) = \frac{d}{dt}v^i(t)$ their acceleration. We make the implicit assumption that vehicles are located in O^i until some release time and remain in D^i once reached. Thus, we let $\mathcal{M}(t) = \{i \in \mathcal{M} : x^i(t) \notin \{O^i, D^i\}\}$ be those vehicles which are currently *inside* the network, that is, they have left their origin but not reached their destination, yet. If we only consider vehicles which belong to a certain group of vehicles (also called a *fleet*) $\Phi \subseteq \mathcal{M}$, we will write $\mathcal{M}^\Phi(t)$. This could, for example, be a fleet of vehicles with the same driving characteristics.

At the core of many microscopic traffic models are *car-following models*. Prominent examples include the Intelligent Driver Model (Treiber *et al.*, 2000), the Optimal Velocity Model (Bando *et al.*, 1994 and 1995), and the Krauß model (Krauß, 1998). Car-following models determine the acceleration behavior of an individual vehicle i along its path on the basis of information on the positions and velocities of the vehicles, typically in a neighborhood of i , and the properties of the system. Often, only the preceding vehicle on the road is relevant, and the acceleration of i is constructed such that vehicles move forward while maintaining a minimal distance. Through specific choices more complex traffic scenarios (e.g., intersections, overtaking) can still be represented in such a manner. Mathematically, car-following models correspond to systems of coupled ordinary differential equations.

Traffic State. We denote by $\gamma(t) = (x^i(t), v^i(t), a^i(t))_{i \in \mathcal{M}}$ the state of the traffic system at time t . It records the position, velocity, and acceleration of any vehicle. The evolution of the traffic system over time is depicted by the (high-dimensional) trajectory $t \mapsto \gamma(t)$.

Macroscopic traffic statistics aggregate these microscopic data. Typical examples include traffic flow (number of vehicles that pass a certain point per time unit), traffic density (number of vehicles per length unit), and average speed. These measures quantify the performance of traffic systems.

Local Traffic Conditions. In order to model the occurrence of accidents depending on local traffic conditions, we partition A in regions. More precisely, we partition A into a finite number of disjoint sets $A_r \subseteq A$ such that $A = \bigcup_{r=1}^R A_r$ and $R \in \mathbb{N}$. We call the elements A_r of the partition a *traffic module*.

We let $\mathcal{M}_r(t) = \{i \in \mathcal{M}(t) : x^i(t) \in A_r\} \subseteq \mathcal{M}(t)$ denote those vehicles that are in A_r at time t (with $\mathcal{M}^\Phi(t)$ defined in analogy). The local traffic state of the module is $\gamma_r(t) = (x^i(t), v^i(t), a^i(t))_{i \in \mathcal{M}_r(t)}$. Key local traffic characteristics (density, flow, speed, etc.) can then be expressed as functions of $\gamma_r(t)$ and its evolution over small time windows.

2.2. Microscopic traffic model with accident losses

So far, the evolution of the traffic system is a deterministic function of time. The advantage of microscopic models is that they enable a detailed simulation of traffic systems. The driving behavior of the vehicles can be varied, likewise their number and paths, road conditions, etc., in order to generate many different scenarios. Such models provide a detailed picture, similar to digital twins of reality, and can be used to analyze potential future traffic systems or to understand the impact of new technologies.

We consider a finite collection of different traffic scenarios $\gamma^k := (\gamma^k(t))_{t \in [0, T]}$ with $k \in \{1, 2, \dots, K\}$ for a short time horizon $T > 0$. The aim is to analyze characteristics of traffic over the long time horizon NT , for example, 1 year, for some large $N \in \mathbb{N}$; this is modeled by a finite sequence of traffic scenarios $(k_1, k_2, \dots, k_N) \in \{1, 2, \dots, K\}^N$. The N subintervals of length T are called time buckets. We will be interested in quantities aggregated or averaged over the whole time horizon NT . Examples include the average traffic flow, the total number of accidents, the aggregate losses due to accidents, etc. These quantities depend not only on the order of the traffic scenarios during this time period but also on their number of occurrences.

We denote by μ^k the number of occurrences of scenario k divided by N , that is, the relative frequency of this traffic scenario over the considered time horizon NT . The vector $\mu = (\mu^1, \mu^2, \dots, \mu^K)^\top$ lies in the simplex $\Delta^{K-1} = \{x \in \mathbb{R}_+^K : \sum_{k=1}^K x^k = 1\}$. We assume that μ is not deterministic, but a random variable. This is to account for the fact that the relative frequencies of traffic scenarios fluctuate over different

years due to varying weather conditions, random changes in traffic demand, or other factors. From a mathematical point of view, this construction leads to a mixture model with exogenous factor μ .

Accident Occurences. We now introduce our traffic accident model that will permit an analysis of aggregate losses and corresponding insurance contracts. The likelihood of the occurrence of an accident is modeled as a function of the traffic scenario. We consider two specifications: a binomial and a Poisson model.

- (i) *Binomial Model.* Accidents are rare events. For a given traffic scenario k , we assume that the probability p^k of an accident is close to zero. This probability may, of course, depend on the evolution of the traffic scenario, that is, on the path $t \mapsto \gamma^k(t)$, and we will discuss concrete specifications later. Given a realization of μ , accidents are assumed to be independent across time buckets. This implies that traffic scenario k occurs for $N\mu^k$ time buckets corresponding to a duration of $NT\mu^k$, and the number of accidents C^k during this period has a conditional binomial distribution with parameters p^k and $N\mu^k$:

$$C^k \mid \mu \sim \text{Bin} (p^k, N\mu^k).$$

- (ii) *Poisson Model.* An alternative model assumes that accidents occur at random times with a distribution governed by an intensity λ^k/T that depends on the traffic scenario k . More specifically, the number of accidents C^k during the period governed by scenario k of duration $NT\mu^k$ is conditionally Poisson distributed with parameter $\lambda^k N\mu^k$:

$$C^k \mid \mu \sim \text{Poiss} (\lambda^k N\mu^k).$$

Accident Losses. Loss sizes conditional on the occurrence of accidents are assumed to be independent across traffic scenarios and across time buckets. We assume that the conditional loss distribution with conditional distribution function F^k depends only on the traffic scenario k . We will discuss examples below. Random total losses over the considered time horizon NT are equal to

$$L = \sum_{k=1}^K \sum_{c=1}^{C^k} X_c^k$$

where the random variables $X_c^k, k = 1, 2, \dots, K, c \in \mathbb{N}$, are independent and $X_c^k \sim F^k, c \in \mathbb{N}$, for any k .

Concrete Specifications. Microscopic traffic models are experimental environments that allow to simulate the behavior of systems where no real data are yet available. Traffic planning can be supported by such models, and the impact of new technologies can be tested in a counterfactual analysis. Here, we specify the general principles how accident occurrences and losses can be based on microscopic traffic models. An implementation will in this paper be based on SUMO, see Section 2.3, but could also rely on any other suitable traffic model.

Initially, K traffic scenarios need to be selected as a basis for the model. While running any deterministic traffic scenario k over the time window $[0, T]$, information can be extracted about the traffic states γ_r^k in each module $r \in \{1, 2, \dots, R\}$. In SUMO typically not complete data on the whole paths are extracted, but only selected information at loop detectors in the network that are part of the implementation.

In reality, the likelihood of accidents typically increases with higher traffic density and higher velocities, ceteris paribus. Also the distribution of losses is influenced by quantities of this type. Examples are described in Section 4. This allows a computation of p^k and λ^k as a function of the data. Using the data associated with the modules, we may specify probabilities and intensities for the modules such that $p^k = \sum_{r=1}^R p_r^k$ and $\lambda^k = \sum_{r=1}^R \lambda_r^k$. In the binomial model, p_r^k/p^k is the conditional probability that the accident is in module r given that an accident occurs. In the Poisson model, the intensities $\lambda_r^k, r = 1, 2, \dots, R$, determine the accident times for each module. The resulting sequence of random times in the whole traffic system possesses the intensity λ^k . Conversely, if we first simulate random times with intensity λ^k and then randomly choose a corresponding module with probability λ_r^k/λ^k in a second step, the random times

associated with each module r possess intensity λ_r^k . Both procedures produce a number of accidents C^k that occur during the period governed by scenario k .

The distributions of losses given a single random event will be chosen as follows. For each traffic scenario, we consider a collection of distribution functions $(F^{k,\psi})_{\psi \in \Psi}$ where ψ corresponds to data that may be extracted from the traffic simulation. In order to do so, we uniformly simulate a random time in $[0, T]$ and extract at this time the data from scenario k that determine ψ . The resulting distribution F^k is a mixture of the distributions $(F^{k,\psi})_{\psi \in \Psi}$. The mixing distribution is derived on the basis of the traffic data of scenario k that are generated from our microscopic traffic model.

Insurance Contracts and Statistical Functionals. The microscopic traffic model with accidents will be the basis for the generation of aggregate losses. We study statistical functionals and insurance contracts. The analysis will be based on Monte Carlo simulations, but we also compare approximation techniques that we describe in Section 3.

