Uncertainty Quantification in Vehicle Velocity Simulations

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)
genehmigte

Dissertation

von

M.Sc. Paul Jakob Strähle

Tag der mündlichen Prüfung:	20. Oktober 2025
Hauptreferent:	Prof. Dr.-Ing. Carsten Proppe
Korreferent:	Prof. Dr.-Ing. Martin Cichon

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als Doktorand bei Bosch Research in Renningen in Kooperation mit dem Institut für Technische Mechanik (Institutsteil Dynamik/Mechatronik) des Karlsruher Instituts für Technologie (KIT).

Mein besonderer Dank gilt meinem Hauptreferenten, Herrn Prof. Dr.-Ing. Carsten Proppe, für die Betreuung dieser Arbeit. Seine fachliche Expertise, die anregenden Diskussionen und die stets gewährte akademische Freiheit haben die Arbeit entscheidend geprägt. Seine Offenheit ermöglichte es mir, die akademischen Anforderungen, die Bedürfnisse von Bosch und meine persönlichen Interessen in Einklang zu bringen.

Herrn Prof. Dr.-Ing. Martin Cichon danke ich herzlich für die Übernahme des Korreferats und sein Interesse an meiner Arbeit. Seine kritischen Fragen und Anmerkungen haben zur Qualität dieser Dissertation maßgeblich beigetragen.

Insbesondere möchte ich mich bei Steve Wolff-Vorbeck, Thomas Leyendecker und Kai Sandmann bedanken, die mich während der Arbeit bei Bosch als Betreuer unterstützt haben. Ihre fachliche Expertise, die Bereitschaft zu tiefgehenden Diskussionen sowie ihre intensive Unterstützung haben entscheidend zum Gelingen dieser Arbeit beigetragen.

Besonders bedanken möchte ich mich auch bei meinen Kollegen bei Bosch sowie am Institut für Technische Mechanik und allen, die mich aus dem Bosch Doktorandennetzwerk begleitet haben. Ihre wertvolle Unterstützung, die vielen Anmerkungen und guten Diskussionen sowie der soziale Austausch haben mir viel Freude bereitet und die Arbeit sehr bereichert.

Zu guter Letzt möchte ich meiner Familie danken, die mich auf meinem bisherigen Lebensweg, in guten wie in schlechten Zeiten, sowie während der Promotion stets unterstützt hat. Vielen Dank dafür, ihr bedeutet mir sehr viel.

Stuttgart, November 2025
Paul Strähle

Kurzfassung

Um zuverlässige Automobilkomponenten ressourceneffizient zu gestalten, ist es unerlässlich, die während des Betriebs zu erwartenden Belastungen zu kennen. In den frühen Entwicklungsphasen ist eine virtuelle Lastgenerierung notwendig, um diese Belastungen effizient und realistisch abzuleiten. Da viele dieser Belastungen von den gefahrenen Geschwindigkeiten des Fahrzeugs abhängen, ist die Bestimmung charakteristischer Geschwindigkeitsprofile für verschiedene Fahrzeugtypen und Fahrweisen wesentlich für die zuverlässigkeitsbasierte Konstruktion in der Automobilindustrie.

Diese Arbeit stellt neue Methoden zur Simulation von Geschwindigkeitsprofilen vor und wendet Methoden der Uncertainty Quantification (UQ) an, um Unsicherheiten in den simulierten Geschwindigkeiten in Bezug auf die Komponentenlasten zu berücksichtigen. Geschwindigkeitsmessungen von verschiedenen Fahrzeugen und Fahrern auf mehreren Routen werden analysiert und die statistischen Eigenschaften von Geschwindigkeitsfluktuationen untersucht. Es wird gezeigt, dass diese Fluktuationen durch einen stochastischen Prozess beschrieben werden können, der an die gemessenen Eigenschaften angepasst ist. Zusätzlich wird Floating Car Data (FCD) analysiert, um den Einfluss des Verkehrs auf Autobahnen zu modellieren, wobei zeitliche und örtliche Abhängigkeiten des Verkehrs berücksichtigt werden, um Staus realistisch zu beschreiben.

Die Verbesserungen in der Geschwindigkeitssimulation führen zu einem stochastischen Modellierungsansatz, der in der Lage ist, charakteristische Geschwindigkeitsprofile für verschiedene Straßentypen, Fahrer und Fahrzeuge zu simulieren. Die abgeleiteten Geschwindigkeiten und Lasten unterliegen verschiedenen Unsicherheiten. Um dies zu adressieren, wird ein Bayessches UQ Framework vorgestellt, das Eingabeparameter als Zufallsvariablen modelliert und so den Umgang mit Unsicherheiten ermöglicht. Generalized Lambda Models (GLaMs) werden angewendet, um stochastische Surrogatmodelle für die Geschwindigkeitssimulation zu erstellen, die eine globale Sensitivitätsanalyse und Bayessche Parameterinferenz ermöglichen, um gültige Beschreibungen von Geschwindigkeits- und Lastunsicherheiten zu erhalten.

Abstract

To design reliable automotive components in a resource-efficient manner, it is essential to understand the loads expected during operation. During the early stages of development, virtual load generation is necessary to derive these loads efficiently and realistically. Since many of these loads depend on the driven velocities of the vehicle, determining characteristic velocity profiles for different vehicle types and driving behaviors is essential for reliability-based design in the automotive industry.

This work presents new methods for simulating velocity profiles and applies Uncertainty Quantification (UQ) methods to account for uncertainties in simulated velocities with respect to the component loads. We analyze velocity measurements from various vehicles and drivers on multiple routes, investigating the statistical properties of velocity fluctuations. We demonstrate that these fluctuations can be described by a stochastic process tailored to the measured properties. Additionally, we analyze Floating car data (FCD) to model traffic influences on motorways, considering time and location dependencies of traffic to describe congestion realistically.

The improvements in velocity simulation lead to a comprehensive stochastic modeling approach, capable of simulating characteristic velocity profiles for diverse road types, drivers, and vehicles. The derived velocities and loads are subject to various uncertainties. To address this, we propose a Bayesian UQ framework that models input parameters as random variables, enabling the handling of uncertainties. We apply Generalized Lambda Models (GLaMs) to create stochastic surrogate models for the velocity simulation, enabling global sensitivity analysis and Bayesian parameter inference to provide valid descriptions of velocity and load uncertainties.

Contents

Appendix

1 Introduction

This work is concerned with the quantification of uncertainties in the context of vehicle velocity simulations as part of a framework for virtual load generation. The framework aims at generating loads, especially for automotive components, which allow for a reliable design without requiring load measurements. In this context, the simulation of vehicle speeds in driving direction, referred to in the following as the velocity of the vehicle, is a central component of the virtual load generation framework, as it allows for the generation of driver and vehicle-specific velocity profiles for arbitrary routes, which are central to most component loads. In the following Section 1.1, the state-of-the-art methods for reliability design in the automotive industry are introduced, based on which Section 1.2 gives an introduction to the current virtual load generation methods in literature. Finally, in Section 1.3, open research needs regarding Uncertainty Quantification (UQ), velocity simulation, and virtual load generation are addressed, while the contributions and outline of this thesis are provided.

1.1 Reliability design in the automotive industry

The design of reliable vehicle components requires information about the strength and loading of the components. Both parts are subject to significant variability due to differences in material properties and production processes, as well as variations in vehicle usage. This variability of strength and loading prevents absolute certainty on the reliability of components and leads to probabilistic statements expressed, e.g., through the failure probability over the intended service life of the product.

Unreliable products are a potential safety hazard, can lead to costly recalls, and may damage the brand reputation. Consequently, considerable effort is spent in the design process to ensure the reliability of components. To do so, the first step is to set a reliability target, which can be prescribed by technical standards, customers' requirements, or

past experiences, and from which the following steps in the design process are derived [6].

Early in the development cycle, simulation methods are used to assess the reliability and make statements on the failure probability of a component or system. These assessments are typically based on the stress-strength interference model, i.e., on a comparison of the competent strength $d_{strength}$ with the stresses d_{stress} acting on it, where failure occurs when the stress exceeds the strength, see [68] for details. The measure of stress and strength can vary depending on the application and is not necessarily a mechanical stress tensor or a related quantity, but a general measure for the component loading, e.g., a damage sum in the case of variable amplitude loading. Both strength and stress are described by a probability distribution to account for their randomness. By comparing the two distributions, the failure probability can be calculated from the probability of the stress distribution exceeding the strength distribution, see Figure 1.1.

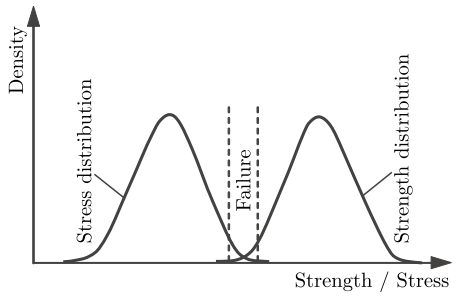


Figure 1.1: Qualitative representation of the failure probability using the stress-strength interference model. Figure modified from [6].

Component strength information is obtained through fatigue testing, yielding S-N curves that describe the relationship between stress amplitude S and cycles to failure N. Fatigue testing ideally is done at the component level, but this requires prototypes of the actual component, which in an early development phase are often not available. Instead, fatigue testing can be done with specimens and design elements, which saves cost and time but requires additional assumptions to transfer the specimen S-N curve to the component S-N curve [46, 79]. Environmental factors such as media, temperature, and corrosion can influence the S-N curve and introduce additional uncertainty when predicting it [39].

The stress on the component is usually variable and depends on the usage of the vehicle, i.e., on the routes driven with the vehicle and the behavior of the driver. The local stress, which can be compared to the component strength, follows from the loads on the component, such as forces and moments, which themselves are a function of the global loads such as the vehicle accelerations, external temperatures, and so on, see Figure 1.2

[41]. In some cases, it is possible to measure the local stress time series directly at the component using strain gauges, but usually this is complex and costly, as it requires access to the relevant highly stressed areas of the component. Therefore, measurements of the loads on the component, such as temperatures, forces, and moments, are taken, which is usually easier, and from which the local stresses can be derived using simulation. Load measurements on the component level do not necessarily require prototypes of the investigated component if older generation components show similar behavior, which allows the use of past measurements conducted with them.

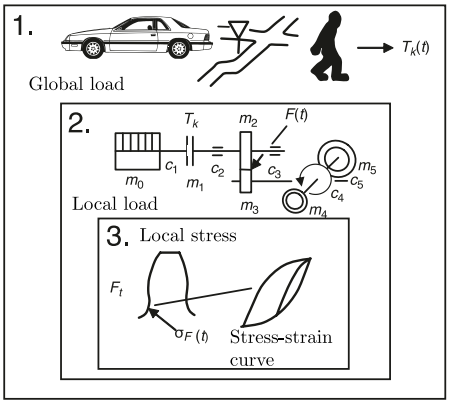


Figure 1.2: Derivation of local stresses from global loads. Figure modified from [41].

Load measurements are conducted by performing test drives on a set of defined routes with a test driver while measuring the loads on the component. To obtain the loads for the complete vehicle lifetime, the measured loads are extrapolated to the desired lifetime target [55]. The measurements can only encompass a limited range of possible usage scenarios. Consequently, certain scenarios, such as drivers predominantly using their cars in the city or primarily on motorways, which may result in a reduced lifetime, are typically not considered. The derivation of loads based on test drives, therefore, relies on the assumption that the combination of the test driver and the considered route set leads to a load that is more critical than all, or at least almost all, real usage scenarios. Moreover, a single load measurement is not reproducible, as the driver never behaves the same, and external influences like the surrounding cars and traffic lights change each time a measurement is conducted on the same route with the same driver. To capture these random influences and derive the full load distribution for the considered route and test driver combination, repeated measurements are needed, which is often not feasible due to the time and cost involved.

To compare the measured or simulated stress time series with the S-N curve describing the component strength, the series is classified with a counting method, e.g., Rainflow

counting, to derive stress cycles which are combined in a load spectrum [3, 103]. Using a fatigue life method and damage accumulation approach, the expected lifetime of the component is calculated by comparing the S-N curve with the load spectrum. Damage accumulation is typically done with Miner's rule [70] or its modifications [40, 60, 108]. By incorporating the scatter of the S-N curve and the randomness of the load spectrum resulting from the random driving behavior, the failure probability of the component is derived under the assumptions made by the engineer, e.g., [55]. Additional uncertainties in the estimated failure probability result from damage accumulation with the empirical Miner's rule, which doesn't consider all influences on the component lifetime and can lead to significant prediction errors.

1.2 Virtual load generation

Virtual load generation aims to generate loads for reliability design virtually, which aims to improve some of the weaknesses in current reliability design methods as described in Section 1.1. The current methods are largely based on measurements of component loads, which, in the automotive context, are complex and expensive. These measurements are often conducted using older product generations or require prototypes that are not available during the early stages of development. Another issue that is addressed with virtual load generation is the small sample size of load measurements, i.e., only a fraction of usage scenarios can be covered by the measurements. By generating the loads virtually, many more scenarios can be investigated in order to derive a representative field load distribution.

As the basis for virtual load generation, multiple usage scenarios representative of the real vehicle usage are defined [7]. Each scenario defines the routes driven by each driver over the lifetime of the vehicle or a representative sample of routes, including the times when the trips take place and the driving behavior of each driver. In [15], a method to combine different data sources to derive the usage scenarios is presented. Field data, which typically includes velocities, accelerations, and position data, can be used to describe the driving behavior and maneuvers. If available in large quantities, it can also be used to define sets of routes representative of all routes driven by a specific vehicle over its lifetime. Geographic data, such as map data including signs, speed limits, and topography, as well as historic weather data, are coupled with simulation models and the usage scenarios to generate driver, vehicle, and route-dependent loads on a large scale [93].

Based on the usage scenarios, component loads are derived virtually through simulation models. Many of these loads depend on the driven velocity, necessitating velocity

simulation to generate vehicle and driver-dependent velocity profiles for the given route sets. By employing load-specific simulation models, referred to as system simulations, local component loads are derived from global loads, such as velocity profiles, see Figure 1.2. There exist different approaches for the modeling of vehicle velocities, which can be classified into microscopic, macroscopic, and stochastic process-based simulation models. Microscopic approaches, such as the intelligent driver model [99], simulate the behavior and interaction of multiple individual vehicles through a system of ordinary differential equations that describes the dynamics of the vehicles' positions and velocities. While these models allow for a detailed description of velocities in a local and bounded area, they are unsuitable for generating velocity profiles on longer routes, as this would require considering all vehicles and their interactions within a much larger domain. In macroscopic traffic models, traffic is treated as a continuous flow, similar to a fluid stream [59, 80]. Instead of focusing on individual vehicles, these models consider the aggregate properties of the traffic stream, such as density, flow, and mean speed, which must adhere to the hydrodynamic relation that flow equals density times speed and the continuity equation, ensuring the conservation of vehicle flow [100]. As the macroscopic models do not model the interactions of individual vehicles with other road users, which strongly influence the short-term velocity fluctuations, as well as the differences in driving behavior of individual drivers and vehicles, they are not suitable for the majority of load generation use-cases. The stochastic velocity simulation, as described in Section 3.1, is based on the description of the velocity of a single vehicle as a stochastic process, which reduces the computational effort but, through the stochastic component, still considers the random influences from the interactions with other road users. Based on information on the speed limit, topography, and road curvature coupled with a model for the behavior of the driver, a stochastic process to model the random influences, and a vehicle dynamics model, velocity profiles for arbitrary routes can be generated within a short runtime, making stochastic velocity simulation suitable for virtual load generation [15, 85, 93]. Existing stochastic velocity simulation models, see e.g. [85], exhibit deficiencies, including insufficient consideration of correlations in the velocity profiles. Since the majority of existing and new use-cases for virtual load generation rely on simulated velocity profiles, enhancing the velocity simulation is crucial for the further development of virtual load generation methods.

The system simulation models must balance the requirement for a detailed vehicle description with short runtimes to enable the simulation of various usage scenarios. For instance, in the case of vibration loads, simplified longitudinal dynamics models are more suitable than full Multibody simulation (MBS) models due to their significantly shorter runtimes [15]. Depending on the load being considered, other models are used instead of or in addition to physical models. This includes data-based approaches that, based on past measurements, model the relationship between component loads

and relevant inputs, such as the vehicle velocity [15]. By coupling these models, i.e., the system simulations, with the velocity simulation, which generates vehicle and driver-dependent velocity profiles on the given route sets, loads can be derived fully virtually without requiring any additional measurements [27, 85].

Using the described methods, the full field load distribution considering diverse usage scenarios can be derived, from which statistically well-qualified quantiles, representing critical usage scenarios, follow, which can be used for reliability-based design as described in Section 1.1. The knowledge of the full load distribution allows for further analysis if necessary, such as the derivation of test cycles for test benches, which are statistically proven to be particularly critical. Besides the advantage of covering more usage scenarios compared to load measurements, leading to less uncertainty regarding the representativity of the derived loads, virtual loads are also available very early in development, only requiring simulation models and no prototypes. Therefore, virtual loads allow for considering reliability targets earlier, reducing the risks for late changes in development and thereby decreasing development time and cost [6]. While virtual loads have less uncertainty regarding their coverage of different usage scenarios, they are still subject to many uncertainties along the simulation chain, a major source being the velocity simulation. To increase the applicability and trust in virtual loads, these uncertainties must be quantified. For the stochastic velocity simulation, this is challenging as the simulation itself is random and the corresponding velocity measurements are subject to random influences which cannot be controlled, requiring methods of UQ that take these random factors into account.

1.3 Contributions and outline

The current state of research in reliability design and virtual load generation in the automotive industry, as well as the deficiencies of current methods addressed in Sections 1.1 and 1.2, motivate the following open research needs:

- Application of UQ methods to the velocity simulation model within a framework for virtual load generation, aiming to quantify the uncertainties of the velocity simulation with regard to the simulated loads.
- Investigation of the additional uncertainty introduced by the randomness of the velocity simulation, which, as a stochastic simulator, requires UQ methods suitable for handling such stochastic behavior.
- Advancing the velocity simulation model to reduce the error and uncertainty regarding the loads by improving the modeling of velocity fluctuations.

- Including traffic influences and stop-and-go traffic on motorways in the velocity simulation by leveraging large amounts of Floating car data (FCD) to investigate the effects of traffic on loads.

In Chapter 2, the theoretical background for the methods applied and developed in this thesis is set. This includes the fundamentals of vehicle dynamics, which are necessary for generating realistic velocity profiles that respect physical bounds. Further, Autoregressive–Moving-Average (ARMA) models, which are used for velocity fluctuation modeling, are introduced, including methods for their estimation and selection. Finally, UQ methods, including Monte Carlo sampling, surrogate modeling, sensitivity analysis, and Bayesian inference, needed to quantify the uncertainties in the velocity simulation are introduced.

Chapter 3 introduces the stochastic velocity simulation model in Section 3.1 and describes the methods developed for its improvement. The main contributions are:

- A novel method for the modeling of velocity fluctuations based on ARMA and red-noise models in combination with a truncation approach for the stochastic process and a model to describe the dependencies of the parameters. The parameters are modeled as a function of the mean velocity and are specific to each driver. Thereby, allowing for the description of the different fluctuation characteristics, such as variance, correlations, and bounds, for city, country, and motorway driving, as well as for a wide range of drivers and vehicles. It is demonstrated that the new method better replicates measured fluctuations compared to the state-of-the-art approach in [85, 93]. See Section 3.2.
- A novel method for the generation of traffic influences on motorways from time and position-dependent traffic speed distributions coupled with congestion length distributions. The distributions of the traffic speed and congestion length are derived from FCD measured in Germany. The method is demonstrated to significantly improve the match of simulated to measured velocities on motorways due to the added inclusion of traffic influences. See Section 3.3.
- Introduction of a new approach to model stop events in the velocity simulation, including an estimation method to derive realistic parameters for the model from measured velocity profiles. By extending the stop model and coupling it with the traffic model, stop events are added on motorways, leading to realistic stop-and-go traffic. See Section 3.4.

Chapter 4 describes the developed framework for UQ of virtual loads and demonstrates it with two applications. The main contributions are:

- Development of a framework which applies UQ methods to virtual load generation to quantify the uncertainties of the velocity simulation. The UQ methods are adapted to the specific requirements of virtual load generation, including the randomness of the loads, their multidimensionality, and the high number of input parameters. See Section 4.1.
- Demonstration of a novel application of Generalized Lambda Model (GLaM) surrogate models in an industrial context. The models serve as surrogates for the loads derived with the velocity simulation, which is a stochastic simulator and thus necessitates the use of stochastic surrogate models such as the applied GLaMs.
- A novel combination of GLaM stochastic surrogates with methods of Bayesian inference for the identification of posterior parameter distributions of stochastic simulators. The method is demonstrated with an application to the velocity simulation and the associated GLaM surrogates to identify the posterior distribution of the velocity simulation parameters.
- Demonstration of the developed UQ framework in two applications. In the first application, the uncertainties of the velocity simulation are identified with respect to a specific system simulation and the associated loads. The second application considers a more generic case where the uncertainties are identified with respect to properties of the velocity profiles without considering a specific system simulation. See Sections 4.2 and 4.3.

In Chapter 5, the work is summarized and the main conclusions are given, leading to an outlook for further research.

2 Fundamentals

This chapter provides the necessary concepts required for the subsequent chapters of this thesis. It begins with the fundamentals of vehicle dynamics in Section 2.1, followed by an overview of ARMA models for time series analysis in Section 2.2, which together form the basis of the stochastic velocity simulation discussed in Chapter 3. Additionally, Section 2.3 introduces methods of UQ, which are the foundation of the UQ framework for virtual load generation presented in Chapter 4.

2.1 Vehicle dynamics

Vehicle dynamics describes the vehicle's behavior in response to driver input, environmental conditions, mass distribution, engine power, and other vehicle properties. Complex vehicle dynamics models can accurately predict longitudinal and lateral behavior for various driving scenarios. [86] lists different types of vehicle dynamics models with varying complexity. These range from linear single-track models to multi-body systems with multiple subsystems such as vehicle body, wheel suspension, tires, and drivetrain. While more complex models allow for a better understanding of the physical details, their computational cost also increases considerably. Simpler models, such as the linear single-track model, are fast to evaluate and can model the main effects on vehicle dynamics. In the considered velocity simulation, see Section 3.1, only the longitudinal motion of the vehicle is taken into account, making linear single-track models sufficient. For the simplified model used in the following, the resistance forces acting on a vehicle are [13]:

- Air resistance $F_{air}(v) = \frac{\rho_{air}}{2} c_w A v^2$ with air density ρ_{air}, coefficient of drag c_w and frontal area A,
- Rolling resistance $F_{roll}(s, v) = m_{veh} g f_{roll} \cos(\alpha(s))$ with vehicle mass m_{veh}, acceleration due to gravity g, rolling resistance coefficient f_{roll} and road slope $\alpha(s)$,
- Slope resistance $F_{slope}(s) = m_{veh} g \sin(\alpha(s))$.