The focus is on functions of aggregate losses L . Letting $h: \mathbb{R} \rightarrow \mathbb{R}$ be an increasing function, we analyze $h(L)$. In particular, we investigate the following functions corresponding to three types of insurance coverage: $h(x) = x$ (full coverage), $h(x) = \max(x - \theta, 0)$, $\theta \geq 0$ (constant deductible), $h(x) = \min(x, \theta)$, $\theta \geq 0$ (stop loss). In each case, we evaluate various statistical functionals:

- (i) *Expectation.* $\mathbb{E}(h(L))$,
- (ii) *Variance.* $\text{Var}(h(L)) = \mathbb{E}((h(L))^2) - \mathbb{E}(h(L))^2$,
- (iii) *Skewness.* $\zeta_{h(L)} = \frac{\mathbb{E}[(h(L) - \mathbb{E}(h(L)))^3]}{(\text{Var}(h(L)))^{3/2}}$,
- (iv) *Value-at-Risk.* $\text{VaR}_p(h(L)) = \inf\{x \in \mathbb{R} : P(h(L) \leq x) \geq p\}$,
- (v) *Expected Shortfall.* $\text{ES}_p(h(L)) = \frac{1}{1-p} \int_p^1 \text{VaR}_q(h(L)) \, dq$.

These functionals allow also the computation of insurance premiums on the basis of premium principles such as the expectation principle, the variance principle, or the standard deviation principle.

2.3. Traffic scenarios in SUMO

2.3.1. A brief overview

A state-of-the-art open-source software that allows us to generate traffic scenarios is SUMO, “Simulation of Urban MObility.” A reference publication on SUMO is Lopez *et al.* (2018); in addition, a detailed user documentation can be found online (see sumo.dlr.de/docs/index.html). Freely available since 2001, SUMO was originally developed by the German Aerospace Center and extended by an active research community. It allows for a plethora of modeling choices at different levels and has been successfully applied to tackle many important research questions addressing (see eclipse.org/sumo/about/), for example, traffic light optimization, routing, traffic forecasting, and autonomous driving.

In the following, we give a short overview. At its core, SUMO is a software which generates a traffic scenario $\gamma = (\gamma(t))_{t \in [0, T]}$ from a given set of input files:

- (i) *Network File.* In SUMO, a traffic network is described by a directed graph whose nodes represent intersections and edges roads. All nodes and edges have attributes, including positions, shapes, speed limits, traffic regulation, etc. As an example, the city of Wildau is represented as a SUMO network in Figure 1.
- (ii) *Route File.* Vehicles are generated on the basis of *traffic demands* between origins and destinations. Routes can either be defined for each vehicle as a *trip* or as recurring *flows* along a specific path. If only origin and destination are provided, the corresponding route is computed when the vehicle enters the system. The problem of allocating traffic demand to routes in a network is referred to as the traffic assignment problem. A standard reference is Patriksson (2015). The *vehicle type* determines the microscopic characteristics of the vehicle such as the governing car-following model, driving parameters (e.g., maximal speed, maximal acceleration,

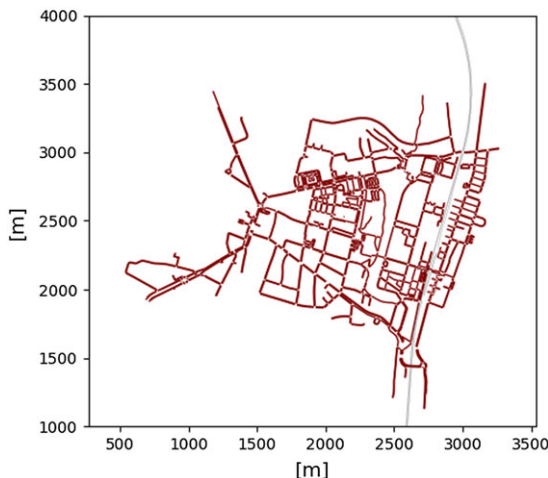


Figure 1. SUMO network of Wildau.

time headway), size, color, etc. By default, vehicles are passenger cars. Other modes, such as pedestrian, bicycle, or truck can also be selected.

- (iii) *Additional Files.* Further components are specified in additional files. An important example are induction loop detectors. These collect time series data on aggregate traffic statistics by counting the vehicles which pass a certain position during a short time interval.

The collection of input files determines the traffic evolution, also called the *SUMO scenario*. The computation can be executed either as a command line application or with a GUI that visualizes the movement of the vehicles through the network over time.

Data Extraction. One particularly appealing extension of SUMO is the “Traffic Control Interface,” TraCI (see Wegener *et al.*, 2008). TraCI provides online access to the microscopic traffic simulation and permits, at each time step, through a comprehensive list of commands (a detailed description can be found at sumo.dlr.de/docs/TraCI.html) to retrieve data and to change the states of objects such as vehicles, roads, traffic lights, etc. Available in standard programming languages (our case studies are based on the Python implementation), TraCI yields easy access to SUMO without the need to modify the underlying code. We use TraCI to extract microscopic data on positions, velocities, and accelerations of randomly selected vehicles.

2.3.2. Generation of traffic scenarios

To represent traffic in a given area over a longer time horizon (e.g., $NT = 1$ year), we generate a diverse collection of traffic scenarios $\gamma^1, \dots, \gamma^K$ of duration T in SUMO by varying the input files. Traffic over a longer time horizon is represented by a random composition of these traffic scenarios. Our general construction has already been discussed in Section 2.2.

SUMO Scenario. We describe the key steps to set up a SUMO scenario and point out specific references to the SUMO documentation in the online appendix in Section B. SUMO provides tools that facilitate the creation of input files, for example, the graphical network editor *netedit* that visualizes a SUMO scenario and allows to modify its properties. In practice, network files are typically imported from other data sources. For example, one can build a real-world traffic network in SUMO from OpenStreetMap data by selecting an area from a map. The route file specifies the trips of the vehicles and the definition of the general vehicle types with their microscopic characteristics. First, there are several options to generate trips in SUMO. These can be obtained from empirical data in the form of traffic counts, imported to SUMO as origin-destination matrices, or modeled via *ad hoc* choices, for

example, using *netedit*. Second, each vehicle is associated with a vehicle type specified on the basis of a comprehensive list of attributes. The corresponding values can be set manually in the route file or accessed and modified via *netedit*. In the absence of detailed traffic data for calibration, SUMO offers an *activity-based demand generation* which deduces traffic demand from general assumptions on the structure of the population (inhabitants, households, etc.) in the considered area. The tool *activitygen* automates the process and produces an artificial route file.

SUMO admits a large variety of modeling choices. Tailored to the needs of the modeler, a highly detailed SUMO scenario can be constructed. Our case studies will be based on publicly available SUMO scenarios; these consist of network and route files which are calibrated to real-world cities.

Varying Traffic Conditions. The input files need to be constructed in such a way that they reflect varying traffic conditions over longer time periods. This includes weather conditions, variation of traffic demand, and other factors.

Maze *et al.* (2006) review empirical studies on the impact of adverse *weather conditions* on traffic. These may induce (i) lower traffic demand, (ii) higher risk of accidents, and (iii) modified driving behavior. Based on empirical findings, Phanse *et al.* (2022) implement reduced velocities due to rainfall. In Weber *et al.* (2019), the idea of introducing into SUMO a friction parameter per road is discussed. Traffic scenarios under adverse weather conditions can be captured by suitable driving parameters in the route file (e.g., by variation of maximal speed, maximal acceleration, etc.), and this may be combined with a weather-dependent model of the occurrence and the severity of accidents.

Traffic demand is traditionally estimated from traffic counts. We refer to Bera and Rao (2011) for an overview. With increasing data availability, the estimation can be enhanced by floating car data (cf., e.g., Nigro *et al.*, 2018), that is, data generated from vehicles over time as they are driving. Traffic demand varies over time, but patterns reoccur over longer time horizons (see, e.g., Soriguera, 2012). Demand depends on the considered traffic network. Weekdays differ from days on the weekend; peaks in demand occur at common commute times. Rush hours are spatio-temporal phenomena that can be analyzed in detail (see, e.g., Xia *et al.*, 2018).

To reflect the heterogeneity of traffic scenarios, two options are available in SUMO: (i) a variety of route files is generated that is consistent with the desired modeling granularity. This process can be automatized via an additional program, a *route file generator*, that produces route files with the desired characteristics. (ii) Another option is to select a medium time horizon (e.g., $T_{\text{SUMO}} = 24$ h) with a corresponding route file that depicts varying traffic demand over time. From the generated SUMO scenario, a selection of small time horizon scenarios (e.g., $T = 1$ min) can be efficiently generated by utilizing SUMO's option to save the state of the running simulation at a priori specified times and load these later.

Besides weather and traffic demand, many other factors influence the traffic dynamics. Wagner (2016) discusses the representation of autonomous vehicles in SUMO. Lücken *et al.* (2019) utilize SUMO to study control transition, that is, selected safety critical situations where the human driver needs to take over control from an autonomously driving vehicle. Pagany (2020) study the impact of wildlife on traffic, an issue that is also relevant in the context of traffic accidents.

3. Evaluation methods

The accident losses L can be simulated using Monte Carlo methods. The simulations may be used to estimate the value of statistical functionals and to price insurance products. We will briefly describe the Monte Carlo methods. In addition, on the basis of the binomial model, we construct a Gaussian approximation to $\mathbb{E}(h(L))$ where the function h corresponds to the three types of insurance coverage that we consider: full coverage, constant deductible, and stop loss. This allows a numerical evaluation similar to Frey *et al.* (2008) and El Karoui and Jiao (2009). The latter paper provides a correction term derived by Stein's method that we will exploit in our application.

3.1. Monte Carlo methods

The Monte Carlo simulation of L requires sampling the number of accidents C^k in either the binomial or Poisson model and sampling the independent conditional losses X_c^k , $c \in \mathbb{N}$, for each traffic scenario $k = 1, 2, \dots, K$ from the corresponding distribution F^k . These tasks can be performed separately but require both a prior evaluation of the microscopic traffic model.

- (i) *Prior Evaluation of Traffic Model.* For each traffic scenario k , a single run delivers data that are the basis for a computation of, respectively, the accidents probabilities p^k and intensities λ^k as well as the corresponding values p_r^k and λ_r^k on the level of the modules $r = 1, 2, \dots, R$.
- (ii) *Number of Accidents.* Sampling from μ and using the results of the prior evaluation allow to sample the number of accidents C^k for each traffic scenario k in both the binomial and Poisson model.
- (iii) *Conditional Accident Losses.* Based on the precomputed values of the accident probabilities and the accident intensities, respectively, we simulate for each traffic scenario k the random locations and times of accidents. These data can be stored. For these locations and times, traffic data ψ are extracted from an additional single run of traffic scenario k . Given ψ , the losses are generated according to conditional loss distributions $F^{k,\psi}$ as described in Section 2.2. Details of the implementation are explained in Section 4.

These Monte Carlo methods can be flexibly applied to all considered functionals. In the special cases of expectation and variance, Wald’s equation can simplify the computation, since the losses L are given in the form of a collective model.

3.2. Gaussian approximation

Another way to compute $\mathbb{E}(h(L))$ for the considered types of insurance coverage is a Gaussian approximation, possibly improved by a correction term. A Gaussian approximation can easily be motivated within the binomial model. For N sufficiently large, the distribution of L given μ is approximately normal, implying that L is a mean-variance mixture of Gaussian distributions. This is an important structural insight from this approximation.

In this section, we condition on μ , that is, suppose that μ is fixed and given. The general results for μ random are then a corollary by considering suitable mixtures according to the distribution of μ . The total random losses in the binomial model can be rewritten as:

$$L = \sum_{k=1}^K \sum_{c=1}^{N\mu^k} \mathbf{1}_c^k \cdot X_c^k$$

where the random variables $\mathbf{1}_c^k, X_c^k, k = 1, 2, \dots, K, c \in \mathbb{N}$, are independent, $X_c^k \sim F^k$, and $\mathbf{1}_c^k$ are Bernoulli random variables taking the value 1 with probability p^k and the value 0 otherwise, $c \in \mathbb{N}$, for any k . Setting

$$Y_c^k := \mathbf{1}_c^k \cdot X_c^k, \quad m^k := \mathbb{E}(Y_c^k), \quad (\sigma^k)^2 := \mathbb{E}\left([Y_c^k - m^k]^2\right), \quad (\zeta^k)^3 := \mathbb{E}\left([Y_c^k - m^k]^3\right),$$

$k = 1, 2, \dots, K, c \in \mathbb{N}$, a classical normal approximation of L is $\sum_{k=1}^K N\mu^k m^k + Z$ with

$$Z \sim \mathcal{N}\left(0, \sum_{k=1}^K N\mu^k (\sigma^k)^2\right).$$

Remark 3.1. The estimation of $m^k, (\sigma^k)^2$, and $(\zeta^k)^3$ requires the simulation of the random variables $X_c^k, c \in \mathbb{N}$. The independent terms $\mathbf{1}_c^k, c \in \mathbb{N}$, factor out, are idempotent and have known expectation p^k .

We focus on three types of insurance coverage, $h(x) = x$ (full coverage), $h(x) = \max(x - \theta, 0)$, $\theta \geq 0$ (constant deductible), $h = \min(x, \theta)$, $\theta \geq 0$ (stop loss), and obtain an approximation $\mathbb{E}(h(Z))$ of $\mathbb{E}(h(L))$ in each of these cases.

On the basis of Stein’s method (see Chen *et al.*, 2011 and Ross, 2011 for an overview), El Karoui and Jiao (2009) suggest correction terms in order to improve the approximation, that is, the approximation $\mathbb{E}\left(h\left(\sum_{k=1}^K N\mu^k m^k + Z\right)\right)$ is replaced by the corrected approximation:

$$\mathbb{E}\left(h\left(\sum_{k=1}^K N\mu^k m^k + Z\right)\right) + C_h.$$

The correction C_h depends on the degree of smoothness of the derivatives of the function h and thus differs (see Theorem 3.1 and Proposition 3.6 in El Karoui and Jiao, 2009) for the three types of coverage. We define

$$d_1 = \sum_{k=1}^K N\mu^k m^k, \quad d_2 = \sum_{k=1}^K N\mu^k (\sigma^k)^2, \quad d_3 = \sum_{k=1}^K N\mu^k (\zeta^k)^3, \quad \tilde{h}(x) = h(x + d_1).$$

We obtain the following correction terms:

- (i) *Full Coverage.* In the case of full coverage, the correction term of El Karoui and Jiao (2009) disappears. In general, if h is some Lipschitz function with bounded third derivative, the correction term equals $C_h = \frac{d_3}{2d_2^2} \cdot \mathbb{E}\left(\left\{\frac{Z^2}{3d_2} - 1\right\} Z\tilde{h}(Z)\right)$.
- (ii) *Constant Deductible.* $C_h = \frac{(\theta - d_1)d_3}{6d_2} \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot d_2}} \cdot \exp\left\{-\frac{(\theta - d_1)^2}{2 \cdot d_2}\right\}$
- (iii) *Stop Loss.* A stop loss $x \mapsto \min(x, \theta)$ can be written as the difference between full coverage $x \mapsto x$ and a constant deductible $x \mapsto \max(x - \theta, 0)$. This implies that the correction term for a constant deductible appears with a negative sign in this case.

The advantage of the (corrected) Gaussian approximation in comparison with pure Monte Carlo is that, once the numbers m^k , $(\sigma^k)^2$, and $(\zeta^k)^3$ have been computed for each traffic scenario $k = 1, 2, \dots, K$, no further data need to be stored or sampled in order to compute $\mathbb{E}(h(L))$. The approximate representation of the distribution of L as a mean-variance mixture is a considerable simplification.

4. Application

We illustrate the application of our microscopic traffic model with accidents on the basis of a publicly available SUMO scenario of a real city.

4.1. SUMO scenario and accident data

Wildau is a small German city of approximately 10,000 inhabitants, located around 30 km south-east of the capital Berlin. A SUMO model of the city was developed within a study project by the Technical University of Applied Sciences Wildau and is publicly available (see github.com/DLR-TS/sumo-scenarios/tree/main/Wildau).

SUMO Scenario. The implemented road network is visualized in Figure 1. It is specified using 646 nodes connected by 1426 edges. The city itself is crossed by the railway; the tracks are represented by the gray line. Vehicles in the present scenario are calibrated from real traffic counts provided by local authorities, see also Behrisch and Hartwig (2022). Such numbers of vehicles per time unit for certain positions are then assigned to different routes through the city. The original scenario has a duration of 7010 s. Empty in the beginning, vehicles enter the system with a peak of approximately 240 vehicles that drive simultaneously. In total, 2502 vehicles are generated.

In the following section, we describe in detail how we adjust this SUMO scenario to obtain suitable ingredients for our case studies. This yields a collection of traffic scenarios that allow us to compare the effects of different driving characteristics and fleet sizes on the total loss and related insurance premiums.

Varying Traffic Conditions. For different collections of model parameters representing different traffic systems we generate adjusted SUMO scenarios. For each choice of parameters, we proceed as follows to produce $K = 100$ traffic scenarios of length $T = 60$ s. Traffic scenarios $k = 1, 2, \dots, 50$ correspond to selected time intervals from the SUMO scenario. Traffic scenarios $k = 51, 52, \dots, 100$ represent higher traffic volumes. They are generated by replacing the original route file by a route file that consists of two copies of the original route file. This simple procedure generates a larger amount of vehicles along the original paths. The traffic scenarios are again selected time intervals from the corresponding SUMO scenario.

To represent the full year, we set $N = 365 \cdot 24 \cdot 60 = 525,600$. We need to specify the random vector $\mu = (\mu^1, \dots, \mu^K)^\top$ describing the number of occurrences of the individual traffic scenarios divided by N . For the purpose of illustration, we specifically assume that two probability measures ν_g, ν_b are given on $\{1, 2, \dots, K\}$ which approximately correspond to the relative frequencies of traffic scenarios in two prototypical years $y = g, b$. In addition, we suppose that the type y of the current year is random where both values g and b have probability $1/2$. Given y , we generate $\mu = (\mu^1, \dots, \mu^K)^\top$ from a multinomial distribution corresponding to ν_y . That is, for all time buckets $n = 1, 2, \dots, N$, a traffic scenario k is chosen independently from the distribution ν_y on $\{1, 2, \dots, K\}$. Dividing the number of occurrences of a scenario k by N , one obtains its random relative frequency μ^k for any $k = 1, 2, \dots, K$. In our case study, the distribution ν_g corresponds to lower traffic densities on average, while ν_b is associated with higher traffic densities, that is, we set

$$\nu_g := \begin{cases} 1/75, & k = 1, 2, \dots, 50, \\ 1/150, & k = 51, 52, \dots, 100 \end{cases}, \quad \nu_b := \begin{cases} 1/150, & k = 1, 2, \dots, 50, \\ 1/75, & k = 51, 52, \dots, 100. \end{cases}$$

Accident Data. The German Accident Atlas provided by Statistical Offices of the Federation and the Länder (2022) depicts the locations of all police-reported accidents *involving personal damage* that occurred within 1 year. In 2020, within the modeled area of Wildau (approximately) 48 accidents were registered. There are also aggregate statistics for Germany for all police-reported accidents. In 2020, approximately 11.8% of all road accidents involved personal damage (see Statistisches Bundesamt, 2023). We use an estimate of $\bar{c}_{\text{year}} = 48/11.8\% \approx 407$ accidents for calibration purposes.