Using the model above, the total driving resistance is $F_{res}(s,v) = F_{air}(v) + F_{roll}(s,v) + F_{slope}(s)$. Figure 2.1 illustrates the set of resistance forces acting on the vehicle.

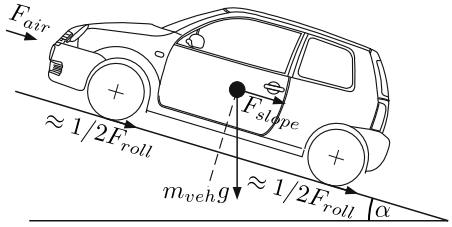


Figure 2.1: Resistance forces acting on vehicle, modified from [13]

The available traction force for accelerating the vehicle is given by

$$F_{tr} = \min(F_{tr,P}, F_{tr,T}, F_{tr,F}) \tag{2.1}$$

and is either limited by the engine power $F_{tr,P}$, the engine torque $F_{tr,T}$ or the road friction $F_{tr,F}$. The force limit resulting from the engine power P_{max} is

$$F_{tr,P} = \frac{P_{max}}{v r_c} \text{ with } r_c = 1 + \frac{\lambda_{max} - 1}{10^{v/v_{max}}} \tag{2.2}$$

where r_c is a drivetrain efficiency factor with the two parameters λ_{max} and v_{max} [94]. The engine torque limit follows directly from the transmission ratio n_g, the wheel diameter r_{wheel} and the maximum torque at the engine shaft $T_{eng,max}$, i.e.

$$F_{tr,T} = \frac{n_g T_{eng,max}}{r_{wheel}}. \tag{2.3}$$

The maximum force that the friction of the road can transmit is [86]

$$F_{tr,F} = m_{veh} \frac{n_{dr,ax}}{n_{ax}} g \mu_k cos(\alpha) \tag{2.4}$$

with the ratio of driven axles $n_{dr,ax}$ to all axles n_{ax} of the vehicle, the friction coefficient μ_k, and the road slope α, see Figure 2.1. The weight is assumed to be evenly distributed between both axles, and weight shift effects during acceleration are ignored. Weight shift effects occur when acceleration or braking causes the vehicle's weight to transfer from one axle to another. These simplifications allow for splitting the weight between the axles using $m_{veh} \frac{n_{dr,ax}}{n_{ax}}$. Figure 2.2 shows an example of the traction force diagram. At lower velocities, the traction force is limited by either the engine torque or the road friction, depending on the vehicle and the road friction conditions, while at higher velocities, the traction force is limited by the engine power.

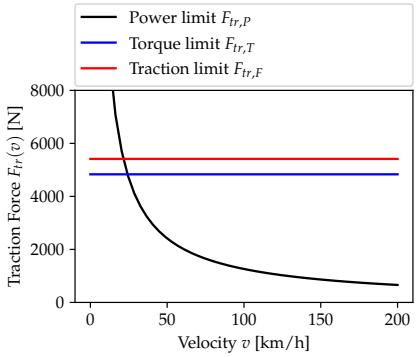


Figure 2.2: Traction force diagram

The traction force for decelerating the vehicle is derived in the same way by replacing the maximum torque and power values of the engine with those of the braking system. From the available traction force and the driving resistance, the maximum possible acceleration of the vehicle follows as

$$a_{x,max,veh}(s,v) = \frac{F_{tr} - F_{res}(s,v)}{m_{veh}}. \tag{2.5}$$

2.2 Autoregressive moving-average models

ARMA models are a class of linear, discrete models for stochastic processes that are widely used in economics, social sciences, and engineering [26, 65, 84]. In Section 2.2.1, ARMA models are defined, and their properties are briefly described. Section 2.2.2 introduces methods to estimate parameters for the models from data, based on which Section 2.2.3 describes ways to choose the model with the best fit to the data. For further insights on ARMA models and time series analysis in general, the reader is referred to [12].

2.2.1 Definition and properties

An ARMA model [71] is a combination of an Autoregressive (AR) and a Moving-Average (MA) model. An AR(p) model of order p describes a stochastic process $\{Z_t\}_{t=1}^{N}$ in the following way

$$Z_t = c + \epsilon_t + \sum_{i=1}^{p} \varphi_i Z_{t-i} \tag{2.6}$$

with the AR coefficients φ_i up to order p, the constant c and the white-noise term $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$. A MA(q) model describes a stochastic process $\{Z_t\}_{t=1}^{N}$ by

$$Z_t = c + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{2.7}$$

with the MA coefficients θ_j up to order q, the remaining terms are defined in the same way as in the AR(p) model. This way an ARMA(p, q) model is defined accordingly as a combination of an AR(p) and a MA(q) model leading to the description

$$Z_t = c + \epsilon_t + \sum_{i=1}^{p} \varphi_i Z_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}. \tag{2.8}$$

of a stochastic process $\{Z_t\}_{t=1}^{N}$.

Based on the general ARMA model we can define a red-noise AR(1) model, see [11], i.e., the white noise term ϵ_t in the AR model is replaced by red-noise which is itself an AR(1) process with zero mean

$$\begin{aligned} &Z_t = c + \eta Z_{t-1} + \epsilon_t \text{ with } \epsilon_t = \theta \epsilon_{t-1} + \zeta_t, \\ &Z_t = c + \eta Z_{t-1} + \theta \epsilon_{t-1} + \zeta_t, \end{aligned} \tag{2.9}$$

where ζ_t is a white-noise term $\zeta_t \sim \mathcal{N}(0, \sigma_{\zeta_t})$. In this case, the process is symmetric in the two autocorrelation parameters η and θ, i.e., the properties of the process, such as mean, variance, and autocorrelation, remain the same when the parameters are

swapped. Starting from the initial values Z_1 and ϵ_1 and iteratively stepping through the process

$$\begin{aligned}
Z_2 &= c + \eta Z_1 + \overbrace{\theta\epsilon_1 + \zeta_2}^{\epsilon_2}, \\
Z_3 &= c + \eta \underbrace{(\eta Z_1 + \theta\epsilon_1 + \zeta_2)}_{Z_2} + \overbrace{\theta \underbrace{(\theta\epsilon_1 + \zeta_2)}_{\epsilon_2} + \zeta_3}^{\epsilon_3}, \\
Z_4 &= c + \eta \underbrace{(\eta^2 Z_1 + \eta\theta\epsilon_1 + \eta\zeta_2 + \theta^2\epsilon_1 + \theta\zeta_2 + \zeta_3)}_{Z_3} + \overbrace{\theta \underbrace{(\theta^2\epsilon_1 + \theta\zeta_2 + \zeta_3)}_{\epsilon_3} + \zeta_4}^{\epsilon_4}, \\
Z_4 &= c + Z_1\eta^3 + \epsilon_1(\eta^2\theta + \eta\theta^2 + \theta^3) + \zeta_4 + \zeta_3(\eta + \theta) + \zeta_2(\eta + \theta)^2,
\end{aligned} \tag{2.10}$$

the red-noise AR(1) process can be generalized as

$$Z_t = c + Z_1\eta^{t-1} + \epsilon_1 \sum_{i=1}^{t-1} \eta^{t-1-i}\theta^i + \sum_{i=2}^{t} \zeta_i(\eta + \theta)^{t-i}. \tag{2.11}$$

For $|\eta| < 1, |\theta| < 1$ the influence of the starting values Z_1 and ϵ_1 quickly diminishes and a MA process with infinite order remains where the coefficients are $\theta_j = (\eta + \theta)^j$. This is in line with all AR processes, as they can generally be written as a MA process of infinite order. Therefore, a red-noise AR(1) process can be approximated by a proper truncation of an infinite order MA process.

Stochastic processes can also be represented as an output from a linear filter [107] which transforms a sequence of independent random variables $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, ...$ (white noise process) into the process Z_t

$$\begin{aligned}
Z_t &= c + \epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + ..., \\
Z_t &= c + \psi(\boldsymbol{B})\epsilon_t,
\end{aligned} \tag{2.12}$$

where the polynomial operator is given by

$$\psi(\boldsymbol{B}) = 1 + \psi_1\boldsymbol{B} + \psi_2\boldsymbol{B}^2 + ... \tag{2.13}$$

with the weight sequence $\psi_1, \psi_2, ...$ and the backward shift operator $\boldsymbol{B}Z_t = Z_{t-1}$. The ARMA model, in (2.8), can thus alternatively be written as

$$(1 - \varphi_1\boldsymbol{B} - \varphi_2\boldsymbol{B}^2 - ... - \varphi_p\boldsymbol{B}^p)Z_t = (1 + \theta_1\boldsymbol{B} + \theta_2\boldsymbol{B}^2 + ... + \theta_q\boldsymbol{B}^q)\epsilon_t \tag{2.14}$$

or more compactly with the polynomial operators $\varphi(\boldsymbol{B})$ and $\theta(\boldsymbol{B})$ of degrees p and q by

$$\varphi(\boldsymbol{B})Z_t = \theta(\boldsymbol{B})\epsilon_t. \tag{2.15}$$

Stationarity

The ARMA model in (2.15) defines a stationary process if the characteristic equation $\varphi(\boldsymbol{B})$ has all its roots outside the unit circle, and thus the stationarity is independent of the MA terms. For an AR(1) process, this implies that the condition $|\varphi_1| < 1$ must be fulfilled for stationarity. For an AR(2) process the stationarity conditions are $\varphi_2 + \varphi_1 < 1$, $\varphi_2 - \varphi_1 < 1$ and $-1 < \varphi_2 < 1$. For the red-noise AR(1) the condition follows from the combination of the two AR(1) processes and thus implies $|\eta| < 1$ and $|\theta| < 1$, see also [11].

Moments

The Autocorrelation function (ACF) for a general ARMA(p,q) model is the following difference equation

$$\begin{aligned} \rho_\tau &= \varphi_1\rho_{\tau-1} + \varphi_2\rho_{\tau-2} + ... + \varphi_p\rho_{\tau-p}, \text{ for } \tau \geq q+1, \\ \varphi(\boldsymbol{B})\rho_\tau &= 0, \text{ for } \tau \geq q+1. \end{aligned} \tag{2.16}$$

In the case of an ARMA(1,1) model the ACF is

$$\begin{aligned} \rho_1 &= \frac{(1 - \varphi_1\theta_1)(\varphi_1 - \theta_1)}{1 + \theta_1^2 - 2\varphi_1\theta_1}, \text{ and} \\ \rho_\tau &= \varphi_1\rho_{\tau-1}, \text{ for } \tau \geq 2, \end{aligned} \tag{2.17}$$

which has a monotone exponential decay for $\varphi_1 > 0$ and an alternating exponential decay for $\varphi_1 < 0$. For the red-noise AR(1) model the ACF is given by (e.g., [11])

$$\rho_\tau = \eta^\tau + \frac{1}{1 + \eta\theta}\sum_{i=1}^{\tau}\eta^{\tau-i}\theta^i(1 - \eta^2). \tag{2.18}$$

Following [11] the mean value of any ARMA model is given by

$$\mathbb{E}[Z_t] = \frac{c}{1 - \sum_{i=1}^{p} \varphi_i} \tag{2.19}$$

and for the red-noise AR(1) in particular by

$$\mathbb{E}[Z_t] = \frac{c}{1 - \eta}. \tag{2.20}$$

In general, the variance of an ARMA process is

$$\mathrm{Var}(Z_t) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2 \tag{2.21}$$

which when solved for an ARMA(1,1) process leads to

$$\mathrm{Var}(Z_t) = \frac{1 + 2\varphi_1\theta_1 + \theta_1^2}{1 - \varphi_1^2}\sigma_\epsilon^2. \tag{2.22}$$

Following [11] the variance of the red-noise AR(1) process is then given by

$$\mathrm{Var}(Z_t) = \frac{\sigma_\zeta^2}{(1 - \eta^2)(1 - \theta^2)} \left(\frac{1 + \eta\theta}{1 - \eta\theta}\right). \tag{2.23}$$

2.2.2 Parameter estimation

There exist several methods to estimate the parameters of an ARMA model. In the following, a few of them are introduced where we focus on the (conditional) Maximum Likelihood Estimation (MLE), the least squares method, and the method of moments (e.g., Yule-Walker equations).

Maximum Likelihood Estimation

Using (2.8) the ARMA(p,q) model can be written as

$$\epsilon_t = Z_t - c - \varphi_1 Z_{t-1} - ... - \varphi_p Z_{t-p} - \theta_1\epsilon_{t-1} - ... - \theta_q\epsilon_{t-q}. \tag{2.24}$$

Assuming that the first p values of $\{Z_t\}_{t=1}^{p}$ and q values of $\{\epsilon_t\}_{t=1}^{q}$ are known then the values $\epsilon_t(\varphi_1, ..., \varphi_p, \theta_1, ..., \theta_q)$ can be calculated recursively using (2.24). Conditional

on the starting values $\{Z_t\}_{t=1}^{p}$, $\{\epsilon_t\}_{t=1}^{q}$ and using the normality and independence assumption of ϵ_t the joint likelihood of the ARMA parameters is determined by

$$\mathcal{L}_{Z_1,\ldots,Z_t}(c,\sigma_\epsilon^2,\varphi_1,\theta_1,\ldots) = \prod_{t=t_{min}}^{N} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_\epsilon^2}\right). \tag{2.25}$$

The starting point is set to $t_{min} = \max(p,q)$ to allow the calculation of ϵ_t using (2.24), which for $t < t_{min}$ is not possible as the equation is not fully defined. The log-likelihood is given by

$$\begin{aligned}\ell_{Z_1,\ldots,Z_t}(c,\sigma_\epsilon^2,\varphi_1,\theta_1,\ldots) &= \log(\mathcal{L}_{Z_1,\ldots,Z_t}),\\ &= -\frac{N-t_{min}}{2}\log(2\pi) - \frac{N-t_{min}}{2}\log\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{t=t_{min}}^{N}\epsilon_t^2,\end{aligned} \tag{2.26}$$

which, accordingly, is also conditional on the unknown starting values. To calculate the log-likelihood, assumptions must be made about the starting values. Here, a sensible choice is to set them to their unconditional expectation, $\mathbb{E}[Z_t] = c$ and $\mathbb{E}[\epsilon_t] = 0$. For reasonably large values of N the influence of the starting values diminishes.

In the case of a red-noise AR(1) model we can expand

$$\begin{aligned}Z_t &= c + \eta Z_{t-1} + \epsilon_t,\\ Z_t &= c + \eta Z_{t-1} + \theta\epsilon_{t-1} + \zeta_t,\\ Z_t &= c + \eta Z_{t-1} + \theta(Z_{t-1} - \eta Z_{t-2} - c) + \zeta_t,\end{aligned} \tag{2.27}$$

which leads to the (unconditional) log-likelihood needed for the MLE of the red-noise AR(1) parameters

$$\begin{aligned}&\ell_{Z_1,\ldots,Z_t}(c,\sigma_\zeta^2,\eta,\theta)\\ &= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\left(\sigma_\zeta^2\right) - \frac{1}{2\sigma_\zeta^2}\sum_{t=1}^{N}(Z_t - (c + \eta Z_{t-1} + \theta(Z_{t-1} - \eta Z_{t-2} - c)))^2.\end{aligned} \tag{2.28}$$

In order to obtain the MLE we use a numerical optimization algorithm (e.g., BFGS [28] or trust-region [95]) to maximize the log-likelihood

$$\hat{c},\hat{\sigma}_\epsilon^2,\hat{\varphi}_1,\hat{\theta}_1,\ldots = \underset{c,\sigma_\epsilon^2,\varphi_1,\theta_1,\ldots}{\arg\max}\ \ell_{Z_1,\ldots,Z_t}(c,\sigma_\epsilon^2,\varphi_1,\theta_1,\ldots). \tag{2.29}$$

The MLE is consistent and converges in probability to the true parameter values, but is shown to be biased [12], meaning that the expected value of the estimator differs from the true value.

Least squares method

As the considered ARMA models assume a normal distribution of ϵ_t the log-likelihood includes the data through the sum of squares

$$S(c, \varphi_1, \theta_1, ...) = \sum_{t=t_{min}}^{N} \epsilon_t^2, \tag{2.30}$$

see (2.26). For a fixed σ_ϵ^2 the log-likelihood $\ell_{Z_1,...,Z_t}$ is therefore a linear function of S. It follows that maximizing the log-likelihood for the MLE is equivalent to minimizing the sum of squares, making the least squares estimate equivalent to the MLE.

Method of moments

To build a method of moments estimator, the moments of the stochastic process are expressed as a function of the model parameters and compared to the sample moments. The number of moments included in the estimation is the same as the number of model parameters. This leads to a system of equations, which is solved for the model parameters. For ARMA models, the ACF, i.e., the second moment, can be used to build the system of equations. A sample ACF estimator is given by

$$\hat{\rho}_\tau = \frac{1}{(N-\tau)\sigma_z^2} \sum_{t=1}^{N-\tau} (Z_t - \mu_z)(Z_{t+\tau} - \mu_z) \tag{2.31}$$

which is biased when μ_z and σ_z^2 are replaced by their estimators $\hat{\mu}_z, \hat{\sigma}_z^2$. For a pure AR model, the ACF as a function of the model parameters is given by

$$\rho_\tau = \sum_{i=1}^{p} \varphi_i \rho_{i-\tau}. \tag{2.32}$$

In the case of an AR model, this gives a system of equations called the Yule-Walker equations, which for an AR(1) model take the form

$$\begin{aligned} \rho_1 &= \varphi_1, \\ \hat{\varphi}_1 &= \hat{\rho}_1, \end{aligned} \tag{2.33}$$

where the autocorrelation ρ_1 is replaced by $\hat{\rho}_1$ to estimate φ_1.

For general ARMA models a method of moments estimator for the AR and MA parameters $\hat{\varphi}_1, \hat{\theta}_1, ...$ can be derived by minimizing the Mean squared error (MSE) between the theoretical and estimated ACF [72] which leads to the estimators

$$\hat{\varphi}_1, \hat{\theta}_1, ... = \underset{\varphi_1, \theta_1, ...}{\arg\min} \sum_{\tau=1}^{\tau_{\max}} \left(\hat{\rho}_t - \rho_t\right)^2 \tag{2.34}$$

with the maximum considered lag $\tau = \tau_{\max}$. To also estimate the mean c and the variance σ_ϵ^2 the expected value $\mathbb{E}[Z_t]$ in (2.19) is substituted by the sample mean $\hat{\mu}_z$

$$\begin{aligned} \mathbb{E}[Z_t] &= \frac{c}{1 - \sum_{i=1}^{p} \varphi_i}, \\ \hat{c} &= \hat{\mu}_z (1 - \sum_{i=1}^{p} \hat{\varphi}_i), \end{aligned} \tag{2.35}$$

and the variance $\mathrm{Var}(Z_t)$ in (2.21) is substituted by the sample variance $\hat{\sigma}_z^2$

$$\begin{aligned} \mathrm{Var}(Z_t) &= \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2, \\ \hat{\sigma}_\epsilon^2 &= \frac{\hat{\sigma}_z^2}{\sum_{i=0}^{\infty} \hat{\psi}_i^2}. \end{aligned} \tag{2.36}$$

In [72], it is noted that the presented method of moments estimator based on the ACF converges to the true parameter values, but is generally biased.

2.2.3 Model selection

Model selection aims at finding the optimal AR and MA orders to model a real stochastic process. The ACF and Partial autocorrelation function (PACF) of the measured sample can be used to identify suitable orders. An ACF estimator $\hat{\rho}_\tau$ is given in (2.31). In [12] the

authors derive a PACF estimator $\hat{\varphi}_{\tau,\tau}$ for lag τ recursively based on the Durbin–Levinson Algorithm [25] and the ACF estimator $\hat{\rho}_\tau$ by

$$\hat{\varphi}_{\tau,\tau} = \frac{\hat{\rho}_\tau - \sum_{t=1}^{\tau-1} \hat{\varphi}_{\tau-1,t}\hat{\rho}_{\tau-t}}{1 - \sum_{t=1}^{\tau-1} \hat{\varphi}_{\tau-1,t}\hat{\rho}_\tau} \text{ where } \hat{\varphi}_{\tau,t} = \hat{\varphi}_{\tau-1,t} - \hat{\varphi}_{\tau,\tau}\hat{\varphi}_{\tau-1,\tau-t} \text{ for } 1 \leq t \leq \tau - 1 \quad (2.37)$$

with $\hat{\varphi}_{p,i}$ being the ith coefficient of an AR model of order p estimated with the Yule-Walker equations, see (2.33). For a pure AR process, the PACF has a cutoff after lag p while the ACF decays exponentially. The ACF of a purely MA process has a cutoff after lag q while the PACF decays exponentially. For a mixed ARMA model, both the ACF and PACF show an exponential decay. With the identified model orders p and q, the parameters can be estimated, as described in Section 2.2.2.

The remaining residual ϵ_t is investigated to assess the quality of the model fit. The residual should have a constant mean and variance, i.e., it should be stationary, and the values should be independent of each other. The ACF and PACF can be used to identify remaining correlations in the residual. Statistical tests like the Ljung–Box test [61] can be applied to test for correlations. Stationarity of the series can be tested with hypothesis tests, e.g., the augmented Dickey–Fuller test [23]. If the chosen model with its parameters is inadequate, the procedure is repeated with another model specification. This iterative procedure to identify the right ARMA model for a time series is known as the Box–Jenkins method [12].

An alternative approach for model comparison is to use an information criterion. Two widely used and closely related criteria are the Akaike information criterion (AIC) [1] and Bayesian information criterion (BIC) [87], which are both based on the likelihood function $\mathcal{L}$ of the model and include a penalty term to avoid overfitting due to too many parameters. The AIC is defined as

$$\text{AIC} = 2k - 2\log\left(\hat{\mathcal{L}}\right) \quad (2.38)$$

with the related BIC given by

$$\text{BIC} = k\log(N) - 2\log\left(\hat{\mathcal{L}}\right). \quad (2.39)$$

Here, $\hat{\mathcal{L}}$ is the maximized likelihood, k is the number of parameters, and N is the sample size. The two criteria differ in their penalty term for the number of parameters. The term in the BIC also includes the sample size and is bigger than the AIC for sample sizes greater than 7. Models with a lower value of the criteria are generally preferred. Both criteria are only valid for model comparison and do not provide an absolute criterion for

rating the quality of a model. Other analyses should still be part of the model selection process, e.g., examination of the residuals and model prediction quality.