4.2. Model specification

Our goal is to analyze accident losses for a fleet Φ over the time horizon of 1 year.

Fleet Definition. In the Wildau scenario, vehicles are defined using repeated flows from origins to destinations. Passenger cars (next to trucks and the train) are defined via 90 different flows. A flow generates vehicles of a given type at a given position. From this position, they navigate to a specified destination.

In SUMO, a passenger car represents a certain vehicle type. Initially, all passenger cars belong to the same vehicle type and, consequently, have the same driving characteristics. To introduce a fleet Φ of vehicles whose driving characteristics we can vary, we define a new vehicle type Φ and construct corresponding new SUMO scenarios. Fixing a fraction $\rho^\Phi \in [0, 1]$ of vehicles belonging to Φ , we retain approximately $1 - \rho^\Phi$ of the existing flow definitions and modify ρ^Φ of the flow definitions suitably in order to model the fleet. In our case studies, we consider $\rho^\Phi = 10\%, 50\%, 90\%$.

Driving Configuration. Vehicles in a fleet Φ are of the same type. Various characteristics can be varied in SUMO; we focus on *maximal speed* v_{\max} , *maximal acceleration* $a_{\max} > 0$, and *time headway* $\zeta > 0$. The time headway is the distance which is kept to the preceding vehicle measured in time, that is, a velocity-weighted safety distance. We refer to a fixed selection of driving characteristics as a *driving configuration*.

In our case studies, we will vary the driving configuration for all vehicles in the fleet Φ and keep all other vehicles as originally introduced; we use the implementation of an Intelligent Driver Model without any speed deviation that does not include further random effects. A driving configuration of vehicles in fleet Φ is denoted by $\xi = (v_{\max}, a_{\max}, \zeta)$. Specifically, we consider

$$\begin{aligned} \xi^{1a} &= (5 \text{ m/s}, 0.8 \text{ m/s}^2, 3.0 \text{ s}), \quad \xi^{2a} = (10 \text{ m/s}, 0.8 \text{ m/s}^2, 2.0 \text{ s}), \quad \xi^{3a} = (15 \text{ m/s}, 0.8 \text{ m/s}^2, 1.0 \text{ s}), \\ \xi^{1b} &= (5 \text{ m/s}, 2.6 \text{ m/s}^2, 3.0 \text{ s}), \quad \xi^{2b} = (10 \text{ m/s}, 2.6 \text{ m/s}^2, 2.0 \text{ s}), \quad \xi^{3b} = (15 \text{ m/s}, 2.6 \text{ m/s}^2, 1.0 \text{ s}). \end{aligned}$$

The configurations 1,2,3 increase in terms of “aggressiveness” from driving slowly with a large headway to fast with a small headway – with two options a and b for the maximal acceleration.

Remark 4.1. The specific choices are inspired by the following considerations: The implemented road speed limit in Wildau is 50 km/h which is approximately 13.9 m/s. Vehicles in the original Wildau scenario have a maximal acceleration of 0.8 m/s², while SUMO’s default value is 2.6 m/s². Similarly, SUMO’s default time headway is 1.0 s.

Accident Occurrence. The best estimate for the total number of accidents in Wildau is $\bar{c}_{\text{year}} \approx 407$. From this, we derive a uniform and a nonuniform accident occurrence model. In both cases, we specify accident probabilities $p^{1,k}$ and $p^{2,k}$ for the binomial model as well as accident intensities $\lambda^{1,k}$ and $\lambda^{2,k}$ for the Poisson model.

- (i) *Uniform Accident Occurrence.* Assuming that accidents occur uniformly over the year, we obtain a probability per time bucket of an accident in the system of $p^1 = \bar{c}_{\text{year}}/N \approx 7.7 \times 10^{-4}$. This is the accident probability that we allocate to each traffic scenario k . We also suppose that accidents occur uniformly across all vehicles in the system. This implies that the probability that any accident occurs in scenario k within the fleet Φ is

$$p^{\Phi,1,k} = \rho^\Phi \cdot p^1, \quad k \in \{1, \dots, K\}.$$

This probability is used in the binomial model. For the Poisson model, we set $\lambda^{\Phi,1,k} = p^{\Phi,1,k}$, $k = 1, 2, \dots, K$, since the intensity approximately equals the probability of an accident per time bucket.

In the case of uniform accident occurrence, we do not consider any spatial variations of the likelihood of accidents due to different traffic conditions. This means that we do not distinguish any modules, that is, we set $R = 1$.

- (ii) *Non-Uniform Accident Occurrence.* In reality, the likelihood of accidents depends on external factors such as weather and local traffic conditions, for example, the velocity of vehicles and traffic density. The quantities vary spatially and over time.

From SUMO runs, we obtain for each traffic scenario $k = 1, 2, \dots, K$ and each module $r = 1, 2, \dots, R$ pairs (d_r^k, \bar{v}_r^k) on the average density and velocity. These statistics can be computed in SUMO, for example, from data that are obtained at induction loop detectors which are placed within the modules; as a proxy for density, we extract the *occupancy* of the loop detector, that is, the fraction of time which it is occupied by a vehicle.

For $r = 1, \dots, R$, we choose benchmark values d_r^* and \bar{v}_r^* for the density and velocity and specify occurrence probabilities and intensities that vary spatially and over time:

$$\lambda_r^{\Phi,2,k} = p_r^{\Phi,2,k} := \frac{p^{\Phi,1,k}}{R} \cdot \frac{\bar{v}_r^k}{\bar{v}_r^*} \cdot \frac{d_r^k}{d_r^*} \cdot e^{-(\zeta^{\Phi-1})}, \quad k \in \{1, \dots, K\}, \quad r \in \{1, \dots, R\}.$$

The last term refers to deviations of the time headway from SUMO’s default value of 1.0 s: a larger time headway is associated with less risky driving. We set $p^{\Phi,2,k} = \sum_{r=1}^R p_r^{\Phi,2,k}$ and $\lambda^{\Phi,2,k} = \sum_{r=1}^R \lambda_r^{\Phi,2,k}$.



Figure 2. Partition of Wildau and placement of induction loop detectors.

In our case studies, we will consider a grid of $R = 4$ modules and compute d_r^k and \bar{v}_r^k as averages over measurements from 10 induction loop detectors that are placed in each module (see also Figure 2). We use the scenario averages $d_r^* = \frac{1}{K} \sum_{k=1}^K d_r^k$ and $\bar{v}_r^* = \frac{1}{K} \sum_{k=1}^K \bar{v}_r^k$. If $\sum_{k=1}^K \mathbb{E}(\mu^k) \bar{v}_r^k = \bar{v}_r^*$, $\sum_{k=1}^K \mathbb{E}(\mu^k) d_r^k = d_r^*$, and $\zeta^\Phi = 1$, then we essentially recover on average the case of uniform accident occurrence.

Accident Losses. The distributions F^k of accident losses associated with a traffic scenario $k = 1, 2, \dots, K$ are constructed on the basis of traffic data that are extracted from the SUMO runs. The general procedure was described in Section 2.2; here, we explain the specific implementation that we use in our case studies.

The likelihood of accident occurrence was discussed in the previous section. F^k is the conditional distribution of an accident loss in traffic scenario k if an accident occurs. Time in traffic scenario k is enumerated by $t \in [0, T]$, and we assume that the time τ of the accident conditional on its occurrence is uniformly distributed on $[0, T]$, that is, $\tau \sim \text{Unif}[0, T]$. We choose a module \mathcal{R} at random in which the accident occurs and assume, respectively, that

$$P(\mathcal{R} = r) = \frac{p_r^{\Phi, \cdot, k}}{p^{\Phi, \cdot, k}}, \quad P(\mathcal{R} = r) = \frac{\lambda_r^{\Phi, \cdot, k}}{\lambda^{\Phi, \cdot, k}}, \quad r \in \{1, \dots, R\}.$$

These ratios depend in the case of nonuniform accident occurrence on the specific fleet, since the properties of the fleet alter the route file that is used to generate the SUMO scenarios; this is true, although the multiplicative terms ρ^Φ appear in both the numerator and denominator and cancel out.