2.3 Uncertainty Quantification

The aim of UQ is to quantify the uncertainties in computational and real-world applications. Understanding these uncertainties helps to aid in the decision-making process and leads to more confidence in respective decisions. Introductions to UQ are for example available in [20, 36, 64, 75, 92, 97].

Uncertainties can be classified into different categories. A common approach is to classify them into aleatoric and epistemic uncertainties [97]. Aleatoric uncertainty is uncertainty about an inherently variable phenomenon and can thus not be reduced. Commonly, aleatory uncertainties are described by parametric distributions. Epistemic uncertainty refers to uncertainty from a lack of knowledge, i.e., it can be reduced by acquiring more knowledge. Epistemic uncertainty can be further divided into model form uncertainty concerning the correct structure of the model and parametric uncertainty about the correct values of the model parameters. The distinction of uncertainties in these two categories is not always clear and can be subjective. The presence of both aleatoric and epistemic uncertainties is commonly referred to as polymorphic uncertainty.

In [97], an overview of objectives that can be solved with methods from UQ is given. They are illustrated with the system f (also denoted as model $\mathcal{M}$) which maps inputs $\boldsymbol{X} = (X_1, ..., X_K)^T$ from some metric space $\boldsymbol{\mathcal{X}}$ to outputs $\boldsymbol{Y} = (Y_1, ..., Y_M)^T = f(\boldsymbol{X})$ in some metric space $\boldsymbol{\mathcal{Y}}$. The output $\boldsymbol{Y}$ can be a single (or multiple), direct output of the system f or another quantity derived from the output(s) to make the problem easier to handle. The output quantities considered in the UQ problem are called Quantities of Interest (QoIs). The following UQ objectives are relevant to this thesis:

- **Forward propagation problem:** Uncertainty about the input parameters is described by a probability distribution $p_{\boldsymbol{X}}(\boldsymbol{x})$ which is propagated to the outputs $\boldsymbol{Y}$. The task is to find the output distribution $p_{\boldsymbol{Y}}(\boldsymbol{y})$ and its properties. Difficulties of this task can be associated with the high dimensionality of either $\boldsymbol{X}$ or $\boldsymbol{Y}$, often also with correlations within inputs and outputs or the black box character of the system f which prevents an analytical description of $p_{\boldsymbol{Y}}(\boldsymbol{y})$.

- **Reliability problem:** Estimate the probability of the QoI belonging to a certain set $\mathcal{Y}_{fail}$ associated with failure

$$\mathbb{P}[f(\boldsymbol{X}) \in \mathcal{Y}_{fail}] = \epsilon. \tag{2.40}$$

- **Prediction problem:** Similar to the reliability problem. Given a maximum acceptable probability of error $\epsilon > 0$ find the set $\mathcal{Y}_\epsilon$ such that failure is avoided with the given probability

$$\mathbb{P}[f(\boldsymbol{X}) \in \mathcal{Y}_\epsilon] \geq 1 - \epsilon. \tag{2.41}$$

- **Inverse problem:** Given (uncertain) observations of the output quantity $\boldsymbol{Y}$ estimate the unobserved input parameters $\boldsymbol{X}$ such that $f(\boldsymbol{X}) = \boldsymbol{Y}$. In a frequentist view, confidence intervals for the parameters can be estimated, while in a Bayesian context, a probability distribution $p_X(\boldsymbol{x})$ is assigned to the input parameters to express their uncertainty.
- **Model reduction:** Replace the full system f by an approximation $\hat{f}$ which itself has some numerical parameters that need to be calibrated to give a good approximation $\hat{f} \approx f$. The approximation (also called surrogate or meta model) usually has less computational cost or some other advantages compared to the full system regarding the UQ objective. The accuracy of the approximation may itself be part of a reliability or prediction problem [97].

2.3.1 Monte Carlo method

Monte Carlo methods [69] are a collection of tools used to solve problems by drawing repeated samples from random variables. In UQ, they can be used to derive properties of the QoI distribution without knowing an analytical description of it. The expected value of the model output

$$\mathbb{E}[f(\boldsymbol{X})] = \int_{\mathcal{X}} p_X(\boldsymbol{x}) f(\boldsymbol{X}) d\boldsymbol{x} \tag{2.42}$$

for example, can be approximated with a (standard) Monte Carlo method by replacing the expected value with the sample mean and drawing N samples from the probability distribution $p_{\boldsymbol{X}}(\boldsymbol{x})$

$$\mathbb{E}[f(\boldsymbol{X})] \approx \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{X}). \tag{2.43}$$

The variance of the unknown distribution $p_{\boldsymbol{Y}}(\boldsymbol{y})$ can be approximated similarly with the sample variance

$$\mathrm{Var}(f(\boldsymbol{X})) \approx \hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f(\boldsymbol{X}) - \hat{\mu})^2. \tag{2.44}$$

The u-th quantile q_u can be approximated by taking the ranked sample $\boldsymbol{X}_{(1)}, \ldots, \boldsymbol{X}_{(N)}$ and using the sample quantile function

$$\hat{q}_u = \hat{Q}(u) = (1-\gamma)\boldsymbol{X}_j + \gamma \boldsymbol{X}_{j+1}, \text{ where } \frac{j}{N} \leq u < \frac{j+1}{N} \tag{2.45}$$

for $0 \leq \gamma \leq 1$ [47]. There exist different definitions for γ which depend on the value of $j = \lfloor uN \rfloor$ and $g = uN - j$. The simplest form is

$$\gamma = \begin{cases} 1 & \text{if } g > 0 \\ 0 & \text{if } g = 0 \end{cases} \tag{2.46}$$

which assigns the next higher sample value if the quantile is between two values of the ranked sample set. The authors in [47] give other, alternative definitions for γ.

Monte Carlo methods have several advantages, e.g., the independence of the estimates, such as $\hat{\mu}$, from the dimensionality K of the input space, making them applicable to higher-dimensional problems. The independence of the evaluations of $f(\boldsymbol{X})$ allows embarrassingly parallel calculations. A disadvantage of Monte Carlo methods is their slow convergence rate $O\left(\frac{1}{\sqrt{N}}\right)$, requiring a large number of samples. Various methods to improve the convergence rate of the standard Monte Carlo method exist, e.g., Latin Hypercube Sampling (LHS) [67], Multilevel Monte Carlo (MLMC) [37], and the quasi-Monte Carlo method [17] which converges with $O\left(\frac{\log(N)^K}{N}\right)$. In [56], the properties of the mentioned and other variants of Monte Carlo methods are compared. Other variants of the Monte Carlo method include variance reduction techniques such as importance sampling [54] which, based on knowledge about $p_{\boldsymbol{Y}}(\boldsymbol{y})$ with $\boldsymbol{y} = f(\boldsymbol{x})$

gives more weight to some samples from $\boldsymbol{X}$ depending on their importance. For more information on Monte Carlo methods, the reader is referred to [14, 63].

2.3.2 Markov Chain Monte Carlo method

A Monte Carlo method of particular importance for Bayesian analysis is the Markov Chain Monte Carlo (MCMC) method. An introduction to it is given in [24, 63]. In Bayesian applications, samples need to be drawn from the posterior distribution, see (2.80), which can be high-dimensional and is only known up to a constant Z_p^*, meaning that $p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{Z_p^*} p_{\boldsymbol{X}}^*(\boldsymbol{x})$. Monte Carlo methods can be used to approximate the integral

$$\begin{aligned} Z_p^* &= \int_{\mathcal{X}} p_{\boldsymbol{X}}^*(\boldsymbol{x}) dx, \\ Z_p^* &\approx V \frac{1}{N} \sum_{i=1}^{N} p_{\boldsymbol{X}}^*(\boldsymbol{x}), \text{ where } V = \int_{\mathcal{X}} d\boldsymbol{x}, \end{aligned} \tag{2.47}$$

to estimate the constant Z_p^* or to draw samples from the posterior distribution to approximate its shape. Distributions in high dimensions are typically concentrated in a small subset of the probability space $\boldsymbol{\mathcal{X}}$. Independent, uniform sampling thus places only a few samples in the relevant subset, making its use of the samples inefficient.

The general idea of MCMC methods is to draw correlated samples that depend on previous draws. New samples are drawn from a transition distribution $T(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i) = p_{\boldsymbol{X}_i}(\boldsymbol{x}_i | \boldsymbol{x}_{i-1})$ which conditions the new sample $\boldsymbol{X}_i$ on the previous one $\boldsymbol{X}_{i-1}$ (i.e., a Markov Chain). The transitions $T(\boldsymbol{x}', \boldsymbol{x})$, where $\boldsymbol{x}'$ is the previous sample of $\boldsymbol{x}$, must be chosen in a way that the resulting distribution

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{\boldsymbol{x}'} T(\boldsymbol{x}', \boldsymbol{x}) p_{\boldsymbol{X}'}(\boldsymbol{x}') \tag{2.48}$$

is invariant under the Markov Chain. In [24] invariance of $p_{\boldsymbol{X}}(\boldsymbol{x})$ is checked by the reversibility condition

$$p_{\boldsymbol{X}}(\boldsymbol{x}) T(\boldsymbol{x}, \boldsymbol{x}') = p_{\boldsymbol{X}'}(\boldsymbol{x}') T(\boldsymbol{x}', \boldsymbol{x}) \tag{2.49}$$

requiring the transitions $T(\boldsymbol{x}', \boldsymbol{x})$ to be chosen such that the Markov Chain is reversible for $p_{\boldsymbol{X}}(\boldsymbol{x})$. Using this condition invariance of $p_{\boldsymbol{X}}(\boldsymbol{x})$ under the Markov Chain can be shown by

$$\sum_{\boldsymbol{x}'} p_{\boldsymbol{X}'}(\boldsymbol{x}') T(\boldsymbol{x}', \boldsymbol{x}) = \sum_{\boldsymbol{x}'} p_{\boldsymbol{X}}(\boldsymbol{x}) T(\boldsymbol{x}, \boldsymbol{x}') = p_{\boldsymbol{X}}(\boldsymbol{x}) \sum_{\boldsymbol{x}'} T(\boldsymbol{x}, \boldsymbol{x}') = p_{\boldsymbol{X}}(\boldsymbol{x}). \tag{2.50}$$

As demonstrated, the reversibility condition is sufficient for invariance of $p_{\boldsymbol{X}}(\boldsymbol{x})$ under the Markov Chain, but it is not a necessary condition [24].

When drawing from the Markov Chain, several initial steps are needed until the distribution converges to the invariant distribution (burn-in). Various MCMC algorithms use different transition distributions $T(\boldsymbol{x}', \boldsymbol{x})$ to draw the next sample. These include the Metropolis-Hastings [42], Gibbs [34], and Slice [74] sampling methods. Slice sampling has the advantage that it does not require the choice of a particular transition distribution. It is based on the fact that sampling from a distribution $p_{\boldsymbol{X}}(\boldsymbol{x})$ is identical to uniformly sampling from the area underneath the Probability Density Function (PDF) of such a distribution and marginalizing over the density axis to retain $p_{\boldsymbol{X}}(\boldsymbol{x})$. This also holds for the unnormalized distribution $p^*_{\boldsymbol{X}}(\boldsymbol{x}) = Z^*_p p_{\boldsymbol{X}}(\boldsymbol{x})$.

To sample from the area under the PDF an auxiliary random variable U supported on $u \in [0, p^*_{\boldsymbol{X}}(\boldsymbol{x})]$ is defined and the joint distribution of $\boldsymbol{X}$ and U is described by

$$p_{\boldsymbol{X},U}(\boldsymbol{x}, u) = \begin{cases} \frac{1}{Z^*_p} & \text{if } 0 \leq u \leq p^*(\boldsymbol{x}) \\ 0 & \text{otherwise} \end{cases}. \tag{2.51}$$

Sampling from $p_{\boldsymbol{X},U}(\boldsymbol{x}, u)$ is performed by alternatively sampling from $\boldsymbol{x}$ and u conditional on the other variable. Figure 2.3 illustrates this procedure for the univariate case. The distribution $p_U(u|x)$ is given by the uniform distribution $p_U(u|x) \sim \mathcal{U}(0, p^*_X(x))$ from which samples can easily be drawn. Sampling from $p_X(x|u)$ is achieved by sampling uniformly from the set $\{x : u < p^*_X(x)\}$, which is the horizontal "slice" in Figure 2.3. The size of the set $\{x : u < p^*_X(x)\}$ is generally not known. The horizontal sampling is performed by defining a region of size w around the point x. This region is extended by w in each direction until both ends are outside the slice, see the lower part of Figure 2.3. Then, uniform sampling within this region is performed. If the drawn sample is within the set $\{x : u < p^*_X(x)\}$, it is accepted, if this is not the case, it is set as the new lower or upper bound for the sampling region. This procedure is repeated until a valid sample for x is found, which then serves as the condition for the next draw from $p_U(u|x)$. This iterative procedure is repeated until the targeted number of samples is drawn.

2.3.3 Sensitivity analysis

Sensitivity analysis investigates how a function $f(x_1, ..., x_K)$ is influenced by the individual inputs x_i as well as their combined effects, which allows for identifying their significance. This helps to reduce the set of parameters $\{x_1, ..., x_K\}$ to only the

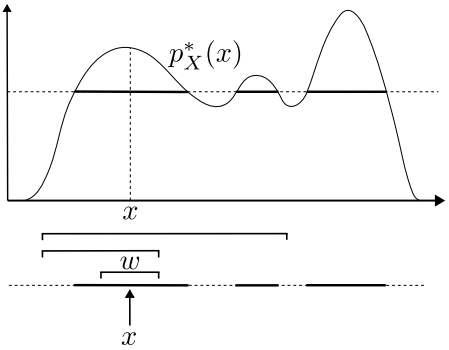


Figure 2.3: Illustration of the slice sampling algorithm. Figure modified from [74].

important ones, which enables further steps such as parameter identification. Sensitivity analysis can be categorized into local and global methods. Local methods study the sensitivity of $f(x_1, ..., x_K)$ at a particular point $\boldsymbol{x}^0$ which involves the calculation of derivatives. For instance, the partial derivative of f with respect to x_i at the fixed point $\boldsymbol{x}^0$ provides insight into how small changes in x_i affect the function's output near that specific point. Global methods, on the other hand, study the overall effect of an input x_i on $f(x_1, ..., x_K)$, which takes their complete definition domain into account and doesn't focus on the effect at a particular point. Due to this global perspective, which makes consideration of the input domain size easier, we focus on global methods in this work.

A popular method for global sensitivity analysis is the variance-based Sobol' method [89]. Sobol' indices decompose the output variance into the individual and combined contributions of the input parameters. Considering a univariate model $Y = f(\boldsymbol{X})$ (extension to multiple outputs possible by performing separate analysis for each output) with uncertain, random inputs $\boldsymbol{X} = (X_1, ..., X_K)^T$, the model can be decomposed as

$$Y = f_0 + \sum_{i=1}^{K} f_i(X_i) + \sum_{i<j}^{K} f_{ij}(X_i, X_j) + ... + f_{1,2,...,K}(X_1, X_2, ..., X_K) \tag{2.52}$$

where all the terms $f_0, f_i, f_{ij}, ..., f_{1,2,...,K}$ are orthogonal [89]. This leads to the following definitions of the conditional expected values

$$\begin{aligned} f_0 &= \mathbb{E}[Y], \\ f_i(X_i) &= \mathbb{E}[Y|X_i] - f_0, \\ f_{i,j}(X_i, X_j) &= \mathbb{E}[Y|X_i, X_j] - f_0 - f_i - f_j, \end{aligned} \tag{2.53}$$

which describe the effect on Y of varying X_i alone and the combined effect of varying multiple inputs simultaneously. The orthogonality allows for the following decomposition of the variance as

$$\mathrm{Var}(Y) = \sum_{i=1}^{K} V_i + \sum_{i<j}^{K} V_{ij} + \ldots + V_{12\ldots K} \tag{2.54}$$

where the individual terms

$$V_i = \mathrm{Var}(\mathbb{E}[Y|X_i]) \tag{2.55}$$

describe the contribution of a single input X_i to the total variance $\mathrm{Var}(Y)$. Higher-order contributions such as

$$V_{ij} = \mathrm{Var}(\mathbb{E}[Y|X_i, X_j]) - V_i - V_j \tag{2.56}$$

describe the combined effect of 2 or more input variables on the output variance. Sobol' indices [89] are defined as normalized versions of the variance contributions

$$\begin{aligned} S_i &= \frac{V_i}{\mathrm{Var}(Y)}, \\ S_{ij} &= \frac{V_{ij}}{\mathrm{Var}(Y)}, \\ &\ldots \\ S_{12\ldots K} &= \frac{V_{12\ldots K}}{\mathrm{Var}(Y)}, \end{aligned} \tag{2.57}$$

where S_i are first order, S_{ij} second order, etc. Sobol' indices. This implies that their sum is

$$1 = \sum_{i=1}^{K} S_i + \sum_{i<j}^{K} S_{ij} + \ldots + S_{12\ldots K}. \tag{2.58}$$

First and higher order Sobol' indices can be combined into the total order Sobol' indices [45]

$$S_{Ti} = \frac{\mathbb{E}[\mathrm{Var}(Y|\boldsymbol{X}_{\sim i})]}{\mathrm{Var}(Y)} = 1 - S_{\sim i} \tag{2.59}$$

where $\boldsymbol{X}_{\sim i}$ are all variables in $\boldsymbol{X}$ except X_i and $S_{\sim i}$ is the sum of all Sobol' indices which do not include i. The total order Sobol' indices therefore measure the total contribution

of X_i to the output variance $\mathrm{Var}(Y)$ across all orders and in combination with any other input variable. This also means that their sum is

$$\sum_{i=1}^{K} S_{Ti} \geq 1 \tag{2.60}$$

as higher order effects between, e.g., X_i and X_j, are included in both, S_{Ti} and S_{Tj}. The sum only equals 1 when the model doesn't include any higher-order effects.

Multiple estimators for the Sobol' indices exist, which allow us to estimate them from Monte Carlo samples, such as those derived in [82]. An accurate estimation often requires hundreds or thousands of runs, which requires techniques such as surrogate modeling to reduce the computational cost, see the following Section.

2.3.4 Surrogate models

Surrogate models are used to replace computationally expensive models with an approximation that is easier and faster to evaluate. The approximation is then used for subsequent, computationally expensive analyses, e.g., sensitivity analysis or parameter estimation. They are constructed by running a Design of Experiments (DoE) with the original model $\mathcal{M}$ to create a set of samples which are used to fit the surrogate $\hat{\mathcal{M}}$. Depending on the choice of the surrogate model, the DoE can be done with a random sampling strategy such as Monte Carlo, quasi Monte Carlo, Latin Hypercube Sampling, etc. Another sampling strategy is to use a deterministic sequence, e.g., full-factorial sampling or Non-Intrusive Spectral Projection (NISP) [57], which allows for a more systematic exploration of the input space, ensuring that certain combinations of inputs are considered, e.g., their extreme value combinations. Depending on the nature of the problem and the properties of the model $\mathcal{M}$, a suitable surrogate is chosen, although it should be noted that in practice this choice is often not known a priori. In principle, every universal function can be used as a surrogate model, but certain functions that have the necessary flexibility and structure to capture the underlying patterns in the data are more suitable. Popular approaches for surrogate models include Polynomial Chaos Expansion (PCE) [10, 35], Gaussian processes [78], Support Vector Machines [21, 101] and artificial neural networks [8, 81]. In the following, we focus on PCE based surrogate models.

We distinguish two types of simulators, standard simulators, which are deterministic, i.e., the same set of input parameters leads to the same QoI value at the output, and

stochastic simulators, which, for a set of input parameters, lead to a distribution of the QoI at the output. Formally, standard simulators can be expressed by

$$\mathcal{M}_d : \mathcal{D}_{\boldsymbol{X}} \rightarrow \mathbb{R} \tag{2.61}$$

which is a mapping of the inputs $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}}$ to the QoI $y = \mathcal{M}(\boldsymbol{x}) \in \mathbb{R}$, this can be extended to multiple QoIs at the output by treating each component separately. Most surrogate models are also deterministic and therefore suited for simulators of this form. Stochastic simulators have an internal source of randomness with sample space Ω and can formally be expressed as

$$\mathcal{M}_s : \mathcal{D}_{\boldsymbol{X}} \times \Omega \rightarrow \mathbb{R}. \tag{2.62}$$

This means that for a given set of input parameters $\boldsymbol{x}$, the model produces a realization from the unknown distribution $Y \sim p_Y(y|\boldsymbol{x})$. To characterize this unknown distribution, repeated runs of the model for a fixed set of parameters $\boldsymbol{x}$ are necessary, which further increases the computational cost when the effect of variations in $\boldsymbol{x}$ on the QoI should be investigated. This requires the use of stochastic surrogate models, which can emulate the QoI distribution $\hat{p}_Y(y|\boldsymbol{x}) \approx p_Y(y|\boldsymbol{x})$ and are easy and cheap to evaluate [109]. Different stochastic surrogate models are described in [62, 110, 112]. One PCE based approach for stochastic surrogate modeling suited to our applications is the GLaM [110].

Polynomial Chaos Expansion

PCE is based on the decomposition of a random variable

$$Y = \mathcal{M}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^K} c_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \tag{2.63}$$

with respect to the basis functions $\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$. This representation can be extended to a multivariate output $\boldsymbol{Y}$ by calculating the PCE for each output separately. Assuming that the components of $\boldsymbol{X}$ are independent, their joint distribution is given by

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{j=1}^{K} p_{X_j}(x_j). \tag{2.64}$$

For independent input $\boldsymbol{X}$, the basis functions $\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$ can be described by orthogonal polynomials which are obtained as a product of the individual polynomials [91]

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{j=1}^{K} \phi_{\alpha_j}^{(j)}(x_j). \tag{2.65}$$

The orthogonal polynomials $\{\phi_k^{(j)} : k \in \mathbb{N}\}$ for each input X_j are defined by

$$\mathbb{E}[\phi_k^{(j)}(X_j)\phi_l^{(j)}(X_j)] = \delta_{kl}, \tag{2.66}$$

with the Kronecker symbol δ defined as $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ otherwise. The components α_j indicate the polynomial degree of ϕ_{α_j}, where the multi-index $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_K) \in \mathbb{N}^K$ combines the polynomial degrees of all variables x_j. For standard distributions of X_j the associated polynomials $\phi_{\alpha_j}^{(j)}$ are known as Hermite (normal distribution), Legendre (uniform distribution), and Laguerre (exponential distribution) polynomials [106]. Truncation schemes are applied to select a subset of all available polynomials $\boldsymbol{\alpha} \in \mathcal{A}$. For a truncated set $\boldsymbol{\alpha}$ the coefficients $c_{\boldsymbol{\alpha}}$ can be estimated for a set of input samples $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_M\}$ and their respective outputs $\{y_1, ..., y_M\}$.

For a correlated distribution $p_{\boldsymbol{X}}(\boldsymbol{x})$, the usual procedure is to transform $\boldsymbol{X}$ into the proxy $\boldsymbol{\xi}$ with independent components using the Nataf or Rosenblatt transform [98]. After the transformation, the above PCE scheme can be applied to the independent components of $\boldsymbol{\xi}$.

With infinite polynomial degrees in $\boldsymbol{\alpha}$, the PCE gives an exact representation of the variable Y. For practical applications, the polynomial degrees need to be truncated to a finite subset $\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{N}^K$ to give an approximate surrogate model $\hat{\mathcal{M}} \approx \mathcal{M}$. This truncation can be done with, e.g., the hyperbolic (q-norm) truncation scheme [9]

$$\mathcal{A}^{p,q,K} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^K, ||\boldsymbol{\alpha}||_q = \left(\sum_{i=1}^{K} |\alpha_i|^q \right)^{\frac{1}{q}} \leq p \right\} \tag{2.67}$$

where p sets the maximum polynomial degree and $q \leq 1$ defines the quasi-norm $||\cdot||_q$. For $q = 1$ the full basis of total degrees less than p is obtained

$$\mathcal{A}^{p,q=1,K} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^K, ||\boldsymbol{\alpha}||_{q=1} = \sum_{i=1}^{K} |\alpha_i| \leq p \right\} \tag{2.68}$$

which includes the maximum coupling terms between the inputs $\boldsymbol{X} = (X_1, ..., X_K)^T$. When only weak or no coupling is expected between the inputs, the degree q can be reduced to decrease the overall number of terms. Table 2.1 gives two examples of the truncation of the polynomial order for two input variables X_1, X_2 with the q-norm truncation scheme.

(a) $p = 3$, $q = 1$, $K = 2$

α_1	0	0	1	0	2	1	0	3	1	2
α_2	0	1	0	2	0	1	3	0	2	1

(b) $p = 3$, $q = 0.5$, $K = 2$

α_1	0	0	1	0	2	0	3
α_2	0	1	0	2	0	3	0

Table 2.1: Polynomial orders for two combinations of p and q with $K = 2$

Generalized lambda distribution

GLaMs [110] are based on the Generalized Lambda Distribution (GLD). The GLD can well approximate most unimodal distributions, for example, uniform, normal, Weibull, and Student's t distributions [30, 52]. Different definitions of the GLD exist [30, 77]. In the following, we focus on the GLD of the Freimer-Kollia-Mudholkar-Lin family [30] which is defined with its quantile function

$$Q(u;\boldsymbol{\lambda}) = \lambda_1 + \frac{1}{\lambda_2}\left(\frac{u^{\lambda_3}-1}{\lambda_3} - \frac{(1-u)^{\lambda_4}-1}{\lambda_4}\right) \tag{2.69}$$

and the four parameters $\{\lambda_l : l = 1, ..., 4\}$. Here, λ_1 is the location parameter, λ_2 is the scale parameter, which must be positive to ensure a valid quantile function, and λ_3 and λ_4 are the shape parameters. Following [110] the PDF of the GLD can be derived based on the quantile function (2.69)

$$p_Y(y;\boldsymbol{\lambda}) = \frac{1}{Q'(u;\boldsymbol{\lambda})} = \frac{\lambda_2}{u^{\lambda_3-1}+(1-u)^{\lambda_4-1}}\mathbb{1}_{[0,1]}(u), \text{ with } u = Q^{-1}(y;\boldsymbol{\lambda}) \tag{2.70}$$

where $Q'(u;\boldsymbol{\lambda}) = \frac{dQ(u;\boldsymbol{\lambda})}{du}$, and $\mathbb{1}_{[0,1]}(u)$ is the indicator function to ensure that the PDF is zero for values outside the defined range $u \in [0, 1]$. There is no closed-form expression to compute the inverse Q^{-1}, i.e., the Cumulative Distribution Function (CDF), analytically, and therefore (2.69) needs to be solved numerically [110].

Figure 2.4 shows different PDFs which can be produced by the GLD. For $\lambda_3 = \lambda_4$ the distribution is symmetric, and $\lambda_3, \lambda_4 < 1$ leads to bell-shaped distributions. A normal distribution, for example, can be approximated by setting $\lambda_3 = \lambda_4 = 0.13$. When $\lambda_3 \leq -\frac{1}{k}$ or $\lambda_4 \leq -\frac{1}{k}$ moments of order greater than k do not exist, i.e., mean and variance only exist for $\lambda_3, \lambda_4 > -0.5$. In this case, the mean is given by

$$\mathbb{E}[Y] = \lambda_1 - \frac{1}{\lambda_2}\left(\frac{1}{\lambda_3 + 1} - \frac{1}{\lambda_4 + 1}\right) \tag{2.71}$$

where the variance is determined by

$$\mathrm{Var}(Y) = \frac{(d_2 - d_1^2)}{\lambda_2^2} \tag{2.72}$$

with the beta function B and

$$\begin{aligned} d_1 &= \frac{1}{\lambda_3}\mathrm{B}(\lambda_3 + 1, 1) - \frac{1}{\lambda_4}\mathrm{B}(1, \lambda_4 + 1), \\ d_2 &= \frac{1}{\lambda_3^2}\mathrm{B}(2\lambda_3 + 1, 1) - \frac{2}{\lambda_3\lambda_4}\mathrm{B}(\lambda_3 + 1, \lambda_4 + 1) + \frac{1}{\lambda_4^2}\mathrm{B}(1, 2\lambda_4 + 1). \end{aligned} \tag{2.73}$$

Generalized lambda models

To derive a surrogate for the stochastic simulator $\mathcal{M}$, the GLaM [110] aims to approximate $Y(\boldsymbol{x})$ by a GLD which is a function of $\boldsymbol{x}$. This means the GLaM needs to find an optimal parameter combination $\boldsymbol{\lambda}(\boldsymbol{x}) = (\lambda_1(\boldsymbol{x}), ..., \lambda_4(\boldsymbol{x}))^T$ as a function of $\boldsymbol{x}$ such that the GLD $p_Y^{\mathrm{GLD}}(y; \boldsymbol{\lambda}(\boldsymbol{x}))$ can well approximate the underlying distribution $p_Y(y|\boldsymbol{x})$. The GLaM uses PCE to model the dependence of the GLD parameters $\boldsymbol{\lambda}(\boldsymbol{x})$ on the model variables $\boldsymbol{x}$ and such allows to approximate the conditional distribution of the QoI

$$Y(\boldsymbol{x}) \sim p_Y^{\mathrm{GLD}}(y; \lambda_1^{\mathrm{PC}}(\boldsymbol{x}), \lambda_2^{\mathrm{PC}}(\boldsymbol{x}), \lambda_3^{\mathrm{PC}}(\boldsymbol{x}), \lambda_4^{\mathrm{PC}}(\boldsymbol{x})). \tag{2.74}$$

The spectral approximations of the lambda parameters are given by

$$\begin{aligned} \lambda_l(\boldsymbol{x}) &\approx \lambda_l^{\mathrm{PC}}(\boldsymbol{x}; \boldsymbol{c}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_l} c_{l,\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \quad l = 1, 3, 4, \\ \lambda_2(\boldsymbol{x}) &\approx \lambda_2^{\mathrm{PC}}(\boldsymbol{x}; \boldsymbol{c}) = \exp\left(\sum_{\boldsymbol{\alpha} \in \mathcal{A}_2} c_{2,\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right), \end{aligned} \tag{2.75}$$

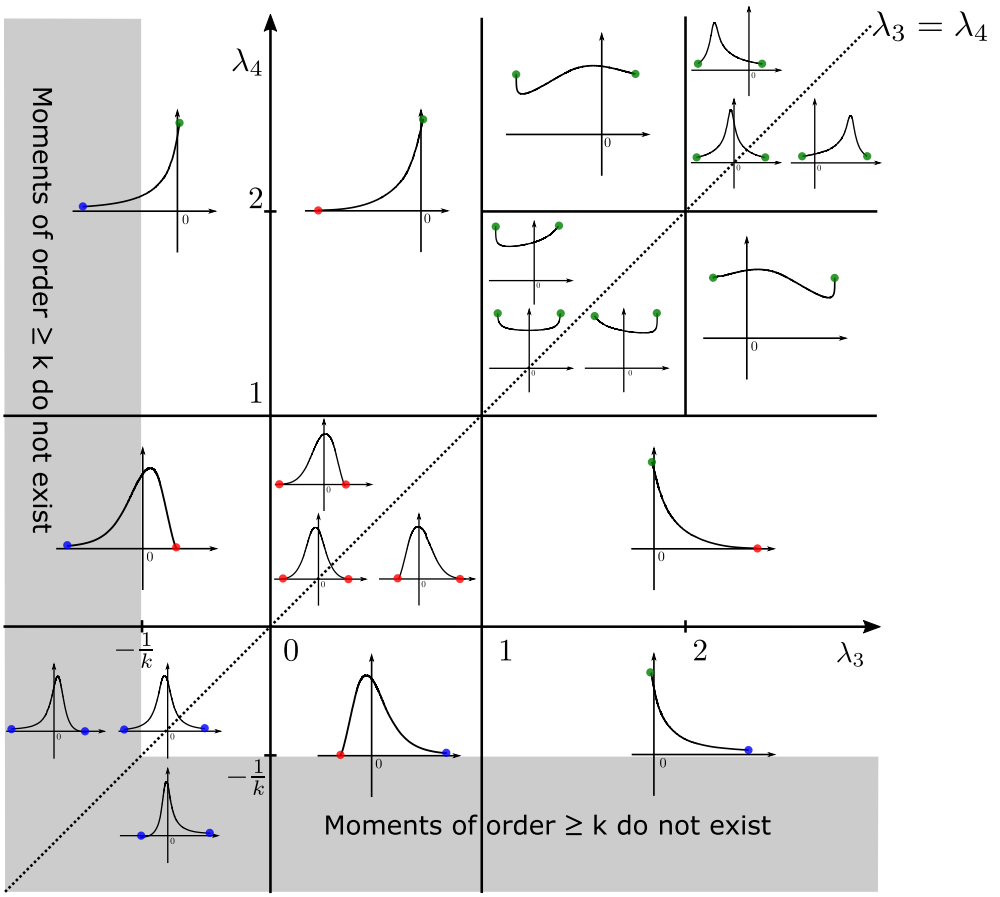


Figure 2.4: Influence of λ_3 and λ_4 on the shape of the GLD. The location is set to $\lambda_1 = 0$ with the scale being $\lambda_2 = 1$. The blue dots indicate that the PDF has unlimited support in the marked direction. Green and red dots denote a bounded support of the PDF with $p_Y(y) = 0$ on the boundary for red dots and $p_Y(y) = 1$ for green dots. [110]

where $\lambda_2(\boldsymbol{x})$ is transformed to $\log(\lambda_2(\boldsymbol{x}))$ to ensure positivity. The truncation sets $\boldsymbol{\mathcal{A}} = \{\mathcal{A}_l : l = 1, ..., 4\}$ and coefficients $\boldsymbol{c} = \{c_{l,\boldsymbol{\alpha}} : l = 1, ..., 4, \boldsymbol{\alpha} \in \mathcal{A}_l\}$ are defined for all $\lambda_l \in \{\lambda_l : l = 1, ..., 4\}$.

A direct approach [109] to estimate the coefficients $\boldsymbol{c}$ of the GLaM is to draw multiple samples from the model distribution $Y \sim p_Y(y|\boldsymbol{x}_i)$ for multiple samples $\boldsymbol{X}_1, ..., \boldsymbol{X}_M$ drawn from $\boldsymbol{X}_i \sim p_X(\boldsymbol{x})$. This allows to fit the GLD to each set of samples $Y_1^{(i)}, ..., Y_R^{(i)} \sim p_Y(y|\boldsymbol{x}_i)$ and estimate the corresponding set of parameters $\{\hat{\lambda}_l(\boldsymbol{x}_i) : l = 1, ..., 4\}$. With the resulting input $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_M\}$ and output sets $\{\hat{\lambda}_l(x_1)..., \hat{\lambda}_l(x_M) : l = 1, ..., 4\}$ the PCE approximations $\{\lambda_l^{\mathrm{PC}}(\boldsymbol{x}; \boldsymbol{c}) : l = 1, ..., 4\}$ can be build.

The described coefficient estimation based on replications for each parameter sample typically requires a large number of samples as for each sample $\boldsymbol{X}_i$ a sufficient amount of replications R is necessary to get stable estimates for the lambda parameters, requiring in total $R \cdot M$ samples. Therefore, the authors in [110] derive a joint fitting estimator that doesn't require replications, i.e., for each $\boldsymbol{X}_i$ only one sample is drawn from $Y_i \sim p_Y(y|\boldsymbol{x}_i)$. The samples Y_i can be considered as draws from the joint distribution $Y_i \sim p_{\boldsymbol{X},Y}(\boldsymbol{x}, y)$ of $\boldsymbol{X}$ and Y. This allows one to derive the joint fitting estimator given by the conditional MLE

$$\begin{aligned} \hat{\boldsymbol{c}} &= \arg\max_{\boldsymbol{c} \in C} \ell(\boldsymbol{c}), \\ \ell(\boldsymbol{c}) &= \sum_{i=1}^{M} \log\left(p_Y^{\mathrm{GLD}}\left(y_i; \boldsymbol{\lambda}^{\mathrm{PC}}(\boldsymbol{x}_i; \boldsymbol{c})\right)\right), \end{aligned} \tag{2.76}$$

which minimizes the Kullback-Leibler divergence between the joint PDF of the GLD $p_{\boldsymbol{X},Y}^{\mathrm{GLD}}(\boldsymbol{x}, y)$ and the true underlying PDF $p_{\boldsymbol{X},Y}(\boldsymbol{x}, y)$ by modifying the PCE coefficients $\boldsymbol{c}$. For more details on the properties of this estimator, see [109, 110]. They also describe a stepwise fitting procedure which first identifies a suitable starting point for $\boldsymbol{c}_0$ and the truncation set $\mathcal{A}$ with Feasible generalized least-squares (FGLS) [105] and then searches for the MLE optimum $\hat{\boldsymbol{c}}$ with the trust-region optimization algorithm [95].

2.3.5 Bayesian inference

Bayesian inference combines measurements with prior knowledge to identify the parameters of a model using Bayes' theorem. It is a suitable method for inverse UQ-problems as it can incorporate the uncertainty of the parameter identification result, taking into account prior knowledge of the parameters and the available data, see Figure 2.5. For an introduction to Bayesian methods and their applications, see also [33, 66].

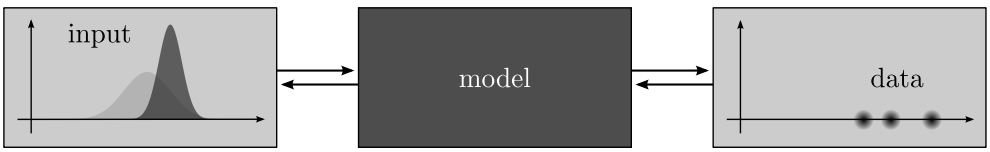


Figure 2.5: Scheme of inverse UQ [73]

The basic idea of Bayes' theorem is to assign probabilities to the parameters of a model, i.e., the inputs $\boldsymbol{x}$ are modeled as random variables which are drawn from the prior distribution $\boldsymbol{X} \sim p_{\boldsymbol{X}}(\boldsymbol{x})$. This distribution can incorporate prior knowledge

about the input parameters' possible values. The random input $\boldsymbol{X}$ is combined with the model $\mathcal{M}$, in our case a stochastic simulator, to represent the random output $Y = \mathcal{M}(\boldsymbol{x})$. The following considerations can be directly extended to multiple outputs by considering their joint distribution. Observations are assumed to be drawn from the output distribution $Y_i \sim p_Y(y)$. The joint distribution of $\boldsymbol{X}$ for N measurements $\boldsymbol{Y} = (Y_1, ..., Y_N)^T$ is given by

$$p_{X,Y}(\boldsymbol{x}, \boldsymbol{y}) = p_X(\boldsymbol{x}|\boldsymbol{y})p_Y(\boldsymbol{y}) = p_Y(\boldsymbol{y}|\boldsymbol{x})p_X(\boldsymbol{x}) \tag{2.77}$$

where the likelihood $\mathcal{L}_y(\boldsymbol{x}) = p_Y(\boldsymbol{y}|\boldsymbol{x})$ expresses the probability of obtaining a set of measurements $\boldsymbol{y}$ given the parameter vector $\boldsymbol{x}$. The conditional distribution $p_X(\boldsymbol{x}|\boldsymbol{y})$ is called the posterior distribution which describes the PDF of the parameters in $\boldsymbol{X}$ conditional on the measurements $\boldsymbol{y} = (y_1, ..., y_N)^T$. Reformulating (2.77) leads to Bayes' theorem

$$p_X(\boldsymbol{x}|\boldsymbol{y}) = \frac{p_Y(\boldsymbol{y}|\boldsymbol{x})p_X(\boldsymbol{x})}{p_Y(\boldsymbol{y})} \tag{2.78}$$

which relates the posterior $p_X(\boldsymbol{x}|\boldsymbol{y})$ to the likelihood $p_Y(\boldsymbol{y}|\boldsymbol{x})$, the prior $p_X(\boldsymbol{x})$, and the marginal distribution $p_Y(\boldsymbol{y})$. The marginal distribution in the denominator is a normalization constant to ensure that the posterior integrates to 1. It is given by

$$p_Y(\boldsymbol{y}) = \int_{\mathcal{X}} p_{X,Y}(\boldsymbol{x}, \boldsymbol{y})d\boldsymbol{x} = \int_{\mathcal{X}} p_Y(\boldsymbol{y}|\boldsymbol{x})p_X(\boldsymbol{x})d\boldsymbol{x} \tag{2.79}$$

and is not a function of the model parameters $\boldsymbol{X}$. The posterior, therefore, is proportional to its numerator

$$p_X(\boldsymbol{x}|\boldsymbol{y}) \propto p_Y(\boldsymbol{y}|\boldsymbol{x})p_X(\boldsymbol{x}). \tag{2.80}$$

To use Bayesian inference, the first step is to select a suitable prior distribution for the input vector $\boldsymbol{X}$. The prior distribution reflects the subjective belief about the probability of certain parameter values. One can also define an uninformative prior by assigning a uniform distribution that only conveys information about the bounds of the parameter, i.e., some physical or model-related constraints to the parameter, and does not favor any value within the bounds. Next, one must define a likelihood function to connect measurements to the model. To reduce the computational effort for evaluating the likelihood, the model is often replaced with a surrogate $\hat{\mathcal{M}} \approx \mathcal{M}$ leading to the likelihood $\hat{\mathcal{L}}_y(\boldsymbol{x}) = \hat{p}_Y(\boldsymbol{y}|\boldsymbol{x})$ for the surrogate $\hat{\mathcal{M}}$. The last step is to sample from the posterior distribution, which is typically done with MCMC methods such as described in Section 2.3.1. Other methods to compute the posterior distribution

include variational inference [88] and Laplace approximations [53], which both use distributional assumptions for the posterior that are not required by the sample-based MCMC methods.

When the uncertainty of input parameters is identified with the posterior distribution $p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{y})$, the effects of it on the model outputs $\boldsymbol{Y}$ can be investigated. A point estimate of the parameter vector $\hat{\boldsymbol{x}}$ is given by the Maximum a posteriori (MAP) estimator

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{y}) \tag{2.81}$$

or the expectation $\hat{\boldsymbol{x}} = \mathbb{E}[\boldsymbol{X}|\boldsymbol{y}]$. Plugging the point estimate into $p_Y(y'|\hat{\boldsymbol{x}})$ gives the distribution of future outcomes y' considering the estimate $\hat{\boldsymbol{x}}$ derived from the posterior distribution. While this approach includes the updated information on the parameters $\boldsymbol{x}$ to some degree, the uncertainty about the parameter estimate is lost through the point estimator. This uncertainty is captured by the posterior predictive distribution, which combines the posterior distribution with the simulation model to assess the uncertainty of the outputs given the identified input uncertainties. The posterior predictive distribution of future outcomes y' is then given by

$$p_Y(y'|\boldsymbol{y}) = \int_{\mathcal{X}} p_Y(y'|\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x} \tag{2.82}$$

which combines the conditional distribution of the simulation model $p_Y(y'|\boldsymbol{x})$ with the uncertainty in the parameters $p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{y})$. While using the posterior predictive distribution provides the advantage of considering the uncertainty about $\boldsymbol{x}$, it has a higher computational cost compared to using the point estimate $\hat{\boldsymbol{x}}$ as the integral in (2.82) must be evaluated.

3 Stochastic velocity simulation

To design reliable automotive components, it is important to understand the expected operational loads. Early in development, virtual load generation helps derive these loads realistically. Many vehicle component loads are based on the driven velocity, which makes velocity simulation an essential component for virtual load generation. The velocity simulation generates velocity profiles on arbitrary routes as a function of driver behavior and vehicle properties, which are then processed with further simulations to derive component loads. The velocity simulation can substitute for physical test drives and enables the completion of virtual test drives for thousands of kilometers within seconds, saving time and resources in an early development stage. To do so, a set of routes is derived for multiple drivers and their vehicles, which should optimally represent the routes driven during the lifetime of the vehicle. Based on these routes, the velocity simulation generates velocity profiles representative of the usage of the vehicle over its lifetime, which then, besides the routes, also depends on the driving behavior of the driver and the properties of the vehicle.