In the chosen module, we pick one or more vehicles at random and extract from the traffic scenario data for these vehicles. In our concrete implementation, we simply choose at time τ a single vehicle I uniformly at random in module \mathcal{R} , that is, its conditional distribution is

$$I \mid \tau, \mathcal{R} \sim \text{Unif}(\mathcal{M}_{\mathcal{R}}^\Phi(\tau)).$$

For the purpose of illustrating our approach, the only data we extract are the velocities v^I of the randomly chosen vehicles that are involved in accidents. We set $\psi = v^I$ and assume that the conditional loss distribution $F^{k, \psi}$ is known (we assume that $F^{k, 0}$ corresponds to a Dirac measure in 0; if $\mathcal{M}_{\mathcal{R}}^\Phi(\tau) = \emptyset$, we set

$\psi = 0$, resulting in 0 losses). If we denote the distribution of ψ by \mathcal{L}^k , we obtain the distribution F^k as a mixture:

$$F^k = \int F^{k,\psi} d\mathcal{L}^k.$$

In our case studies, we will assume that $F^{k,\psi} = F^\psi$ for all k ; however, the mixing distribution \mathcal{L}^k will depend on the traffic scenario k . We consider the following examples for F^ψ :

- (i) *Gamma Distribution.* We define distributions with varying levels of dispersion. A measure for the dispersion of a random variable X is the coefficient of variation defined by $c_v = \sqrt{\text{Var}(X)}/\mathbb{E}(X)$. For $c_v \in \{1/2, 1, 2\}$, we choose $F^\psi = \Gamma\left(\frac{1}{c_v^2}, \frac{1}{c_v^2\psi^2}\right)$. The expectation of this distribution is ψ^2 and increases quadratically with ψ , the velocity of the vehicle involved in an accident; this is consistent with the fact that losses scale with kinetic energy. The variance of the distribution F^ψ equals $c_v^2\psi^4$; hence, the coefficient of variation is indeed c_v .
- (ii) *Log-Normal Distribution.* We consider log-normal distributions with expectation ψ^2 and variance $c_v^2\psi^4$, implying that the coefficient of variation is again $c_v \in \{1/2, 1, 2\}$. This log-normal distribution is obtained as the distribution of $\exp(Z)$ for a normal random variable Z with expectation $\ln(\psi^2/\sqrt{1+c_v^2})$ and variance $\ln(1+c_v^2)$, that is,

$$F^\psi = \mathcal{LN}\left(\ln\left(\frac{\psi^2}{\sqrt{1+c_v^2}}\right), \ln(1+c_v^2)\right).$$

Remark 4.2. The conditional frequency and severity given traffic scenarios is an essential ingredient to our modeling approach. Its flexibility comes from the fact that traffic can be micro-simulated, and its impact on aggregate losses can be analyzed if conditional loss distributions are available. These need to be studied in more detail on the basis of empirical data and structural considerations. This paper focuses on an illustration of the simulation methodology. Reviews on different methodologies for the assessment of accident frequencies and severities based on underlying covariates are provided by Lord and Mannering (2010), Savolainen *et al.* (2011), Mannering and Bhat (2014), and Theofilatos and Yannis (2014). Lian *et al.* (2020) reviews the use of big data for the analysis of traffic conditions and their relationships to accident frequencies and severities. Retallack and Ostendorf (2019) review articles on the relationship of traffic congestion and accidents. Malin *et al.* (2019) investigate the relative accident risk of different road and weather conditions and combinations of conditions using data for major roads in Finland; their analysis is based on the notion of Palm probability. Using generalized additive models, Becker *et al.* (2022) quantify the combined effects of traffic volume and meteorological parameters on probabilities of 78 different crash types. Comi *et al.* (2022) investigate the suitability of various data mining techniques in analyzing the factors underlying accidents and predicting these in case studies based on data collected in Rome.

4.3. Case studies

4.3.1. Overview

We illustrate our modeling approach in case studies on multiple levels. A selection of case studies is discussed in detail in Sections 4.3.2 & 4.3.3. All numerical results for the following choices are documented online in tables in Appendix F:

- (i) *Fleet Models.* We analyze six driving configurations with three different fleet proportions.
- (ii) *Accident Occurrence.* In our traffic system, accidents occur uniformly or nonuniformly in space. Their number is given by a binomial or a Poisson model with parameters depending on traffic condition.

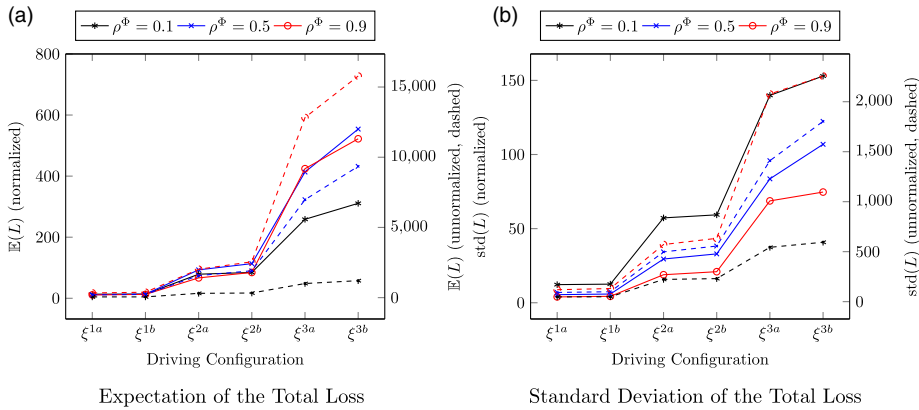


Figure 3. Impact of fleet size and driving configuration on the total loss L (solid lines represent normalized values (left y-axis) and dashed lines unnormalized ones (right y-axis)).

- (iii) *Accident Losses.* We study two parametric families of loss distributions with three different choices for the coefficient of variation.
- (iv) *Insurance Design.* Insurance losses are a function of the total losses; we distinguish three contract designs.

Denoting aggregate losses by L , we evaluate for each type of insurance coverage h the resulting insurance losses $h(L)$ in terms of their expectation, variance, and skewness, and the monetary risk measures Value-at-Risk and Average Value-at-Risk, also called Expected Shortfall. To analyze the distributions in detail, we provide qq-plots and estimates of cumulative distribution functions (CDFs) and densities. The main tool to access the random variable $h(L)$ is Monte Carlo sampling; we provide a pseudocode how we obtain samples of L in Algorithm 1 in Appendix E. In Section 4.3.3, we compare this approach to the normal mean-variance mixture approximation introduced in Section 3.2.

This paper explores the analysis and management of risks that occur in vehicle fleets in traffic systems. We distinguish to perspectives:

- (i) *The Engineering Perspective.* In these case studies, we fix the accident occurrence and accident loss distributions and vary the fleet models, that is, driving configurations and fleet proportions. We focus on $\mathbb{E}(L)$, $\text{Var}(L)$, and complement these with analyses of the performance of traffic system.
- (ii) *The Actuarial Perspective.* In these case studies, we fix the fleet model and vary accident occurrence and accident loss distributions as well as the insurance design. We study the distribution of L and the insurance prices $\mathbb{E}(h(L))$.

4.3.2. *The engineering perspective*

Our micro-modeling approach allows us to study the effects of different traffic-related controls on total losses L ; we investigate the effects of fleet size and traffic configuration. Throughout this section, we consider nonuniform accident occurrence in the binomial model with Gamma-distributed accident losses and a coefficient of variation $c_v = 1$.

Losses. We evaluate expected loss $\mathbb{E}(L)$ and standard deviation $\text{std}(L)$ for different fleet models. To compare losses for different fleet sizes, we normalize losses per 100 expected insured vehicles: for each traffic scenario k , the number of insured vehicles is the number of vehicles belonging to Φ as given in the underlying route file. We refer to Appendix F for more details. The model specification was explained in Section 4.2 which includes in particular a description of the driving configurations. The results are documented in Figure 3. The solid lines are the normalized quantities.

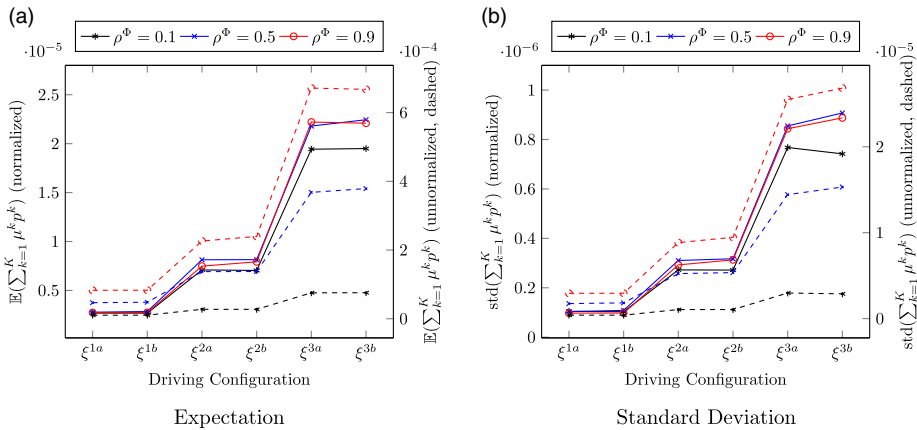


Figure 4. Impact of fleet size and driving configuration on expectation and standard deviation of average accident frequency (solid lines represent normalized values (left y-axis) and dashed lines unnormalized ones (right y-axis)).

In Figure 3(a), we see that increasing the aggressiveness of driving increases both the total and normalized expected loss. An impact of the maximal acceleration on losses is only substantial for the most aggressive driving configurations ξ^3 . Increasing the fleet size increases the expected loss which is primarily due to the fact that we count losses only within the fleet and a higher volume is associated with higher losses. More interesting is the normalized case: apparently higher speeds also increase the normalized losses.