Velocity profiles are random, i.e., for a given route, driver, and vehicle, the resulting velocity is not deterministic, but instead is given by a stochastic process $\{V_s : s \in S\}$ with respect to position s along the route. In the following, the stochastic process is divided into a piecewise constant part $m_V(s)$, which models changes in the mean velocity, and a stochastic part X_s to model the fluctuations around the mean velocity. The stochastic process for the velocity is then given by

$$\{V_s\} = \{m_V(s)X_s\} = m_V(s)\{X_s\} \tag{3.1}$$

which allows for splitting the modeling of velocities in the simulation into two parts. The piecewise constant part is influenced by, e.g., speed limits, traffic jams, stops, or curves. The stochastic part, on the other hand, models random fluctuations that are caused by driver inattention and the actions of nearby pedestrians, cyclists, and vehicles. In the further course, the piecewise constant part $m_V(s)$ is denoted as the route velocity $v_{route}(s)$ as it primarily depends on the route properties. Some component loads are

primarily dependent on the mean velocity, as they are not sensitive to short-term fluctuations, e.g., road-induced vibrations do generally not depend on sudden changes in velocity. Other loads, especially on the drivetrain components of a vehicle, such as shafts, gears, bearings, etc., are highly dependent on the random velocity fluctuations, as the fluctuations cause additional acceleration and deceleration cycles, which lead to spikes in the forces and torques acting on the drivetrain components.

The chapter is structured as follows. In Section 3.1, the existing velocity simulation model is introduced. In Section 3.2, a new approach to model the velocity fluctuations X_s based on the stochastic process in (3.1) is developed. The following Sections concern changes to the mean velocity $m_V(s)$. In Section 3.3, the influence of traffic on the mean velocity is described by analyzing FCD and deriving a statistical model for traffic jams from it. Finally, in Section 3.4, stops are analyzed to derive stop probabilities and durations for various stop reasons, such as traffic lights and traffic jams.

3.1 Velocity simulation model

In order to model velocities using (3.1), we follow the approaches described in [85, 93], where the authors derive a simulation model that uses descriptions of the driver, vehicle, and road to generate realistic velocity profiles for a route. The following Sections describe the general simulation procedure and the most important steps involved.

3.1.1 Simulation procedure

The velocity simulation follows a defined procedure that, considering the driver and vehicle properties, step-by-step generates velocity profiles on a route. The route is described by several properties along its path: the road geometry, including slope $\alpha(s)$ and curvature $\kappa(s)$, the legal speed limit $v_{max,le}(s)$, the road type $r_{road}(s)$ such as city, country, or motorway, and the locations of stop events s_{Sr,r_S} with their reasons r_S like traffic lights, yield signs, stop signs, and pedestrian crossings. All these properties depend on the position s along the route. The route properties are obtained from a database with map data, which, given the route coordinates, can be queried and returns the mentioned route properties. The database is provided by map data providers such as HERE [43]. Moreover, in [85], a range of parameters is used to describe the driver and his behavior with properties such as the tolerated accelerations and the standard deviation of random velocity fluctuations, see Table A.2 for a complete list. Incorporating the vehicle dynamics model described in Section 2.1, the vehicle is described by its basic mechanical properties such as mass, power, etc., see Table A.1 for

a complete list. Figure 3.1 gives an overview of the velocity simulation procedure. The main steps during the simulation are described in the following.

1. **Pre-processing**
   The road description is given by the road properties, such as slope $\alpha(s)$ and legal speed limit $v_{max,le}(s)$, obtained from a map database at predefined reference points. For an arbitrary point, the road properties can be interpolated with respect to the reference points using linear interpolation. As a preparation for subsequent simulation steps, see Section 3.1.2, the route is split into segments indexed by u where splitting occurs when the road type changes or the maximum length of a segment is exceeded, see Figure 3.2a.

2. **Generate goal velocity profile**
   The idea is to generate a goal velocity series that the driver tries to follow, irrespective of limitations in longitudinal dynamics. The goal velocity

   $$v_{goal}(s) = v_{route}(s)X_s \tag{3.2}$$

   is assembled from the piecewise constant route velocity $v_{route}(s)$ and a stochastic component X_s, where the route velocity is the minimum of the legal speed limit $v_{max,le}(s)$, traffic speed limit $v_{max,tr}(s)$ (see Section 3.3), maximum driver speed $v_{max,dr}(s)$ and the curvature velocity limit $v_{max,\kappa}(s)$

   $$v_{route}(s) = \min(v_{max,le}(s), v_{max,tr}(s), v_{max,dr}(s), v_{max,\kappa}(s)). \tag{3.3}$$

   See Section 3.2 for a description of the stochastic component X_s.

3. **Stop events**
   The positions of stop events s_{Sr,r_S} with reason r_S are taken from a map database. Stop events are assigned a probability for a stop $p_S(s|r_S)$ as well as a stop duration distribution $p_{T_S}(\tau_S|r_S)$ conditional on the reason r_S of the stop event. When a stop event happens, the route velocity $v_{route}(s)$ is set to zero at the position of the stop, and the stop duration is drawn from $p_{T_S}(\tau_S|r_S)$.

4. **Longitudinal dynamics**
   The longitudinal dynamics are limited by the traction F_{tr} and resistance F_{res} forces of the vehicle, based on the model described in Section 2.1, and the tolerated accelerations $a_{x,max,dr}$ and decelerations $a_{x,min,dr}$ of the driver.

In Section 3.1.2, a detailed description of the process to generate the route velocity $v_{route}(s)$ is given. Section 3.1.3 gives an introduction to available solutions for the

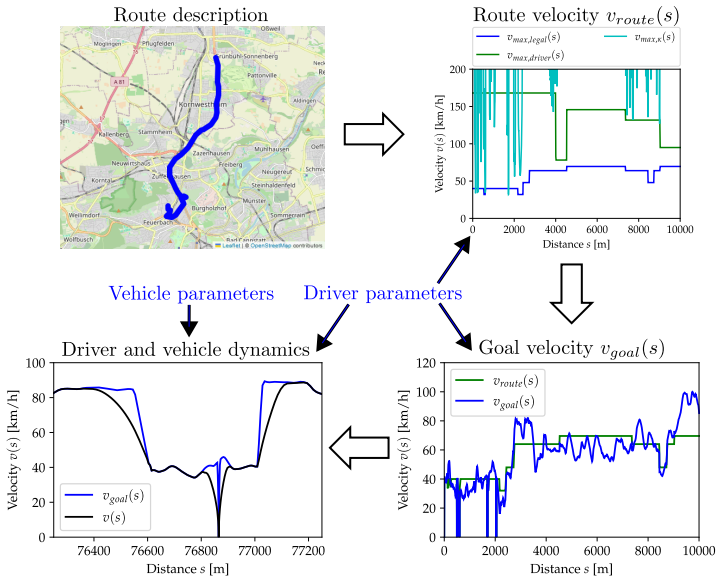


Figure 3.1: Overview velocity simulation model

generation of velocity fluctuations. The last Section 3.1.4 explains how the longitudinal dynamics are included in the velocity simulation model.

3.1.2 Route velocity

The route velocity is derived from the properties of the road. As described in (3.3), it is the minimum of multiple other quantities, each characterizing another limit to the velocity. The individual limits are:

1. **Legal speed limit** $v_{max,le}(s)$
   The legal speed limit is given by the product of the speed limit $v_{lim}(s)$ taken from map data, the road type, and driver-dependent speeding factor $o_{v,lim}(r_{road}(s))$ which describes how closely the driver, on average, follows the speed limit and the slope factor $o_{v,\alpha}(s)$

$$v_{max,le}(s) = v_{lim}(s) o_{v,lim}(r_{road}(s)) o_{v,\alpha}(s). \tag{3.4}$$

   The slope factor adapts the speed to changes in the road slope $\alpha(s)$, which generally means the speed is decreased on uphill sections and increased on

downhill sections. The slope factor $o_{v,\alpha}(s)$ is given as a linear function of the road slope

$$o_{v,\alpha}(s) = \begin{cases} o_{v,\alpha,min} & \text{if } m_\alpha \alpha(s) < o_{v,\alpha,min} \\ m_\alpha \alpha(s) & \text{if } o_{v,\alpha,min} \leq m_\alpha \alpha(s) \leq o_{v,\alpha,max} \\ o_{v,\alpha,max} & \text{if } m_\alpha \alpha(s) > o_{v,\alpha,max} \end{cases} \tag{3.5}$$

where the slope adaption coefficient m_α controls the sensitivity to changes in the road slope, while the lower $o_{v,\alpha,min}$ and upper limit $o_{v,\alpha,max}$ control the limits of the slope factor for steep road sections. The parameters m_α, $o_{v,\alpha,min}$, and $o_{v,\alpha,max}$ are set to default values independent of the driver.

2. **Traffic speed limit** $v_{max,tr}(s)$
   The traffic speed limit models random limitations of the velocity due to high traffic densities, which are not included in the available models [85, 93]. The newly developed approach is described in Section 3.3.

3. **Maximum driver speed** $v_{max,dr}(s)$
   Describes how fast, on average, the driver wants to drive on each road segment u if there were no other limitations. This limit is especially relevant for unrestricted segments on German motorways. For each segment, it is drawn from the truncated normal distribution

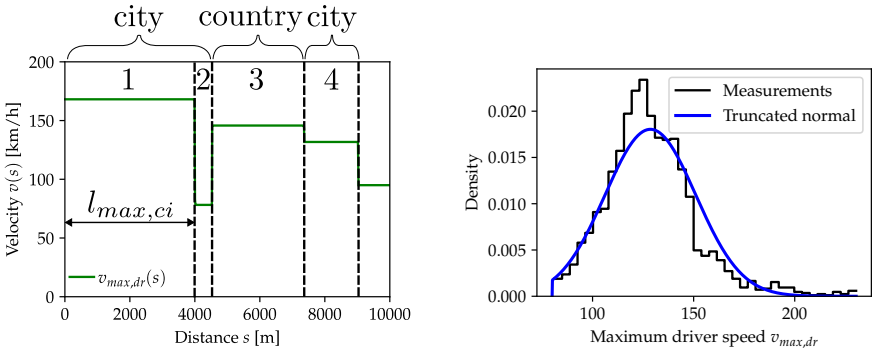
$$v_{max,dr,u} \sim p_{V_{max,dr}}(v_{max,dr}; \mu_{v_{max,dr}}, \sigma_{v_{max,dr}}, a_{v_{max,dr}}, b_{v_{max,dr}}) \tag{3.6}$$

   where $\mu_{v_{max,dr}}$, $\sigma_{v_{max,dr}}$ and $b_{v_{max,dr}}$ are the mean, standard deviation and maximum of the maximum driver speed, all three quantities are derived from measurements of the mean velocity in unrestricted motorway segments of the respective driver. The minima $a_{v_{max,dr}}$ is set to a fixed value of 80 km/h for cars and 60 km/h for trucks to prevent smaller maximum driver speeds, which are likely caused by dense traffic and do not reflect the drivers request. The difference of $a_{v_{max,dr}}$ between cars and trucks is due to the generally lower speeds of trucks. Figure 3.2b compares the measured maximum driver speed distribution of one driver with the truncated normal distribution approximation. There are distinct deviations between the measured distribution and the approximation, primarily a density accumulation between 120 km/h and 130 km/h in the measurement. This accumulation is caused by the presence of legal speed limits in some segments in reality, which are marked as unrestricted in the map data, leading to the mistaken inclusion of the corresponding road segments in the analysis. Since the legal speed limit $v_{max,le}$ is separately handled in the derivation of the route velocity, see (3.3), this deviation

in the distribution approximation is not significant for modeling the maximum driver speed. Figure 3.2a depicts how the maximum driver speed $v_{max,dr}$ changes between road segments when drawing from (3.6). As the maximum driver speed is not specific to the road type, it can take values significantly above the respective speed limits, e.g., on city roads, which limits its influence mainly to motorways.

4. **Curvature velocity limit** $v_{max,\kappa}(s)$
   The tolerated lateral accelerations of the driver $a_{y,max,dr}$ and the curvature of the road $\kappa(s)$ define the curvature velocity limit

$$v_{max,\kappa}(s) = \sqrt{\frac{a_{y,max,dr}}{|\kappa(s)|}}. \tag{3.7}$$



(a) Segmentation of route into segments. Splitting is done when the road type changes or the maximum segment length, in this case $l_{max,ci}$ for city roads, is exceeded.

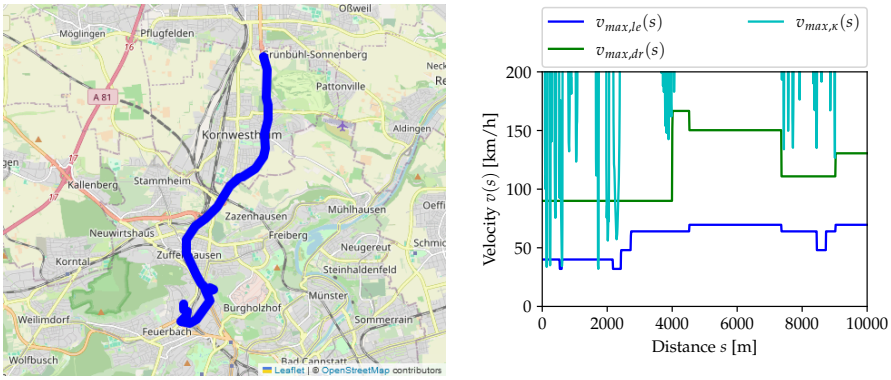
(b) Distribution of maximum driver speed measurements for one driver, compared with the truncated normal distribution approximation

Figure 3.2: Route segmentation and derivation of maximum driver speed distribution

In Figure 3.3, an example of the components of the route velocity $v_{route}(s)$ is given. The curvature limit $v_{max,\kappa}(s)$ drops sharply at road bends, while the maximum driver speed $v_{max,dr}(s)$ changes randomly at every change in road type and after a certain distance. The legal speed limit $v_{max,le}(s)$ mainly follows the speed limit of the road while taking into account the speeding $o_{v,lim}(r_{road}(s))$ and slope $o_{v,\alpha}(s)$ factors.

3.1.3 Velocity fluctuations

Velocity fluctuations are a central component of the velocity simulation as they are one of the main sources of randomness in the simulation. The goal is to simulate the random fluctuations in velocity caused by both internal and external factors, including driver



(a) Map of the 10 km route driven northbound

(b) Legal speed limit $v_{max,le}(s)$, maximum driver speed $v_{max,dr}(s)$ and curvature velocity limit $v_{max,\kappa}(s)$ on the route

Figure 3.3: Components of the route velocity $v_{route}(s)$ on an example route

inattention and the actions of nearby pedestrians, cyclists, and vehicles, e.g, [58]. These influences contribute to the stochastic nature of the velocity. The velocity fluctuations X_s are modeled by a stochastic process $\{X_s : s \in S\}$ for the distance s along a route. In [85], an approach for a discrete stochastic process to model velocity fluctuations is given. The process is a discrete time series $\{X_{t_i}\}_{i=1}^{N}$ of random variables X_{t_i} with constant step width $\Delta t = t_i - t_{i-1} = const$. The discrete time series can be transformed to a discrete distance series by taking into account the goal velocity $v_{goal}(s) = v_{route}(s)X_s$ and summing up over all previous time steps to calculate the distance series

$$s_i = s_1 + \sum_{j=1}^{i-1} v_{goal}(s_j)\Delta t = s_1 + \sum_{j=1}^{i-1} v_{route}(s_j)X_{t_j}\Delta t, \tag{3.8}$$

which allows to express the velocity fluctuations $X_{s_i} = X_{t_i}$ as a function of the distance steps s_i with non constant step width. Values of X_s, or $v_{goal}(s)$ respectively, between the discrete distance steps s_i can be obtained by linear interpolation.

The logarithm of the process $\{X_{t_i}\}_{i=1}^{N}$ is an AR(1) process

$$\begin{aligned} X_{t_i} &= \exp\left(Z_{t_i} - \frac{\omega^2}{2}\right), \\ Z_{t_i} &= \eta Z_{t_{i-1}} + \epsilon_{t_i}, \text{ with } \epsilon_{t_i} = \mathcal{N}(0, \sigma_\epsilon), \end{aligned} \tag{3.9}$$

with the mean correction $\omega = \frac{\sigma_\epsilon}{\sqrt{1-\eta^2}}$ to ensure that $\mathbb{E}[X_{t_i}] = 1$. This gives two parameters to influence the properties of the velocity fluctuations, the AR parameter η to set the correlation of the velocity fluctuations and σ_ϵ to set the variance of the noise. Figure 3.4 gives an example of the goal velocity $v_{goal}(s)$ composed of route velocity and velocity fluctuations. At the beginning of the route, some stop events are also added to the goal velocity, i.e., the velocity is set to zero at the position of the stop.

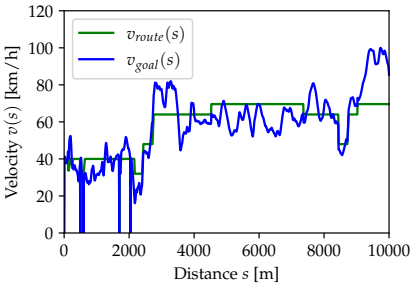


Figure 3.4: Example of goal velocity $v_{goal}(s) = v_{route}(s)X_s$

3.1.4 Longitudinal dynamics

The velocity simulation dynamics module takes the longitudinal dynamics of the driver and the vehicle into account, where the longitudinal vehicle dynamics are based on the model described in Section 2.1. The longitudinal dynamics are simplified to only take the main factors into account, which is sufficient to model them in the context of velocity simulation [93]. The model calculates the acceleration of the vehicle taking into account the goal velocity $v_{goal}(s)$, the tolerated acceleration of the driver $a_{x,max,dr}$, and the maximum acceleration possible with the given vehicle $a_{x,max,veh}(s,v)$. The acceleration

$$a_{x,max,veh}(s,v) = \frac{F_{tr} - F_{res}(s,v)}{m_{veh}}. \tag{3.10}$$

is derived from the resistance forces $F_{res}(s,v)$ and the available traction force F_{tr}, which is a function of engine torque, power, and road friction, see Section 2.1 for more details.

The maximum acceleration tolerated by the driver is drawn for each road segment u using a normal distribution

$$a_{x,max,dr}(s) \sim \mathcal{N}(\mu_{a_{x,max,dr}}, \sigma_{a_{x,max,dr}}) \tag{3.11}$$

where the mean and variance are estimated from various acceleration events of the driver. To prevent a binary driving style with instant acceleration changes while driving, the maximum driver acceleration is scaled depending on the difference $\Delta v(s)$ between the current velocity $v(s)$ and the goal velocity $v_{goal}(s)$, e.g., when the driver wants to accelerate from a standstill to the goal velocity, the acceleration decreases progressively to ensure a smooth transition towards the goal velocity. The scaled maximum driver acceleration is given by

$$a'_{x,max,dr}(\Delta v(s)) = o_{a,dr}(\Delta v(s)) a_{x,max,dr}(s) \tag{3.12}$$

where the scaling factor $o_{a,dr}(\Delta v(s))$ is defined by the bounded logarithmic function

$$o_{a,dr}(\Delta v(s)) = \begin{cases} o_{a,dr,min} & \text{if } o_{a,dr} < o_{a,dr,min} \\ \alpha_y \log\left(\max \frac{|\Delta v(s)|}{\alpha_x} - \beta_x\right) + \beta_y & \text{if } o_{a,dr,min} \leq o_{a,dr} \leq 1 \\ 1 & \text{if } o_{a,dr} > 1 \end{cases} \tag{3.13}$$

with the minimum factor $o_{a,dr,min}$ and the scale α_x, α_y and shift β_x, β_y parameters, which are chosen to achieve the desired acceleration behavior on city, country, and motorway roads, see Figure 3.5.

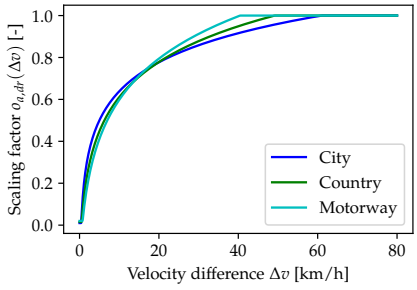
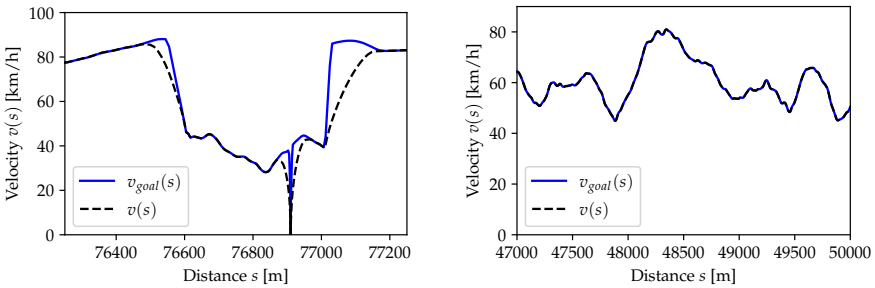


Figure 3.5: Driver acceleration scaling factor $o_{a,dr}(\Delta v)$ for city, country, and motorway roads

Algorithm 1 Iterative simulation of longitudinal dynamics

1: $v(s_1) \leftarrow v_{goal}(s_1)$
2: **for** $i \leftarrow 2$ to N **do**
3: $\quad \Delta s \leftarrow s_{i+1} - s_i$
4: $\quad \Delta v \leftarrow v_{goal}(s_i) - v(s_{i-1})$
5: $\quad \Delta v_{possible} \leftarrow \frac{\Delta s}{v(s_{i-1})} \min(a_{x,max,veh}(s_{i-1}, v(s_{i-1})), a'_{x,max,dr}(\Delta v))$
6: $\quad v(s_i) \leftarrow \min(v_{goal}(s_i), v(s_{i-1}) + \Delta v_{possible})$
7: **end for**

The velocity series $v(s)$ at the discrete positions $s_1, ..., s_N$ along a route is determined from the goal velocity $v_{goal}(s)$ in (3.2) and the longitudinal dynamics iteratively with Algorithm 1. Velocities between the discrete positions are then obtained by linear interpolation. The same algorithm is applied in the reverse direction, starting from the last distance step s_N to consider the longitudinal dynamics for deceleration with the maximum tolerated driver deceleration $a_{x,min,dr}$ and the maximum deceleration possible with the given vehicle $a_{x,min,veh}$. Figure 3.6 shows how the vehicle and driver dynamics influence the simulated velocity series $v(s)$. In Figure 3.6a, a road section with stop events and abrupt changes in the goal velocity is depicted, the velocity series $v(s)$ is therefore strongly influenced by the longitudinal dynamics. Figure 3.6b depicts a road section where the goal velocity is altered by the velocity fluctuations without any abrupt changes. In this case, the dynamics do not prevent the velocity series from following the goal velocity, and therefore $v(s) = v_{goal}(s)$ follows.



(a) Road section with stop events and abrupt changes

(b) Road section without stop events and abrupt changes

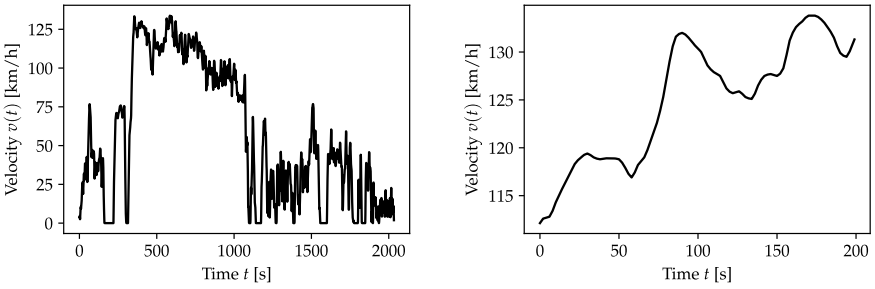
Figure 3.6: Simulated velocity series $v(s)$ and goal velocity $v_{goal}(s)$

3.2 Modeling of velocity fluctuations

As described in Section 3.1, there is a stochastic component within the velocity simulation to model random fluctuations of the velocity profiles. These fluctuations result from internal and external influences, such as the driver's inattentiveness and the behavior of surrounding pedestrians, bikes, cars, etc., which all cause the velocity to contain a stochastic component [58]. The stochastic component is given by the discrete stochastic process $\{X_{t_i}\}_{i=1}^{N}$. The approach given in [85] uses a logarithmic AR(1) process to describe X_{t_i}, see (3.9). Although this approach allows modeling of the general characteristics of velocity fluctuations on a wide range of driver, car, and route combinations, it shows some deficiencies. These are addressed in the following

Sections to build a more general model of human driving behavior. The remarks of these Sections are already published in a compact form in [113].

3.2.1 Data preparation



(a) Velocity profile of a trip

(b) Velocity profile of a window

Figure 3.7: Example of velocity profiles

Following [85], we aim to model the velocity fluctuations by a stochastic process $\{X_{t_i}\}_{i=1}^{N}$ which is piecewise stationary, meaning that for a fixed time interval the criterion for stationarity is fulfilled. To identify the model and its parameters for one driver, a dataset with approximately 5000 km of velocity profiles from one driver is analyzed. However, on long time intervals, the measured velocity profiles, see Figure 3.7a, show instationary behaviors, such as changes in mean, variance, and correlation, caused by acceleration spikes or stop events. To handle this, the profiles are split into short windows to approximately achieve stationarity, see Figure 3.7b, whereby windows with non-stationary characteristics are filtered out. The procedure to achieve this is described in the following.

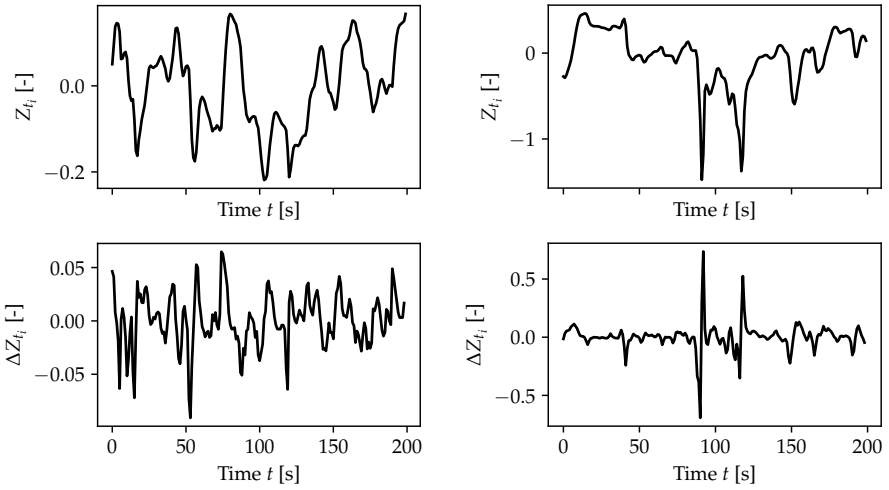
To extract the fluctuations from the measured velocities of each window, they are first discretized into steps of $\Delta t = t_i - t_{i-1} = 1\,\mathrm{s}$ and then normalized, logarithmized, and centered

$$\begin{aligned} X_{t_i} &= \frac{v_{t_i}}{\bar{v}}, \\ Z'_{t_i} &= \log\left(X_{t_i}\right), \\ Z_{t_i} &= Z_{t_i} - \bar{Z}, \end{aligned} \tag{3.14}$$

where $\bar{v}$ is the mean velocity of the window and $\bar{Z}$ the mean of the uncentered stochastic process Z'_{t_i} of the window. The length of the windows is a tradeoff between having

enough data points to accurately estimate the stochastic process parameters and the justification of the needed stationarity assumption for the application of the chosen models. For the fluctuations of a complete trip with multiple stop events, city, country, and motorway driving, etc., the stationarity assumption is clearly violated, but locally for sufficiently short windows, the assumption is justified, and the fluctuations can be approximated by stationary stochastic processes.

For further analyses the window length is set to $t_{w,l} = 200\,\mathrm{s}$, which at $\bar{v} = 50\,\mathrm{km/h}$ corresponds to a window length of $s_{w,l} = 2778\,\mathrm{m}$. The windows are shifted by $t_{w,o} = 25\,\mathrm{s}$, which allows for better capturing of short-term differences in the velocity fluctuations compared to shifting them without overlap. The resulting database includes approximately 9 000 windows to analyze for the investigated driver. The chosen window length is a compromise between the amount of data available per window and the stationarity assumption, while shortening the window shift further would increase the amount of data to analyze without significantly improving the ability to capture short-term differences.



(a) Window without outliers

(b) Window with outliers

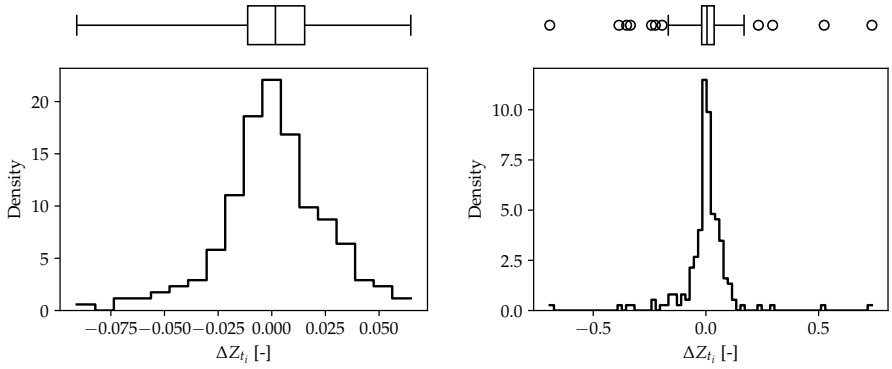
Figure 3.8: Fluctuations Z_{t_i} and their difference ΔZ_{t_i} for one window

Analyzing the fluctuation windows reveals some issues leading to non-stationarities in them. As depicted in Figure 3.8b, sharp drops in the velocity fluctuations Z_{t_i}, e.g., due to sudden braking of the vehicle, can occur and lead to instationary time series. When differentiating the series, these drops become increasingly visible. Analyzing the difference $\Delta Z_{t_i} = Z_{t_i} - Z_{t_{i-1}}$ distribution, the drops can be identified as outliers, see Figure 3.9b. Therefore, we compute the Interquartile range (IQR), and set the length of

the whiskers of the box-plot to 3 IQR(ΔZ_{t_i}) = 3($q_{0.75}(\Delta Z_{t_i}) - q_{0.25}(\Delta Z_{t_i})$). The windows in which at least one outlier

$$\Delta Z_{t_i} \notin (q_{0.25}(\Delta Z_{t_i}) - 3\ \mathrm{IQR}(\Delta Z_{t_i}), q_{0.75}(\Delta Z_{t_i}) + 3\ \mathrm{IQR}(\Delta Z_{t_i})) \tag{3.15}$$

is present are dropped from the analysis. Figures 3.8 and 3.9 give the velocity fluctuations Z_{t_i}, their difference ΔZ_{t_i}, and the distribution of the difference for a window where no outliers are present and a window with outliers. After the procedure of removing the outliers, the remaining dataset contains approximately 4 000 windows of measured velocities, which will be used in the following for the estimation of the driver-specific stochastic process parameters.



(a) Window without outliers

(b) Window with outliers

Figure 3.9: Box-plot and distribution of ΔZ_{t_i} for one window

3.2.2 Model selection

The existing velocity fluctuation model in (3.9), using an AR(1) process with white-noise, shows some deficiencies, as in general there are strong remaining correlations in the residuum ϵ_{t_i} after fitting the process. Figure 3.10 depicts the ACF of the residuum for an exemplary window, showing the remaining correlations. Therefore, based on the existing AR(1) model, other approaches to model the velocity fluctuations are investigated to identify an improved one. The new model should be able to better represent the correlation structure of human driving behavior while allowing for a flexible parametrization to differentiate between driving scenarios such as city, country, and motorway driving.

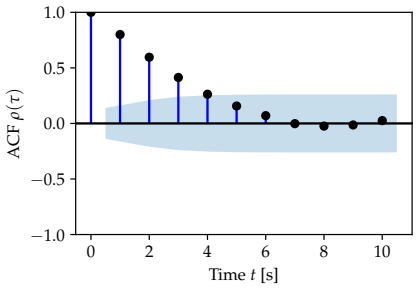


Figure 3.10: ACF $\hat{\rho}_{\epsilon_{t_i}}(\tau)$ of white-noise AR(1) residuum for an exemplary window [113]

The candidate models for the discrete process Z_{t_i} are ARMA models [71] with AR order up to $p = 3$ and MA order up to $q = 3$. Additionally, the red-noise AR(1) model [11] is considered. The BIC is used to judge the quality of the model fits, see (2.39) for its definition. The BIC is calculated for all N_w velocity windows. To assess the overall fit of a model, the mean of all BIC values from the windows is taken

$$\overline{\mathrm{BIC}} = \sum_{w}^{N_w} \mathrm{BIC}_w. \tag{3.16}$$

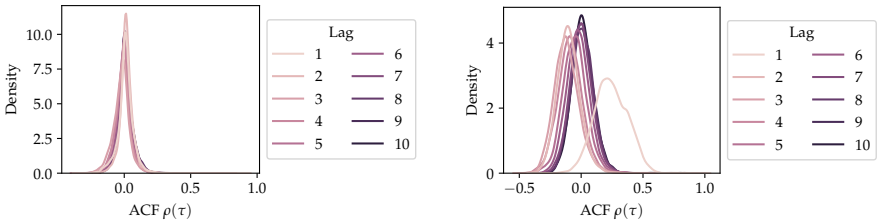
Table 3.1 gives the mean BIC values for the investigated p and q combinations as well as the red-noise AR(1) model. The models are fitted with a MLE approach as described in Section 2.2.2.

$p \downarrow q \rightarrow$	0	1	2	3	red-noise AR(1)
0	−100	−345	−551	−699	−1 014
1	−868	−995	−1 032	−1 040	
2	−1 026	−1 047	−1 045	−1 042	
3	−1 045	−1 043	−1 034	−1 033	

Table 3.1: Mean BIC values for different ARMA and red-noise AR(1) models

The best fit is achieved with an ARMA(2, 1) model. Compared to the AR(1) model described in [85], a significant improvement is achieved. The red-noise AR(1) model shows a comparable performance to that of the ARMA(2, 1) model. These two models are further investigated and compared. As another indicator of the quality of the model fit, the ACF in (2.31) of the residues is analyzed. This analysis yields an ACF for each window where lags up to $\tau = 10\,\mathrm{s}$ are considered. Figure 3.11 depicts the distribution of the estimated ACF of the residuum for each of the lags across all windows for the

ARMA(2, 1) and red-noise AR(1) models. The ARMA(2, 1) model for all lags and across all windows has fewer remaining correlations than the red-noise AR(1) model, which, especially for lag $\tau = 1$ has remaining correlations in the residuum.

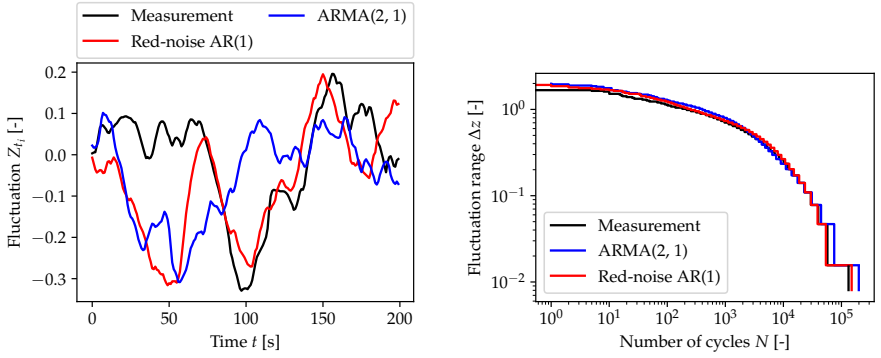


(a) ACF $\hat{\rho}_{\epsilon_{t_i}}(\tau)$ of ARMA(2, 1) residuum

(b) ACF $\hat{\rho}_{\zeta_{t_i}}(\tau)$ of red-noise AR(1) residuum

Figure 3.11: Distribution of the ACF of the residuum per window

For each window, one sample is drawn from the two stochastic processes with their estimated parameters for that window, see Figure 3.12a with an example for one window. These samples are then compared to the measured fluctuations of that window. A rainflow counting algorithm [3] is applied to each window separately. All counted cycles are added up across all windows and summarized in a histogram without mean correction, see Figure 3.12b. Both models achieve a good replication of the measured rainflow histogram.



(a) Measured values and model samples of $\{Z_{t_i}\}$ for one window

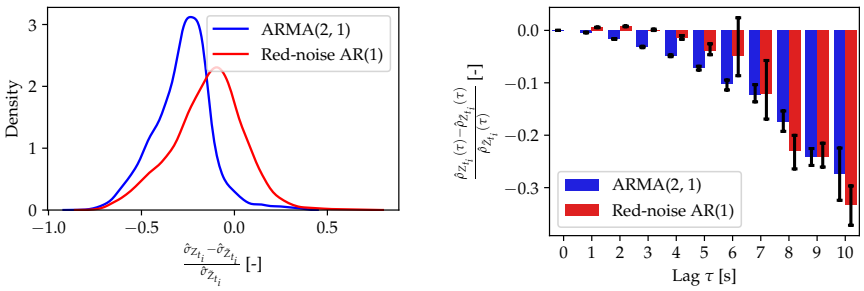
(b) Accumulated rainflow histogram of window fluctuations $\{Z_{t_i}\}$

Figure 3.12: Comparison of fluctuations between ARMA(2, 1) and red-noise AR(1) models and measurements

Further comparisons are based on the relative error

$$\epsilon_b = \frac{\hat{b}_{Z_{t_i}} - \hat{b}_{\tilde{Z}_{t_i}}}{\hat{b}_{\tilde{Z}_{t_i}}} \tag{3.17}$$

of a statistic estimated from the model draws $\hat{b}_{Z_{t_i}}$ and the measured velocity $\hat{b}_{\tilde{Z}_{t_i}}$ for each window. In Figure 3.13a, the standard deviations of the velocity fluctuations are compared. $\hat{\sigma}_{Z_{t_i}}$ is the standard deviation of the generated sample and $\hat{\sigma}_{\tilde{Z}_{t_i}}$ that of the measured velocity series in the window. On average, the red-noise AR(1) model better replicates the original standard deviation, while the error distribution of the ARMA(2, 1) model has a smaller variance. Another comparison is done based on the ACF of the stochastic process $\{Z_{t_i}\}$. Unlike in the analysis of the residuals, see Figure 3.11, correlations are expected here. The two models are compared based on the error of their ACF relative to the measurement. Figure 3.13b depicts the mean of the relative ACF error for all windows and its 95% Confidence interval (CI). Up to lag $\tau = 6\,\mathrm{s}$, there is a significantly smaller mean relative error for the red-noise AR(1) model. For high lags, there are no more significant differences.



(a) Standard deviation error

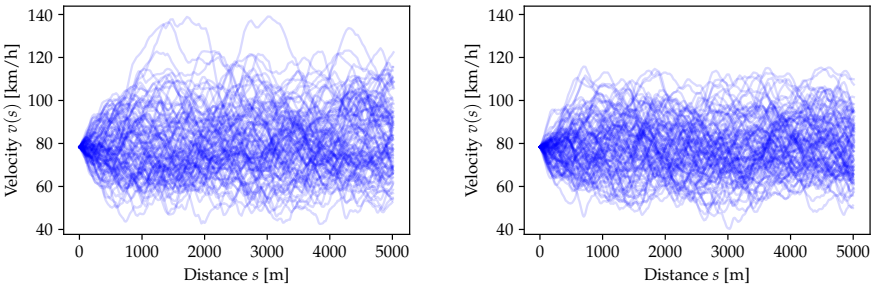
(b) ACF error

Figure 3.13: Relative error of standard deviation and ACF between generated and measured velocity fluctuations

3.2.3 Truncation of stochastic processes

ARMA or related models like the red-noise AR(1) model by default have a normal distribution, the velocity fluctuations X_{t_i}, see (3.9), thus have a log-normal distribution. The log-normal distribution is supported on $X_{t_i} \in (0, \infty)$ while also having heavy tails on the right side [29], i.e., velocities can theoretically reach any value in the range from 0 to ∞. In reality, velocities are bounded by driving behavior, road properties,

and surrounding traffic, i.e., velocities stay within certain thresholds even when the vehicle's dynamics allow higher or lower velocities. The considered stochastic processes regularly exceed these bounds, which restricts their application in velocity simulations due to the excessively strong fluctuations in the resulting velocity profiles. This issue is depicted in Figure 3.14a, which shows 100 samples of $v_{goal}(s)$ created with a red-noise AR(1) process and a route velocity of $v_{route} = 80\,\text{km/h}$. The red-noise AR(1) process is parameterized with typical values for the given route velocity. The samples frequently reach peaks of up to 140 km/h, which, when compared to measurements with similar mean velocity, is significantly out of distribution. It is thus necessary to modify the stochastic process to truncate its distribution to be consistent with velocities obtained from measurements.



(a) Not bounded **(b)** Bounded

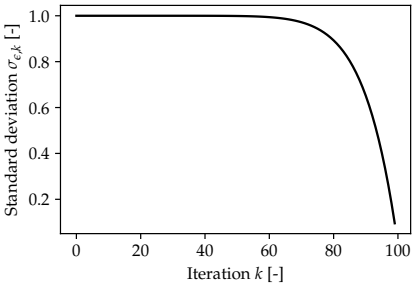
Figure 3.14: 100 goal velocity $v_{goal}(s)$ samples generated with a bounded and a not bounded red-noise AR(1) process [113]

The authors in [18] describe an approach for a bounded AR process. Their approach is based on an iterative truncation of the AR noise term ϵ_{t_i} (or ζ_{t_i} in the red-noise AR(1) case) to ensure that the process Z_{t_i} stays within its bounds $\underline{z}$ and $\overline{z}$ at each sampling step. This leads to a sharp bound on the distribution of Z_{t_i} as depicted in Figure 3.15b. In measured velocity fluctuations, no sharp lower or upper bounds are observed, the approach from [18] would therefore lead to an unrealistic accumulation of velocities near the bounds when sampling from the truncated distribution. In load generation applications, this has proven to be too conservative as the share of extreme velocity fluctuations and thus strong accelerations is overestimated. An alternative, which does not accumulate velocity fluctuations near the bounds, is to truncate the process by resampling until a feasible solution is found. In practice, this could lead to undesired runtimes when the probability of achieving a feasible solution is low. By decreasing the standard deviation σ_ϵ of ϵ_{t_i} at each resampling iteration until zero, a solution within a

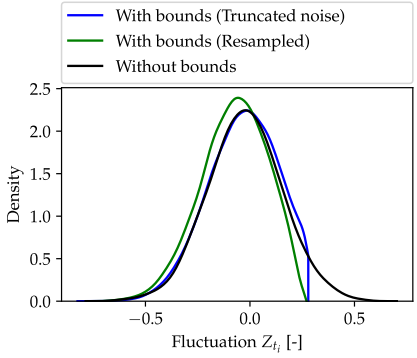
specific number of iterations N_{max} can be guaranteed. In the implemented solution σ_ϵ is reduced in each iteration k using the function

$$\sigma_{\epsilon,k} = \sigma_{\epsilon,0}\left(1 - \left(\frac{k}{N_{max}}\right)^a\right) \tag{3.18}$$

where in the following applications we set $a = 10$ and $N_{max} = 100$. As depicted in Figure 3.15a, this parametrization leads to an almost unchanged standard deviation $\sigma_{\epsilon,k}$ for the first approximately 60 iterations during which, in most cases, a feasible solution is found. After 60 iterations, an increasingly steep decline of the standard deviation begins, guaranteeing a feasible solution within a maximum of 100 iterations. By choosing other values of the parameters a and N_{max}, the shape of the truncated distribution can be modified. Decreasing the exponent a or the maximum number of iterations N_{max} leads to a narrower distribution, while increasing them has the opposite effect. Resulting from the resampling algorithm is a distribution of the stochastic process Z_{t_i} which is bounded by $\underline{z}$ and $\overline{z}$, where the distribution is not cut off at the bounds but, as desired, is gradually tapered towards its limits while accumulating more density in its center, see Figure 3.15b, leading to more realistic and less conservative truncated velocity fluctuations.



(a) Iterative reduction of standard deviation $\sigma_{\epsilon,k}$ according to (3.18) for $\sigma_{\epsilon,0} = 1$, $a = 10$ and $N_{max} = 100$



(b) Empirical distribution of a red-noise AR(1) process without truncation and truncated at $\overline{z} = 0.3$ by the approach described in [18] and by resampling

Figure 3.15: Properties and distribution of truncated stochastic process

3.2.4 Parameter modeling

As described in Section 3.2.2, sets of parameters are estimated from the velocity profile in each window. For the red-noise AR(1) model, those parameters are the correlation parameters η, θ and the white noise standard deviation σ_ζ. For the ARMA(2, 1) model these are the AR parameters φ_1, φ_2, the MA parameter θ_1 and the white noise standard deviation σ_ϵ. Those parameter sets are stored in a tuple, $\boldsymbol{p}_w = (\eta_w, \theta_w, \sigma_{\zeta,w})_w$ or $\boldsymbol{p}_w = (\varphi_{1,w}, \varphi_{2,w}, \theta_{1,w}, \sigma_{\epsilon,w})_w$ respectively, for each window w. For the truncation of the stochastic processes, see 3.2.3, an additional tuple $\boldsymbol{z}_w = (\underline{z}_w, \overline{z}_w)_w$ with the minima and maxima of the measured series $\{Z_{t_i}\}$ in each window w is stored. The parameter tuples $\boldsymbol{p}_w$ and $\boldsymbol{z}_w$ are binned according to the mean velocity $\bar{v}_w$ of the window.