In Figure 3(b), we have the corresponding standard deviations. Increasing the aggressiveness of driving increases the standard deviation of the total loss. The standard deviations of the normalized losses are decreasing in the fleet size. The main reason is that fluctuations normalized for a fixed volume are larger for smaller pools than for larger pools; a rationale for this is provided by the law of large numbers and the central limit theorem.

The frequency and severity of accidents are, of course, increasing in the aggressiveness of driving. To demonstrate this, we evaluate the expectation and the standard deviation of the average accident frequency $\sum_{k=1}^K \mu^k p^k$ (both normalized and unnormalized) and the average accident severity $\sum_{k=1}^K \mu^k X_1^k$, as displayed in Figures 4 and 5. A larger fleet increases frequency and, in aggressive scenarios, also the expected accident severity. This is, of course, due to the specific choice of the driving behavior of the considered fleets in comparison with the driving behavior of the remaining vehicles and does not necessarily hold for all traffic systems in general. In this section, we discussed losses for selected special cases. A comprehensive set of tables for all other cases and statistical functionals is provided online in Appendix F.

Changing the characteristics of the fleet not only affects the losses. At the same time, this has an impact on the performance of the traffic system. We present further results in Appendix C.

4.3.3. The actuarial perspective

From an actuarial perspective, it is relevant to understand the *risk* that corresponds to the insurance losses. This requires a more detailed analysis of the *probability distributions*. To do this, we pick a particular fleet model and use probabilistic techniques to evaluate the distribution of $h(L)$. From now on, we consider $\rho^\phi = 0.5$ with driving configuration ξ^{2a} and nonuniform accident occurrence.

Distributional Analysis of Losses. We start our investigations with the total losses L . Table 1 shows the evaluation of statistical functionals for different accident losses in the case of the binomial model. These numbers quantify the risk entailed in the total losses. Both the distributional family and the chosen coefficient of variation for the accident loss model have a substantial effect on the risk.

Table 1. Statistical Functionals of L for $\rho^\Phi = 0.5$ and ζ^{2a} .

Binomial model						
	Gamma			Log-normal		
	$c_v = 0.5$	$c_v = 1.0$	$c_v = 2.0$	$c_v = 0.5$	$c_v = 1.0$	$c_v = 2.0$
$\mathbb{E}(L)$	1577.8	1571.5	1578.4	1581.7	1576.2	1582.7
$\text{Var}(L)$	160,179.5	247,943.9	626,190.6	162,067.5	247,069.8	614,713.4
ζ_L	0.333	0.561	0.993	0.377	0.660	2.152
$\text{VaR}_{0.9}(L)$	2111.9	2219.1	2634.2	2112.9	2242.4	2526.6
$\text{ES}_{0.9}(L)$	2331.7	2538.8	3233.3	2342.4	2557.5	3280.3
$\text{VaR}_{0.95}(L)$	2275.9	2471.3	3070.5	2288.4	2468.1	2996.2
$\text{ES}_{0.95}(L)$	2468.9	2748.0	3628.9	2491.4	2772.5	3838.4
$\text{VaR}_{0.99}(L)$	2587.3	2891.3	3999.6	2608.8	2943.8	4228.6
$\text{ES}_{0.99}(L)$	2755.9	3188.6	4490.2	2809.0	3234.4	5424.2

The statistical functionals of the total loss are approximated using 10,000 independent samples of L .

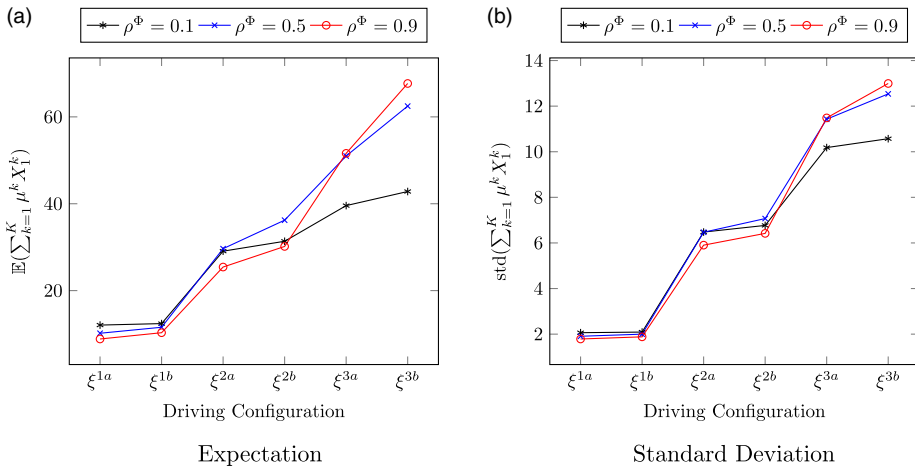


Figure 5. Impact of fleet size and driving configuration on expectation and standard deviation of average accident severity.

A visual impression of the distributions is provided in Figure 6. We compare different accident loss models while fixing the coefficient of variation $c_v = 2$. We plot the empirical distribution functions as estimates of the CDF and a kernel density estimate of the corresponding densities. Moreover, Figure 6(c) shows qq-plots for the quantiles of standardized values of the losses against quantiles of a standard normal distribution (samples are standardized by subtracting their sample average and dividing by their sample standard deviation).

We find that the binomial and the Poisson model do not differ too much. Yet, log-normal accident losses produce heavier tails than the corresponding Gamma losses. The qq-plots reveal that the right tails are heavier compared to a normal distribution, while the left tails are lighter. The latter observation simply relates to the fact that the original losses are nonnegative, while the normal distribution takes values on the whole real line.

In Figure 7, we analyze the impact of the coefficient of variation while fixing the log-normal distribution for the accident losses. We see again that binomial and Poisson model do not differ substantially.

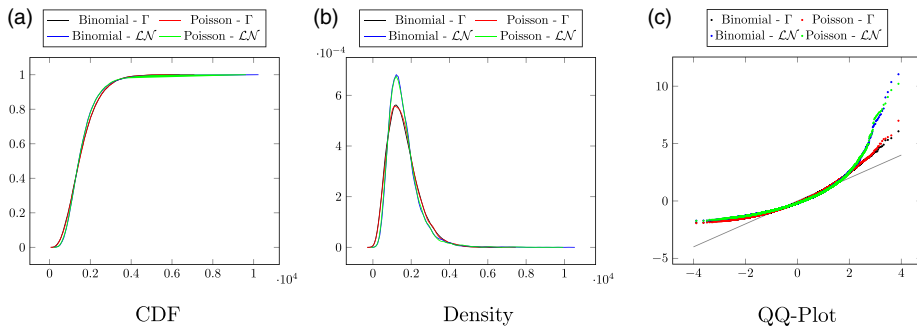


Figure 6. Distribution of the total loss for fixed coefficient of variation $c_v = 2$.

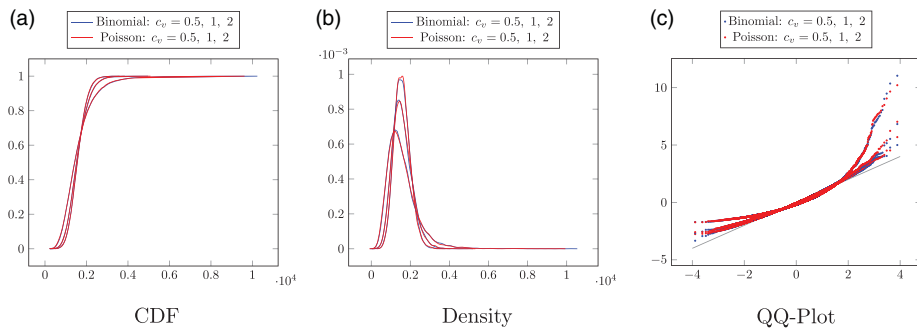


Figure 7. Distribution of the total loss for log-normal accident losses and varying coefficient of variation.

However, the effect of the coefficient of variation is clearly visible: increasing c_v produces heavier right tails. Introducing dispersion to accident losses substantially changes the distribution of the total losses.

Comparison of Losses and Normal Mean-Variance Mixture Approximation. In Section 3.2, we suggested a normal mean-variance mixture approximation for the total loss. To study the quality of this approximation, we generate 10,000 samples from the approximation; in the following, we focus on the case of Gamma-distributed accident losses with coefficient of variation $c_v = 1$.

To sample from the approximation, we rely on the following computations of m^k and $(\sigma^k)^2$. Using $\mathbb{E}(\mathbf{1}_c^k) = p^k$ and $\mathbb{E}(X_c^k) = \mathbb{E}(\mathbb{E}(X_c^k | \psi)) = \mathbb{E}(\psi^2) = \int \psi^2 d\mathcal{L}^k$, we obtain for the Gamma losses $m^k = p^k \cdot \int \psi^2 d\mathcal{L}^k$ and $(\sigma^k)^2 = p^k \cdot c_v^4 \cdot \int \psi^4 d\mathcal{L}^k \cdot \left(1 + \frac{1}{c_v^2}\right) \frac{1}{c_v^2} - (m^k)^2$.

Remark 4.3. For the notation, we refer to Sections 3.2 & 4.2. The computations are valid for any coefficient of variation c_v , but we use only $c_v = 1$ in the numerical case study.