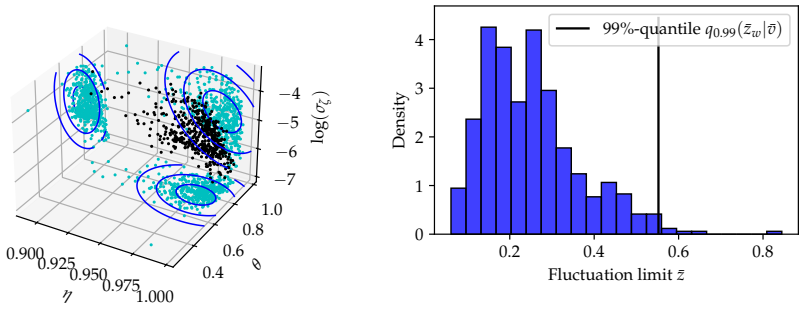
From the tuple with the stochastic process parameters $\boldsymbol{p}_w$ a multivariate distribution is estimated for each velocity bin. The distribution is described by a multivariate normal distribution

$$p_{\boldsymbol{P}|\bar{v}}(\boldsymbol{p}|\bar{v}) \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{P}|\bar{v}}, \boldsymbol{\Sigma}_{\boldsymbol{P}|\bar{v}}) \tag{3.19}$$

where the mean vector $\boldsymbol{\mu}_{\boldsymbol{P}|\bar{v}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{P}|\bar{v}}$ are estimated from the available parameter tuples $\boldsymbol{p}_w$ within the respective bin. Here, for each $\boldsymbol{p}_w$ in the bin we consider the logarithm $\log(\sigma_{\zeta,w})$ ($\log(\sigma_{\epsilon,w})$ respectively) of the standard deviation instead of the standard deviation itself such that the conditional distribution of $\sigma_{\zeta,w}$ ($\sigma_{\epsilon,w}$ respectively) with all other parameters fixed will be given by a one-dimensional log-normal distribution. Taking the logarithm and using the log-normal distribution provides a better fit of the distribution and ensures that no samples of the standard deviation < 0 are drawn. Certain parameter combinations violate the stability criterion for the two stochastic processes, as outlined in Section 2.2.1. The implementation of the criterion for the red-noise AR(1) process is straightforward, as it only requires the truncation of the multivariate normal distribution $p_{\boldsymbol{P}|\bar{v}}(\boldsymbol{p}|\bar{v})$ to stay within the bounds $|\eta| < 1$ and $|\theta| < 1$. In Figure 3.16a, this procedure is illustrated for the velocity bin including the mean velocities $100\,\mathrm{km/h} < \bar{v}_w < 110\,\mathrm{km/h}$, where the measured parameter tuples $\boldsymbol{p}_w$ of the red-noise AR(1) process and the projected 1, 2 and 3-sigma ellipses of the fitted normal distribution are depicted.

For each bin of the parameter tuples $\boldsymbol{z}_w$ the 99%-quantile $q_{0.99}(\bar{z}_w|\bar{v})$ and the 1%-quantile $q_{0.01}(\underline{z}_w|\bar{v})$ are determined. These quantiles are used to adapt proper bounds to the stochastic process Z_{t_i} as described in Section 3.2.3. Taking quantiles of the minima and maxima, the estimation of the bounds will be less sensitive to outliers, compared to taking the absolute minima and maxima. In Figure 3.16b the histogram of the

upper limits $\bar{z}_w$ and the corresponding 99%-quantile for the mean velocity range $100\,\mathrm{km/h} < \bar{v}_w < 110\,\mathrm{km/h}$ are depicted.



(a) Measured parameter tuples $\boldsymbol{p}_w$ and fitted multivariate normal distribution

(b) Histogram of upper limits $\bar{z}_w$ and the corresponding 99%-quantile

Figure 3.16: Distribution of $\boldsymbol{p}_w$ and $\bar{z}_w$ for the mean velocity bin $100\,\mathrm{km/h} < \bar{v}_w < 110\,\mathrm{km/h}$. Figure modified from [113].

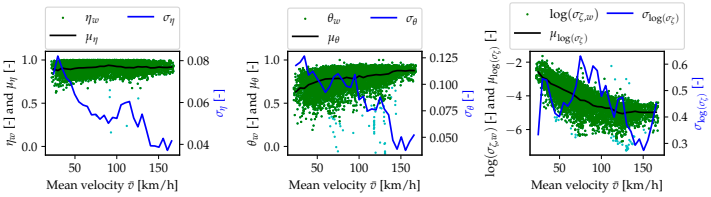
The estimation of the multivariate normal distribution $p_{\boldsymbol{P}|\bar{v}}(\boldsymbol{p}|\bar{v})$ for the stochastic process parameters and the derivation of the lower $q_{0.01}(\underline{z}_w|\bar{v})$ and upper quantiles $q_{0.99}(\bar{z}_w|\bar{v})$ is repeated for all mean velocity bins. Using continuous interpolation of the parameters $\boldsymbol{\mu}_{\boldsymbol{P}|\bar{v}}, \boldsymbol{\Sigma}_{\boldsymbol{P}|\bar{v}}$ and the bounds $q_{0.01}(\underline{z}_w|\bar{v})$ and $q_{0.99}(\bar{z}_w|\bar{v})$ along the mean velocity bins they can be described as a continuous function of the mean velocity $\bar{v}$.

Each point in the diagrams of Figure 3.17 indicates the estimated parameters $\boldsymbol{p}_w$ for one window. Cyan points are classified as outliers and are omitted in the estimation of the mean vector $\boldsymbol{\mu}_{\boldsymbol{P}|\bar{v}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{P}|\bar{v}}$. Outliers are classified as such when the parameter X_w (e.g., η_w) is outside the range

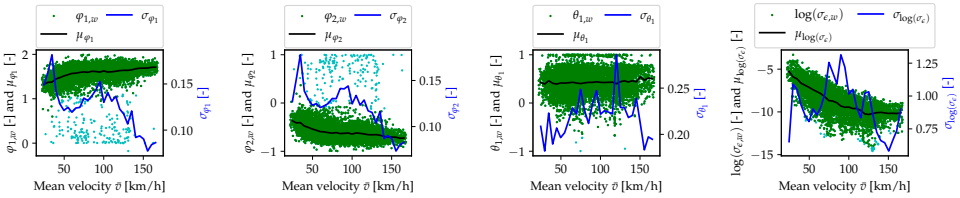
$$X_w|\bar{v} \notin (q_{0.25}(X_w|\bar{v}) - 3\ \mathrm{IQR}(X_w|\bar{v}), q_{0.75}(X_w|\bar{v}) + 3\ \mathrm{IQR}(X_w|\bar{v})) \tag{3.20}$$

with the $\mathrm{IQR}(X_w|\bar{v}) = q_{0.75}(X_w|\bar{v}) - q_{0.25}(X_w|\bar{v})$ specified for each velocity bin. It can be observed that the number of outliers is significantly increased for the ARMA(2, 1) model compared to the red-noise AR(1) model. This indicates that the MLE parameter estimation procedure is less robust due to the higher number of parameters of the ARMA(2, 1) model.

Significant trends of the parameters concerning the mean velocity can be observed. The mean of the red-noise AR(1) correlation parameter θ, for instance, increases with rising mean velocity, indicating less random velocity fluctuations corresponding to smoother driving behavior, see Figure 3.17a. This is sensible as on trips in urban areas, random events are usually less predictable and on short-term notice due to the higher



(a) Red-noise AR(1) parameters



(b) ARMA(2, 1) parameters

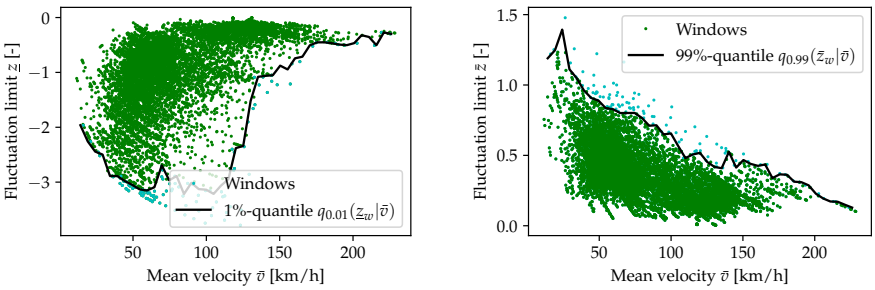
Figure 3.17: Measurements for the red-noise AR(1) and ARMA(2, 1) parameter tuples $\boldsymbol{p}_w$. For each measured window, a point is depicted in the diagrams along with the values of the mean vector $\boldsymbol{\mu}_{\boldsymbol{P}|\bar{v}}$ of each bin. The diagrams also show the standard deviation of each bin, i.e., the square root of the values on the diagonal of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{P}|\bar{v}}$. Cyan points are classified as outliers and are not included in calculating the mean vector and covariance matrix.

interaction with other road users and the worse long-distance visibility. On motorways, the velocities show higher correlations, i.e., are smoothed out, as the random events there are usually more predictable, such as cars braking in the distance or upcoming lane changes, and generally better long-distance visibility. Similar behavior can be observed for the AR parameters φ_1 and φ_2 of the ARMA(2, 1) model, which in absolute terms also increase with rising mean velocity, see Figure 3.17b. On the other hand the mean of the logarithmic standard deviations $\log(\sigma_\zeta)$ and $\log(\sigma_\epsilon)$ decreases when the mean velocity increases, indicating a decreasing amplitude of the velocity fluctuations with increasing velocity in line with the expectation of smoother driving behavior at higher velocities on motorways and vice versa. No mean trends can be observed for the parameters η and θ_1 as they stay relatively constant with rising mean velocity.

The behavior of the standard deviations of the parameters with respect to the mean velocity is less clear than for the mean of the parameters. The standard deviation of both red-noise AR(1) correlation parameters η and θ decreases with rising mean velocity, see Figure 3.17a. Therefore, the difference in velocity fluctuation correlation between windows with similar mean velocities decreases. At lower velocities, the fluctuations can vary greatly for the same mean velocity, as someone could be driving in a chaotic scenario with many random interactions or rather smoothly on a wide road with minimal interactions with pedestrians, bicycles, and other vehicles. At higher

velocities, driving tends to happen on increasingly similar roads (e.g., motorways), hence the standard deviation of the fluctuation correlation parameters decreases. The same tendency is observed for the standard deviation of the AR parameters φ_1 and φ_2 of the ARMA model as depicted in Figure 3.17b. For all other parameters, no significant trends in the standard deviations can be observed.

The trends of the stochastic process bounds $q_{0.01}(\underline{z}_w|\bar{v})$ and $q_{0.99}(\bar{z}_w|\bar{v})$, as depicted in Figure 3.18, follow the general tendencies observed for the stochastic process parameters mean $\boldsymbol{\mu}_{\boldsymbol{P}|\bar{v}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{P}|\bar{v}}$. As the mean velocity increases, the bounds converge towards smaller values (in absolute terms) as driving behavior becomes smoother and the standard deviation of the stochastic process decreases. The lower bounds $q_{0.01}(\underline{z}_w|\bar{v})$ increase towards small mean velocities $\bar{v} < 50\,\mathrm{km/h}$ as the velocity fluctuations $X_{t_i} = v_{t_i}/\bar{v}$ are bounded by the increasing ratio to small velocities v_{t_i} as $\bar{v}$ decreases and v_{t_i} cannot get any smaller as $v_{t_i} > 0$.



(a) $q_{0.01}(\underline{z}_w|\bar{v})$ **(b)** $q_{0.99}(\bar{z}_w|\bar{v})$

Figure 3.18: Measurements for the stochastic process bound tuples $\boldsymbol{z}_w$. For each measured window a point is depicted along with the values of the lower $q_{0.01}(\underline{z}_w|\bar{v})$ and upper quantiles $q_{0.99}(\bar{z}_w|\bar{v})$ for each bin. Cyan points are outside the defined lower and upper quantile bounds.

3.2.5 Relation to vehicle dynamics

The velocity fluctuations are related to the dynamics of the vehicle, which limit the amplitude and frequency of the fluctuations. The power that is necessary to change the velocity rises for increasing velocity, and thus puts a limit on the parameter relationships to the mean velocity observed in Figure 3.17. In the following Section, a theoretical bound on the velocity fluctuations as a function of the mean velocity and the power of the vehicle is derived.

Considering a simplified version of the vehicle dynamics model described in Section 2.1, where all resistance forces are neglected, results in the force

$$F_{acc}(t) = m_{veh} a_x(t) \tag{3.21}$$

needed to accelerate the vehicle with a_x being equal to the inertial force. The mechanical power needed for the acceleration of the mass m_{veh} is given by

$$P_{acc}(t) = v(t) F_{acc}(t) \tag{3.22}$$

with the force $F_{acc}(t)$ and velocity $v(t) > 0$. The velocity and acceleration are given by the mean velocity $\bar{v}$, the stochastic process $X_t > 0$ and the derivative dX_t/dt which leads to the description of the power

$$\begin{aligned} P_{acc}(t) &= v(t) F_{acc}(t), \\ P_{acc}(t) &= v(t) m_{veh} a_x(t), \\ P_{acc}(t) &= \bar{v} X_t m_{veh} \bar{v} \frac{dX_t}{dt}, \\ P_{acc}(t) &= m_{veh} \bar{v}^2 X_t \frac{dX_t}{dt}. \end{aligned} \tag{3.23}$$

Hereafter, we consider the absolute power used to drive the system (i.e., the vehicle)

$$|P_{acc}(t)| = \left| m_{veh} \bar{v}^2 X_t \frac{dX_t}{dt} \right| = m_{veh} \bar{v}^2 X_t \left| \frac{dX_t}{dt} \right| \tag{3.24}$$

as otherwise, accelerations and decelerations would partly cancel each other out, which is not the case, as no form of energy storage within the system is considered.

The stochastic process is given in its discrete form $\{X_{t_i}\}_{i=1}^{N}$ which leads to the discrete derivative of the stochastic process

$$\frac{dX_{t_i}}{dt} = \frac{X_{t_i} - X_{t_{i-1}}}{t_i - t_{i-1}}. \tag{3.25}$$

The mean absolute power needed to accelerate the vehicle is

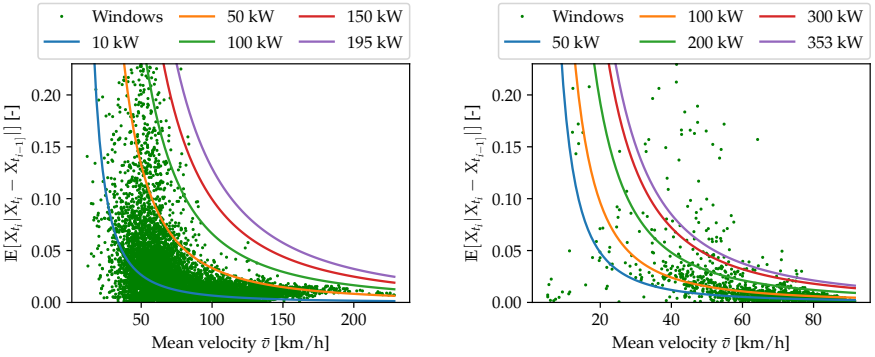
$$\begin{aligned} |\bar{P}_{acc}| &= \int_0^\infty p_{X_{t_i}}(x_{t_i}) m_{veh} \bar{v}^2 X_{t_i} \left| \frac{X_{t_i} - X_{t_{i-1}}}{t_i - t_{i-1}} \right| dx_{t_i}, \\ |\bar{P}_{acc}| &= \frac{m_{veh} \bar{v}^2}{t_i - t_{i-1}} \int_0^\infty p_{X_{t_i}}(x_{t_i}) X_{t_i} \left| X_{t_i} - X_{t_{i-1}} \right| dx_{t_i}, \\ |\bar{P}_{acc}| &= \frac{m_{veh} \bar{v}^2}{t_i - t_{i-1}} \mathbb{E}[X_{t_i} \left| X_{t_i} - X_{t_{i-1}} \right|], \end{aligned} \tag{3.26}$$

with the expected value $\mathbb{E}[X_{t_i} \left| X_{t_i} - X_{t_{i-1}} \right|]$ which is directly related to the properties and the parameters of the stochastic process $\{X_{t_i}\}_{i=1}^{N}$. The expected value can be expressed as a function of the mass of the vehicle m_{veh}, the mean velocity $\bar{v}$, the mean absolute power $|\bar{P}_{acc}|$ and the step width of the stochastic process $t_i - t_{i-1}$, i.e.,

$$\beta_{X_{t_i}} = \mathbb{E}[X_{t_i} \left| X_{t_i} - X_{t_{i-1}} \right|] = \frac{|\bar{P}_{acc}|(t_i - t_{i-1})}{m_{veh}\bar{v}^2}. \tag{3.27}$$

Figure 3.19a depicts the estimated expected value $\hat{\beta}_{X_{t_i}}$ for each measured window of the investigated driver and compares them with calculated values as a function of the mean velocity $\bar{v}$ and mean absolute power $|\bar{P}_{acc}|$. The observed shape of the point cloud is similar to that of the calculated curves. This observation relates to the trends observed for the process parameters, see Figure 3.17. The log-standard deviations $\log(\sigma_\zeta)$ and $\log(\sigma_\epsilon)$, for example, show a decline with rising mean velocity, as does the expected value $\beta_{X_{t_i}}$. The standard deviations directly relate to $\beta_{X_{t_i}}$, the trends observed in Figure 3.19a as a function of the mean absolute power $|\bar{P}_{acc}|$ therefore also influence the trends observed for the log-standard deviations $\log(\sigma_\zeta)$ and $\log(\sigma_\epsilon)$ in Figure 3.17. The curve characterized by the maximum power $P_{max} = 195\,\text{kW}$ of the vehicle should theoretically never be exceeded by the measured windows. Some windows still exceed this limit as a consequence of the simplifications in the model, including the neglect of resistance forces and not taking into account the directionality of the maximum power, i.e., limiting the braking power by the engine power P_{max}.

The velocity fluctuations of cars are significantly influenced by the limits of the vehicle dynamics, as shown in the precedent analysis, as well as the behavior of the driver. Velocity fluctuations of trucks are limited to a much greater extent by the vehicle dynamics. In Figure 3.19b, the measured values of $\hat{\beta}_{X_{t_i}}$ are compared to the calculated values for a heavy-duty truck. Although the measured values show the same tendencies overall, they are on average much closer to the curve specified by the maximum power of 353 kW compared to the car and exceed it more frequently, thereby indicating that the fluctuations are often limited by the vehicle dynamics, or even exceed those limits due to the simplified model. The more frequent exceedance of the maximum power limit for the heavy-duty truck compared to the car is also a consequence of the significantly higher resistance forces for trucks, resulting in a greater error when these forces are neglected.



(a) Car: $m_{veh} = 1955\,\text{kg}$, $P_{max} = 195\,\text{kW}$

(b) Heavy-duty truck: $m_{veh} = 33\,700\,\text{kg}$, $P_{max} = 353\,\text{kW}$

Figure 3.19: Measurements of $\mathbb{E}[X_{t_i}|X_{t_i} - X_{t_{i-1}}|]$ as a function of the mean velocity $\bar{v}$ compared to calculated curves for different values of the mean absolute power $|\bar{P}_{acc}|$, see (3.27).

3.2.6 Generation of velocity fluctuations

Based on the identified model(s), the truncation approach, and the parameter dependencies, an algorithm to generate velocity fluctuations including these findings is derived, see Algorithm 2 for an implementation of the red-noise AR(1) model. For the ARMA(2, 1) model, the algorithm can easily be adjusted accordingly by replacing the components concerning the stochastic process. The algorithm generates the goal velocity $v_{goal}(s) = v_{route}(s)X_s$ as a function of the route velocity $v_{route}(s)$, see also Section 3.1 for the integration into the velocity simulation. The fluctuations are generated separately for each road segment u with the distance increments $s_j \in \mathcal{S}_u$, $N_j = |\mathcal{S}_u|$ where $\Delta s = s_j - s_{j-1} = const.$, see Figure 3.2a for an example of the segmentation. The mean velocity $\bar{v}$ used to model the fluctuation parameter dependence is set to the mean route velocity of the segment, given by the harmonic mean

$$\bar{v}_u = \frac{N_j}{\sum_{s_j \in \mathcal{S}_u} \frac{1}{v_{route}(s_j)}} \tag{3.28}$$

which accounts for the varying size of the time steps $t_j - t_{j-1}$.