The involved moments of ψ are approximated using 10,000 samples from the traffic simulation, for each $k = 1, \dots, K$. A sample from the normal mean-variance mixture approximation is generated by, first, sampling μ and, second (conditional on μ), sampling the normal random variable $\sum_{k=1}^K N\mu^k m^k + Z$ with $Z \sim \mathcal{N}\left(0, \sum_{k=1}^K N\mu^k (\sigma^k)^2\right)$.

Figure 8 shows the qq-plot comparing quantiles of the crude Monte Carlo simulation with quantiles of the approximation. This demonstrates the quality of our suggested approximation. It is almost exact between the 5% and 95% quantile as the values lie on the half-line. It is still very good for the 1–5% and 95–99% quantiles and is only less accurate in the extreme tails where also in the Monte Carlo simulation only few data points are available. These analyses, on the one hand, confirm the postulated structural model insight. On the other hand, they also validate the implementation of our crude Monte Carlo sampling.

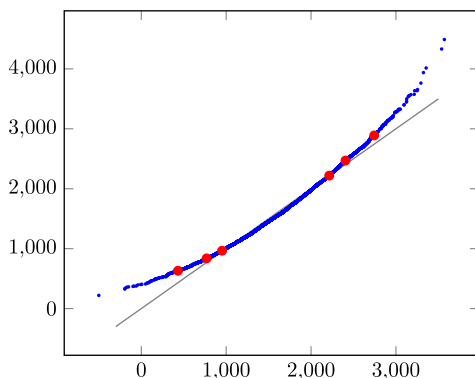


Figure 8. QQ-Plot of 10,000 Monte Carlo samples (y-axis) versus 10,000 samples of mixture approximation (x-axis) (the red dots mark the 1%, 5%, 10%, 90%, 95%, 99% quantiles).

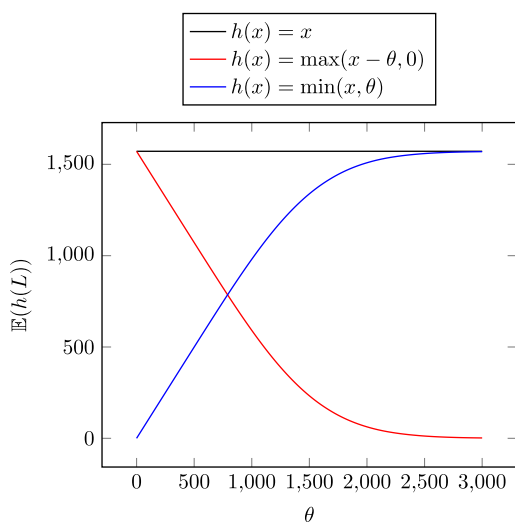


Figure 9. Insurance prices.

Pricing and Evaluation Methods. To conclude our case studies, we study prices for various insurance contracts. We compare $\mathbb{E}(L)$ (full coverage), $\mathbb{E}(\max(L - \theta, 0))$ (constant deductible), and $\mathbb{E}(\min(L, \theta))$ (stop loss) for different values of θ . The results are given in Figure 9. We obtain the typical hockey stick profiles satisfying the parity $\mathbb{E}(L) = \mathbb{E}(\max(L - \theta, 0)) + \mathbb{E}(\min(L, \theta))$. We note that other insurance contracts can easily be represented in our framework; also deductibles per accident can be implemented by changing the accident loss distributions accordingly.

Besides Monte Carlo methods, our normal mean-variance mixture approximation, and the correction suggested in Section 3.2 provide alternative techniques to compute the prices $\mathbb{E}(h(L))$ for the different types of coverage h . To compute the correction term C_h , we need to compute $(\zeta^k)^3$ (for the notation we refer to Sections 3.2 & 4.2). For Gamma-distributed losses, we obtain

$$(\zeta^k)^3 = p^k \cdot c_v^6 \cdot \int \psi^6 d\mathcal{L}^k \cdot \left(2 + \frac{1}{c_v^2}\right) \left(1 + \frac{1}{c_v^2}\right) \frac{1}{c_v^2} - 3m^k(\sigma^k)^2 - (m^k)^3.$$

Since $\mathbb{E}(h(L)) = \mathbb{E}(\mathbb{E}(h(L) | \mu))$, we may generate samples of μ and evaluate the corresponding conditional expectations $\mathbb{E}(h(L) | \mu)$; in the normal mean-variance mixture approximation, these are

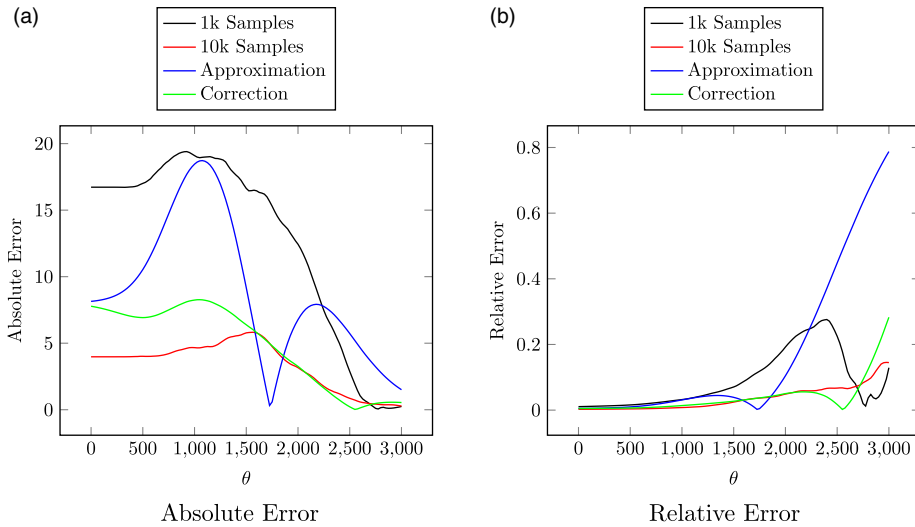


Figure 10. Comparison of estimation errors.

expectations of functions of normally distributed random variables. We compute these expectations numerically as integrals with respect to Lebesgue measure using a normal density.

We compare the estimation errors of the different approaches in Figure 10. We produce 100,000 samples of L to approximate the “true” value of $\mathbb{E}(\max(L - \theta, 0))$ of coverage with constant deductible, for different values of θ . Independently, we generate 10,000 samples and consider Monte Carlo approximations based on all 10,000 samples and based only on the first 1000 samples. We also study the normal mean-variance mixture approximation with and without correction using 1000 samples of μ .

While the absolute error generally decreases in the deductible θ for all methods (apart from some local effects), the relative error increases for larger values of θ which are associated with a lower price of the contract. At the same time, we observe that, in terms of the relative error, the normal mean-variance mixture approximation produces reasonable estimation results (compared to 1000 samples) for moderate values of θ . This is in line with our previous observations on the quality of our normal mean-variance mixture approximation which becomes worse in the extreme tail of L . Interestingly, the estimation error can largely be reduced using the correction; with the correction, the estimation becomes quite good even for large values of θ .

5. Conclusion

This paper developed a methodology to study accident losses based on microscopic traffic simulators. An adaption of the digital twin paradigm enabled us to test the impact of fleet sizes and their driving configuration on system efficiency, accident losses, and insurance premiums. It was shown that – on a 1-year horizon – total losses can be approximated by a mean-variance mixture of normal distributions. This offered an alternative technique to evaluate the model; the numerical efficiency can be increased adding a correction term that is derived by Stein’s method. The proposed methodology can be extended and modified, for example, and utilized to study future traffic systems. We illustrated in counterfactual case studies that were based on the software SUMO how accident risk can be successfully analyzed, both from an engineering and an actuarial perspective.

Future research should also address the important issues of calibration and validation. While real data can be used to calibrate models that describe historical and current transportation systems, simulation models that generate artificial data are essential to evaluate new technologies in future transportation

systems. An important question is to what extent and how historical data can be methodically used to calibrate and validate such simulation models. For example, real microscopic traffic data, for example, on accident patterns, collected by means of telematics technologies, could be applied to optimize models in the future. The comparison of simulated and real data will also allow to investigate the impact of traffic conditions on accident occurrence, financial losses, and the probability distributions thereof.

Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/asb.2023.36>

References

- Acemoglu, D., Makhdoui, A., Malekian, A. and Ozdaglar, A. (2018) Informational Braess' paradox: The effect of information on traffic congestion. *Operations Research*, **66** (4), 893–917.
- Afonso, L.B., Cardoso, R.M.R., dos Reis, A.D.E. and Guerreiro, G.R. (2017) Measuring the impact of a bonus-malus system. *ASTIN Bulletin*, **47** (2), 417–435.
- Arief, M., Glynn, P. and Zhao, D. (2018) An accelerated approach to safely and efficiently test pre-production autonomous vehicles on public streets. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE.
- Bando, M., Hasebe, K., Nakayama, A., Shibata, A. and Sugiyama, Y. (1994) Structure stability of congestion in traffic dynamics. *Japan Journal of Industrial and Applied Mathematics*, **11** (2), 203–223.
- Bando, M., Hasebe, K., Nakayama, A., Shibata, A. and Sugiyama, Y. (1995) Dynamical model of traffic congestion and numerical simulation. *Physical Review E*, **51** (2), 1035–1042.
- Becker, N., Rust, H.W. and Ulbrich, U. (2022) Weather impacts on various types of road crashes: a quantitative analysis using generalized additive models. *European Transport Research Review*, **14** (1), 37.
- Behrisch, M. and Hartwig, P. (2022) A comparison of SUMO's count based and countless demand generation tools. *SUMO Conference Proceedings*, Vol. 2, pp. 125–131.
- Bera, S. and Rao, K.V.K. (2011) Estimation of origin-destination matrix from traffic counts: The state of the art. *European Transport/Trasporti Europei*, **49**, 2–23.
- Casas, J., Ferrer, J.L., Garcia, D., Perarnau, J. and Torday, A. (2010) Traffic simulation with aimsun. *International Series in Operations Research & Management Science*, pp. 173–232. Springer New York.
- Chen, L.H., Goldstein, L. and Shao, Q.-M. (2011) *Normal Approximation by Stein's Method*, Springer Berlin Heidelberg.
- Colini-Baldeschi, R., Cominetti, R., Mertikopoulos, P. and Scarsini, M. (2020) When is selfish routing bad? The price of anarchy in light and heavy traffic. *Operations Research*, **68** (2), 411–424.
- Comi, A., Polimeni, A. and Balsamo, C. (2022) Road accident analysis with data mining approach: evidence from rome. *Transportation Research Procedia*, **62**, 798–805.
- Corradin, A., Denuit, M., Detyniecki, M., Grari, V., Sammarco, M. and Trufin, J. (2021) Joint modelling of claim frequencies and behavioral signals in motor insurance. *ASTIN Bulletin*, **52** (1), 33–54.
- Denuit, M., Maréchal, X., Pitrebois, S. and Walhin, J.F. (2007) *Actuarial Modelling of Claim Counts: Risk classification, Credibility and Bonus-Malus Systems*. Chichester: John Wiley & Sons.
- El Karoui, N. and Jiao, Y. (2009) Stein's method and zero bias transformation for cdo tranche pricing. *Finance & Stochastics*, **13**, 151–180.
- Fellendorf, M. and Vortisch, P. (2010) Microscopic traffic flow simulator VISSIM. *International Series in Operations Research & Management Science*, pp. 63–93. Springer New York.
- Freund, D., Henderson, S.G. and Shmoys, D.B. (2022) Minimizing multimodular functions and allocating capacity in bike-sharing systems. *Operations Research*, **70** (5), 2715–2731.
- Frey, R., Popp, M. and Weber, S. (2008) An approximation for credit portfolio losses. *The Journal of Credit Risk*, **4** (1), 3–20.
- Gao, G., Meng, S. and Wüthrich, M.V. (2022) What can we learn from telematics car driving data: A survey. *Insurance: Mathematics and Economics*, **104**, 185–199.
- Gazi, D.C. (2002) The origins of traffic theory. *Operations Research*, **50** (1), 69–77.
- Helbing, D. (2001) Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, **73** (4), 1067–1141.
- Henckaerts, R. and Antonio, K. (2022) The added value of dynamically updating motor insurance prices with telematics collected driving behavior data. *Insurance: Mathematics and Economics*, **105**, 79–95.
- Husnjak, S., Perakovic, D., Forenbacher, I. and Mumdziev, M. (2015) Telematics system in usage based motor insurance. *Procedia Engineering*, **100**, 816–825.
- Jones, D.C.S., Nassehi, A., Yon, J. and Hicks, B. (2020) Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, **29**, 36–52.
- Krauß, S. (1998) *Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics*. Ph.D. Thesis, DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

- Kusari, A., Li, P., Yang, H., Punshi, N., Rasulis, M., Bogard, S. and LeBlanc, D.J. (2022) Enhancing SUMO simulator for simulation based testing and validation of autonomous vehicles. *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE.
- Lemaire, J., Park, S.C. and Wang, K.C. (2015) The use of annual mileage as a rating variable. *ASTIN Bulletin*, **46** (1), 39–69.
- Lian, Y., Zhang, G., Lee, J. and Huang, H. (2020) Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles. *Accident Analysis & Prevention*, **146**, 105711.
- Lopez, P.A., Wiessner, E., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flotterod, Y.-P., Hilbrich, R., Lucken, L., Rummel, J. and Wagner, P. (2018) Microscopic traffic simulation using SUMO. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE.
- Lord, D. and Mannering, F. (2010) The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, **44** (5), 291–30.
- Lücken, L., Mintsis, E., Porfyri, K.N., Alms, R., Flötteröd, Y.-P. and Koutras, D. (2019) From automated to manual - modeling control transitions with sumo. *SUMO User Conference 2019*.
- Malin, F., Norros, I. and Innaamaa, S. (2019) Accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention*, **122**, 181–188.
- Mannering, F.L. and Bhat, C.R. (2014) Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, **1**, 1–22.
- Maxwell, M.S., Restrepo, M., Henderson, S.G. and Topaloglu, H. (2010) Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, **22** (2), 266–281.
- Maze, T.H., Agarwal, M. and Burchett, G. (2006) Whether weather matters to traffic demand, traffic safety, and traffic operations and flow. *Transportation Research Record: Journal of the Transportation Research Board*, **1948** (1), 170–176.
- McNeil, A.J., Frey, R. and Embrechts, P. (2015) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, revised edition.
- NHTSA (2022) Early estimates of motor vehicle traffic fatalities and fatality rate by sub-categories in 2021. *US Department of Transportation*, pp. 1–10.
- Nigro, M., Cipriani, E. and del Giudice, A. (2018) Exploiting floating car data for time-dependent origin–destination matrices estimation. *Journal of Intelligent Transportation Systems*, **22** (2), 159–174.
- Nikolova, E. and Stier-Moses, N.E. (2014) A mean-risk model for the traffic assignment problem with stochastic travel times. *Operations Research*, **62** (2), 366–382.
- Norden, J., O’Kelly, M. and Sinha, A. (2019) Efficient black-box assessment of autonomous vehicle safety. [arXiv: 1912.03618](https://arxiv.org/abs/1912.03618) [cs.LG].
- Pagany, R. (2020) Wildlife-vehicle collisions - influencing factors, data collection and research methods. *Biological Conservation*, **251**, 108758.
- Patriksson, M. (2015) *The Traffic Assignment Problem: Models and Methods*. Mineola, New York: Dover Publ Inc.
- Phanse, S., Chaturvedi, M. and Srivastava, S. (2022) Modelling and simulation of road traffic under rainy conditions. *2022 14th International Conference on Communication Systems & NETWORKS (COMSNETS)*. IEEE.
- Retallack, A.E. and Ostendorf, B. (2019) Current understanding of the effects of congestion on traffic accidents. *International Journal of Environmental Research and Public Health*, **16** (18), 3400.
- Ross, N. (2011) Fundamentals of stein’s method. *Probability Surveys*, **8**, 210–293.
- Savolainen, P.T., Mannering, F.L., Lord, D. and Qaddus, M.A. (2011) The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis & Prevention*, **43** (5), 1666–1676.
- Schwarz, C. and Wang, Z. (2022) The role of digital twins in connected and automated vehicles. *IEEE Intelligent Transportation Systems Magazine*, **14** (6), 41–51.
- So, B., Boucher, J.-P. and Valdez, E.A. (2021) Cost-sensitive multi-class adaboost for understanding driving behavior based on telematics. *ASTIN Bulletin*, **51** (3), 719–751.
- Soriguera, F. (2012) Deriving traffic flow patterns from historical data. *Journal of Transportation Engineering*, **138** (12), 1430–1441.
- Statistical Offices of the Federation and the Länder (2022) *German accident atlas*. Data licence Germany – attribution – Version 2.0.
- Statistisches Bundesamt (2023) *Accidents registered by the police by type of damage/location*.
- Szalai, M., Varga, B., Tettamanti, T. and Tihanyi, V. (2020) Mixed reality test environment for autonomous cars using unity 3d and SUMO. *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMII)*. IEEE.
- Theofilatos, A. and Yannis, G. (2014) A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, **72**, 244–256.
- Treiber, M., Hennecke, A. and Helbing, D. (2000) Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, **62** (2), 1805–1824.
- Verbelen, R., Antonio, K. and Claeskens, G. (2018) Unravelling the predictive power of telematics data in car insurance pricing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **67** (5), 1275–1304.
- Wagner, P. (2016) Traffic control and traffic management in a transportation system with autonomous vehicles. In *Autonomous Driving*, pp. 301–316. Springer Berlin Heidelberg.
- Weber, T., Driesch, P. and Schramm, D. (2019) Introducing road surface conditions into a microscopic traffic simulation. *SUMO User Conference 2019*.
- Wegener, A., Piórkowski, M., Raya, M., Hellbrück, H., Fischer, S. and Hubaux, J.-P. (2008) TraCI. *Proceedings of the 11th Communications and Networking Simulation Symposium*. ACM Press.

- Wüthrich, M.V. (2013) Non-life insurance: Mathematics & statistics. *SSRN Electronic Journal*. <https://ssrn.com/abstract=2319328>
- Xia, F., Wang, J., Kong, X., Wang, Z., Li, J. and Liu, C. (2018) Exploring human mobility patterns in urban scenarios: A trajectory data perspective. *IEEE Communications Magazine*, **56** (3), 142–149.
- Zhao, D., Huang, X., Peng, H., Lam, H. and LeBlanc, D.J. (2018) Accelerated evaluation of automated vehicles in car-following maneuvers. *IEEE Transactions on Intelligent Transportation Systems*, **19** (3), 733–744.