3.2.7 Discussion and results

The presented red-noise AR(1) and ARMA(2, 1) models are compared against an existing approach [85], see (3.9), and validated with velocity measurements. The results

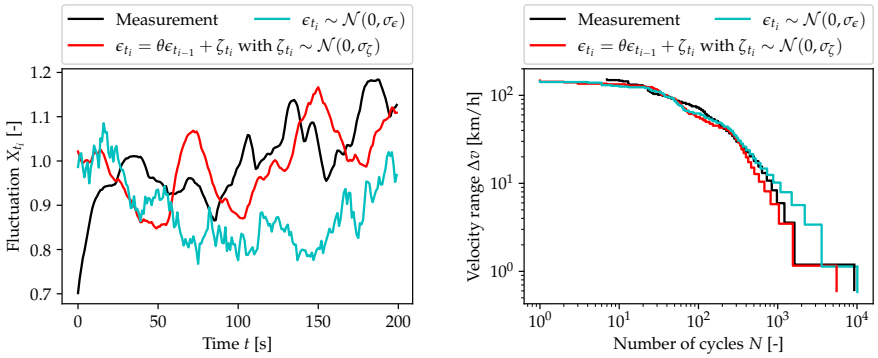
Algorithm 2 Generation of velocity fluctuations X_s and goal velocity $v_{goal}(s)$

Initialize route segment of length l_u with positions $s_j \in \mathcal{S}_u$
$\bar{v}_u \leftarrow \frac{N_j}{\sum_{s_j \in \mathcal{S}_u} \frac{1}{v_{route}(s_j)}}$
Draw $\boldsymbol{p}_u \sim p_{\boldsymbol{P}|\bar{v}}(\boldsymbol{p}|\bar{v}_u)$ with $\boldsymbol{p}_u = (\eta_u, \theta_u, \sigma_{\zeta,u})_u$
$\underline{z}_u \leftarrow q_{0.01}(\underline{z}_w|\bar{v}_u)$, $\bar{z}_u \leftarrow q_{0.99}(\bar{z}_w|\bar{v}_u)$
$\Delta t \leftarrow 1\,\mathrm{s}$
$\epsilon_s \leftarrow \frac{\Delta s}{10}$
$k \leftarrow 0$
$Z_{min} \leftarrow -\infty$, $Z_{max} \leftarrow \infty$
while $(Z_{min} < \underline{z}_u) \mid (Z_{max} > \bar{z}_u)$ **do**
    $\sigma_{\zeta,u,k} \leftarrow \sigma_{\zeta,u}\left(1 - \frac{k}{N_{max}}\right)^a$
    $\sigma_{Z,u,k} \leftarrow \sigma_{\zeta,u,k}\sqrt{\frac{\eta_u\theta_u+1}{(1-\eta_u^2)(1-\theta_u^2)(1-\eta_u\theta_u)}}$
    $\omega \leftarrow \sigma_{Z,u,k}$
    $i \leftarrow 1$
    $t_i \leftarrow 0$, $s_i \leftarrow 0$
    $\epsilon_{t_i} \leftarrow 0$, $Z_{t_i} \leftarrow 0$
    while $s_i < l_u$ **do**
        $i \leftarrow i + 1$
        $t_i \leftarrow t_{i-1} + \Delta t$
        Draw $\zeta_{t_i} \sim \mathcal{N}(0, \sigma_{\zeta,u,k})$
        $\epsilon_{t_i} \leftarrow \theta_u \epsilon_{t_{i-1}} + \zeta_{t_i}$
        $Z_{t_i} \leftarrow \eta_u Z_{t_{i-1}} + \epsilon_{t_i}$
        $X_{t_i} \leftarrow \exp\left(Z_{t_i} - \frac{\omega^2}{2}\right)$
        $v'_{goal} \leftarrow v_{route}(s_{i-1})X_{t_i}$
        $r \leftarrow 1$
        $s_{i,r} \leftarrow s_{i-1} + v'_{goal}\Delta t$
        $s_{i,r-1} \leftarrow s_{i,r} - \Delta s$
        while $s_{i,r} - s_{i,r-1} > \epsilon_s$ **do**
            $v'_{goal} \leftarrow \frac{v_{route}(s_{i,r})X_{t_i} + v_{route}(s_{i-1})X_{t_{i-1}}}{2}$
            $r \leftarrow r + 1$
            $s_{i,r} \leftarrow s_{i-1} + v'_{goal}\Delta t$
        end while
        $s_i \leftarrow s_{i,r}$
    end while
    $N_i \leftarrow i$
    $Z_{min} \leftarrow \underset{i=1,\ldots,N_i}{\arg\min}\, Z_{t_i} - \frac{\omega^2}{2}$
    $Z_{max} \leftarrow \underset{i=1,\ldots,N_i}{\arg\max}\, Z_{t_i} - \frac{\omega^2}{2}$
    $k \leftarrow k + 1$
end while
Get X_{s_j} for $s_j \in \mathcal{S}_u$ by linear interpolation between the pairs $\{(s_i, X_{t_i}) : i = 1, \ldots, N_i\}$
$v_{goal}(s_j) \leftarrow v_{route}(s_j)X_{s_j}$ for $s_j \in \mathcal{S}_u$

of the red-noise AR(1) model are presented, but as the previous Sections 3.2.2 and 3.2.4 demonstrate, similar results can be achieved for the ARMA(2, 1) model. The velocity fluctuation model alters step 2 of the velocity simulation as described in Section 3.1 by changing the stochastic component X_s in (3.2).

In Figure 3.20a we illustrate the simulated velocity fluctuations using both the existing white-noise AR(1) model described in (3.9), and the presented red-noise AR(1) model. Moreover, the measured velocity fluctuations for the respective route section are depicted, revealing that the red-noise AR(1) approach can reproduce the correlation structure of the fluctuations much better. Moreover, as an important property of the velocity series relevant for load generation, we consider the number and amplitude of rainflow cycles included in the series, see, for instance, [3]. To evaluate the rainflow cycles of the velocity series, we simulate multiple routes driven by the same driver using the velocity simulation model described in Section 3.1 and compare the white and red-noise AR(1) models, see Figure 3.20b. Comparing the rainflow histograms indicates that the red-noise AR(1) model leads to velocity profiles that well reproduce the number and amplitude of rainflow cycles compared to measured velocities. When using the

white-noise approach, a distinct overestimation of the number of small velocity cycles becomes visible, a result of the improper modeling of the correlation structure leading to more small fluctuations as visible in Figure 3.20a.



(a) Samples of logarithmic AR(1) processes with white- and red-noise compared to measurement **(b)** Sum of velocity rainflow cycles for all trips of a user

Figure 3.20: Comparison of measurements to simulation with white- and red-noise AR(1) fluctuation models

Another comparison is made based on a damage sum derived from the torque load on the drivetrain of an electric vehicle. The torque load is calculated from the measured and simulated velocities using a system simulation that models the vehicle and its drivetrain and, under consideration of the vehicle dynamics, outputs the respective torque series in the drivetrain. As outlined in [113], the torque series are split into windows of length $s_{w,l} = 5000\,\text{m}$ and shift $s_{w,s} = s_{w,l}$ to assess the component load. For each window w the gear damage sum $d_{G,w}$, as it could be used for gear dimensioning according to [50], is calculated. As we only investigate the influence of the load generation on the damage sum $d_{G,w}$ and do not consider a particular component and its properties the damage sum is normalized with the measured mean damage of all windows $\bar{d}_{G,meas}$ and the window length $s_{w,l}$

$$d'_{G,w} = \frac{d_{G,w}}{s_{w,l}\bar{d}_{G,meas}}. \tag{3.29}$$

Figure 3.21 depicts the CDF of the window damage sums $d'_{G,w}$ derived from the measured and simulated velocities of all trips of a user using the white-noise AR(1) and red-noise AR(1) models. The stronger fluctuations of the white-noise approach lead to an overestimation of the damage sum, while the red-noise approach accurately reproduces the damage sum distribution derived from the measured velocities.

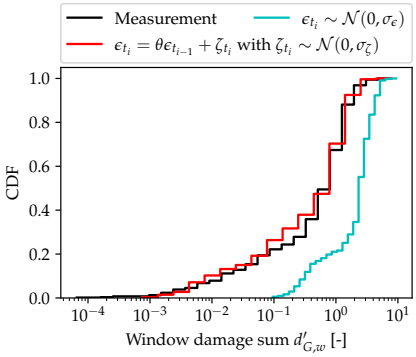
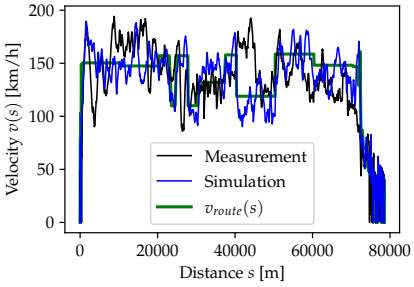
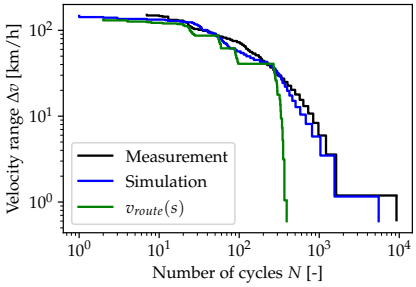


Figure 3.21: Comparison of window damage sum $d'_{G,w}$ distributions derived from measured and simulated velocities using white- and red-noise AR(1) fluctuation models

In Figure 3.22a we compare the simulated velocity series for a full motorway trip of length ~ 80 km with the respective measurement, displaying a good replication of the fluctuations in the measurement due to the correlation structure of the red-noise AR(1) process and the flexible parameter modeling which adjusts the variance and correlation parameters to the mean velocity. In Figure 3.22b, the rainflow histogram of the simulated velocities is compared to the simulation result $v_{route}(s)$ with deactivated fluctuations, i.e., $X_s = 1$. Demonstrating the need for proper modeling of velocity fluctuations, as otherwise the total number of velocity cycles is considerably underestimated.



(a) Comparison of simulated (with red-noise AR(1)) and measured velocity series for a motorway trip



(b) Sum of velocity rainflow cycles for all trips of a user

Figure 3.22: Comparison between simulated and measured velocity profiles

3.3 Modeling of traffic influence

The velocity simulation described in Section 3.1 does not consider any traffic influences, such as congestion, road closures, or construction activities, on the velocity. Traffic can influence the load on the components, as it can lead to lower mean velocities and more stops along the route. This can have an especially big influence on motorways, as there, the otherwise smooth driving with high velocities can switch to slow driving with high fluctuations and many stops. We therefore focus on the modeling of traffic influences on motorways, as this limits the length of the routes to be analyzed. Additionally, the available database for traffic analysis is larger on motorways, and the velocity profiles in general have a simpler structure. The analysis is further restricted to German motorways, although an extension to other countries would be straightforward when the necessary data is available. For further information on traffic modeling and the underlying theory, see [31, 100].

The available data to determine the velocity driven on motorways is FCD from the provider INRIX [48] and accompanying data collected from the Bundesanstalt für Straßen und Verkehrswesen (Federal Highway and Traffic Research Institute) (BASt) [5] at automatic permanent counting stations on motorways, where both datasets are collected from September 2018 to August 2019. The FCD consists of approximately $20 \cdot 10^9$ km of Global Positioning System (GPS) traces from roads in Germany, Austria, and Switzerland, although only the subset on German motorways is considered. The GPS sampling interval is variable, but for the majority of the dataset is in the range $30\,\mathrm{s} - 60\,\mathrm{s}$. From the GPS traces, the velocities can be derived as a function of time and position. The BASt data consist of traffic counting data collected at 2 114 automatic counting stations, of which 1 228 counting stations are on motorways. Each counting station is directional, and there are usually two counting stations at each location to count traffic in both directions. The dataset contains the total number of vehicles passing the stations in hourly resolution, thereby providing the ground truth for the total number of vehicles at each of the stations, but no information about the velocity of the vehicles. The counting stations are distributed across Germany in a non-uniform fashion, in order to support decision making in traffic planning [5]. Therefore, they are generally located more frequently in areas with high variability in traffic density and less frequently in areas where the traffic density is low and steady. See Figure 3.23 for the locations of all counting stations on German motorways.

Figure 3.23: Locations of BASt automatic permanent counting stations on German motorways

3.3.1 Location and time-dependent traffic speed distributions

The traffic speed, or space mean speed, denoted as $v_{tr}(s, t_h)$, is defined as the harmonic mean of the velocities of the vehicles passing a location s during a time interval t_h and is therefore given by

$$v_{tr}(s, t_h) = \frac{N_v}{\sum_{i=1}^{N_v} \frac{1}{v_i}} \tag{3.30}$$

with the number of vehicles N_v passing the location during the time interval, and the velocities of the individual vehicles v_i [31]. The size of the time interval is a compromise between the temporal resolution of the traffic speed and the available data to calculate the harmonic mean. We chose a length of 1 hour for the interval, which is sufficient to capture differences in the traffic speed between rush hour and light traffic. To simplify the analysis of the available FCD and maximize the number of data points, we consider each FCD velocity measurement at position s and time t during the time interval t_h individually and set $v_{tr}(s, t_h) = v_{FCD}(s, t)$ as $N_v = 1$.

The traffic speed at a specific location and time interval is considered as a random variable $V_{tr}(s, t_h)$ where the measured traffic speeds $v_{tr}(s, t_h)$ in the FCD dataset are samples from this distribution. Accordingly, the traffic speed at multiple positions and time intervals, is described by a discrete random field $\{V_{tr}(s, t_h) : s \in \mathcal{S}, t_h \in \mathcal{T}_h\}$ which is a function of the discrete time intervals t_h and positions s. Ideally, one would

estimate the distribution of the entire random field from the FCD samples, but this is associated with certain challenges described in the following.

For a unique time interval, i.e., an hour-long interval on a specific day, the FCD measurements contain at most a few random samples from the traffic speed $V_{tr}(s, t_h)$. To increase the available data, the time intervals t_h are clustered according to the time of day and weekday

$$t_h \in \mathcal{T}_h = \{\text{Mon. 0-1 am, Mon. 1-2 am, ..., Sun. 11 pm-12 am}\}, \tag{3.31}$$

as a repetitive pattern of the traffic speed for each week can be assumed, where we neglect any public holidays and other singular events. Concerning the position s we cannot apply any clustering as there is generally no similarity between the traffic speed on different road sections. Conditional on the time t_h the random field $\{V_{tr}(s, t_h)\}$ becomes a one-dimensional stochastic process $\{V_{tr}(s)|t_h\}$ with respect to position s along a route. This stochastic process cannot be assumed to be stationary, as the dependence of the traffic speed distribution $p_{V_{tr}}(v_{tr}|s, t_h)$ on the position s must especially be considered to distinguish between sections with low and high traffic density. The estimation of the stochastic process $\{V_{tr}(s)|t_h\}$ from the available data thus becomes more difficult, especially because the FCD is often fragmented into small pieces as drivers, e.g., only drive to the next slip road on the motorway. Therefore, it is difficult to get a sufficient number of samples $\{v_{tr}(s_k)|t_h : s_k \in \mathcal{S}_k\}$ on a specific road segment with positions $\mathcal{S}_k$ (e.g., from one motorway interchange to the next) in the time interval t_h, to describe the stochastic process on this segment $\{V_{tr}(s_k)|t_h : s_k \in \mathcal{S}_k\}$.

Due to the described challenges, the correlation of traffic speed distribution is neglected in a first step and only the stationary traffic speed distributions $p_{V_{tr}}(v_{tr}|s_m, t_h)$ for specific positions $\{s_m \in \mathcal{S}_m \subset \mathcal{S}\}$ are estimated from the FCD. These distributions indicate how the traffic speed can fluctuate for a given time interval t_h and position s_m, e.g., due to changes in the number of vehicles, roadworks, and weather. It is therefore not sufficient to analyze the mean traffic speed $\bar{v}_{tr}(s_m, t_h)$, as this would only lead to moderate differences between different positions and times and neglect the tails of the traffic speed distributions, which provide information about the frequency and intensity of congestion.

The position subset $\mathcal{S}_m$ is a reduction of all positions on the motorway network to handle the amount of data that must be analyzed to derive the distributions $p_{V_{tr}}(v_{tr}|s_m, t_h)$. Instead of uniformly distributing the positions within $\mathcal{S}_m$ on the motorway network in Germany, the positions of the automatic traffic counting stations are used. This gives a higher spatial resolution in parts of the network where changes in traffic speed are more frequent, and leads to a lower resolution when traffic jams are less common and

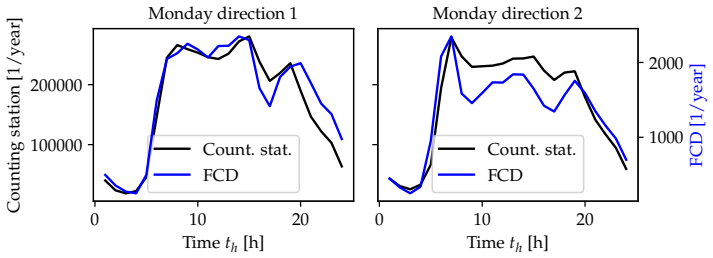
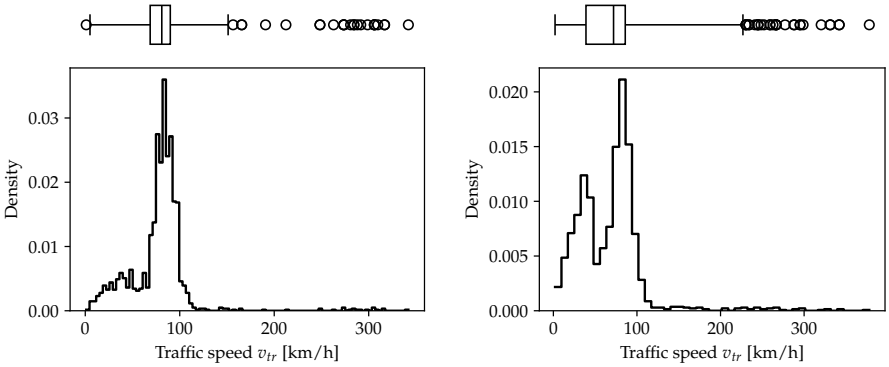


Figure 3.24: Total number of vehicles in a year at a traffic counting station and number of FCD measurements at the same location

the traffic speed typically stays constant. Each position in $\mathcal{S}_m$ is assigned all velocity measurements within the FCD database at the respective position, taking into account the correct driving direction. Comparison of the individual FCD velocity measurements with the traffic counting station data allows to assess their validity, i.e., if the measured FCD subset of all vehicles is sufficient to derive the traffic speed distribution for a given position and time. On average, 0.75% of all vehicles passing a counting station are included in the FCD. In Figure 3.24, the total number of vehicles per hour on Monday from the FCD measurements is compared to the traffic counting station data, separated by driving direction. The number of vehicles is the sum of all measurements across a full year. The Figure depicts how the number of vehicles changes in a day, with no significant changes in the share of FCD measurements relative to the total number of vehicles as measured by the counting stations. For periods of low traffic, the number of FCD vehicles drops to approximately 200 which is still sufficient to estimate the respective traffic speed distributions. For other locations in $\mathcal{S}_m$ the total number of vehicles is lower, and hence the number of FCD measurements also decreases. When less than 30 measurements are available for a specific time t_h and position s_m the data needs to be aggregated. Aggregation is done by first combining data from low traffic density times, as the distribution $p_{V_{tr}}(v_{tr}|s_m, t_h)$ shows little change during these times. On weekdays, the aggregation is done from 9 pm to 5 am, on weekends from 9 pm - 7 am. If this is not sufficient, all weekdays and the weekend are combined, as generally similar traffic patterns can be observed from Monday to Friday and on Saturday and Sunday. As the last aggregation step, all times are combined, leading to a single distribution $p_{V_{tr}}(v_{tr}|s_m, t_h)$ for the respective location, irrespective of the weekday and time. Of the 1 228 analyzed locations in $\mathcal{S}_m$ there are 9 remaining locations where, after all aggregation steps, no traffic speed distribution can be determined.

Figure 3.25 depicts the traffic speed histograms for a location at two different times determined with the FCD. Above the histograms are the corresponding box plots with

whisker length 3 IQR(V_{tr}). Both histograms show a significant amount of outliers towards high velocities > 130 km/h, caused by the simplification to approximate the traffic speed v_{tr} with the velocity v_{FCD} of the individual FCD cars. For this reason, in the further analysis of the traffic speed distribution $p_{V_{tr}}(v_{tr}|s_m, t_h)$, values outside the whiskers are classified as outliers and are not taken into consideration. This heuristic doesn't allow the removal of all outliers, as evident from Figure 3.25b, but removes the strongest ones for most distributions.



(a) Monday 6 pm - 7 pm

(b) Thursday 6 pm - 7 pm

Figure 3.25: Histograms of the measured traffic speed v_{tr} for a location s_m in two different time intervals t_h

Both histograms in Figure 3.25 are bimodal, typical for many observed traffic speed distributions, especially on heavily used roads at rush hour. Traffic flow near the maximum capacity of the road is unstable, meaning that small disturbances like a slow vehicle or an inattentive driver can cause smooth traffic to turn into a traffic jam [32, 44]. This instability causes the observed bimodality as the distribution is composed of two distributions, one for the traffic speed in congestion and one for the traffic speed in free flow. Gaussian mixture models (GMMs) are used to model this bimodality in the traffic speed distribution with the PDF given by

$$\begin{aligned} p_{V_{tr}}(v_{tr}|s_m, t_h) = &\frac{w_{v_{tr},s_m,t_h}}{\sqrt{2\pi\sigma^2_{1,v_{tr},s_m,t_h}}} \exp\left(-\frac{(v_{tr}-\mu_{1,v_{tr},s_m,t_h})}{2\sigma^2_{1,v_{tr},s_m,t_h}}\right) \\ &+ \frac{1-w_{v_{tr},s_m,t_h}}{\sqrt{2\pi\sigma^2_{2,v_{tr},s_m,t_h}}} \exp\left(-\frac{(v_{tr}-\mu_{2,v_{tr},s_m,t_h})}{2\sigma^2_{2,v_{tr},s_m,t_h}}\right). \end{aligned} \tag{3.32}$$

The parameters include the expectations μ_{1,v_{tr},s_m,t_h}, μ_{2,v_{tr},s_m,t_h} and variances $\sigma^2_{1,v_{tr},s_m,t_h}$, $\sigma^2_{2,v_{tr},s_m,t_h}$ of the two underlying Gaussian distributions as well as the mixture coeffi-

cient w_{v_{tr},s_m,t_h}. This set of parameters for the GMM is estimated using the Expectation–maximization (EM) algorithm [22]. A GMM is fitted to the FCD for each location s_m and, if enough data is available, for all time intervals t_h, otherwise, the previously mentioned aggregation scheme is used. Figure 3.26 depicts the histograms without outliers and the fitted GMMs for one location in the hourly time intervals from 1 pm to 5 pm. The GMM can capture the bimodality of the distribution and how it changes from off-peak times to rush hour. When the variance of the traffic speed increases during rush hour, e.g., 4 pm to 5 pm, outliers are not reliably removed, leading to long tails and a worse fit of the GMM.

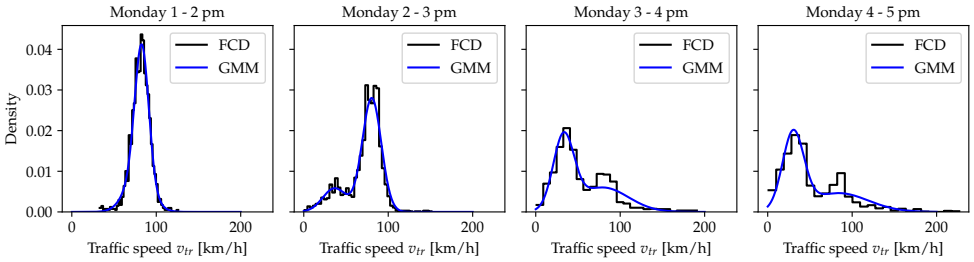
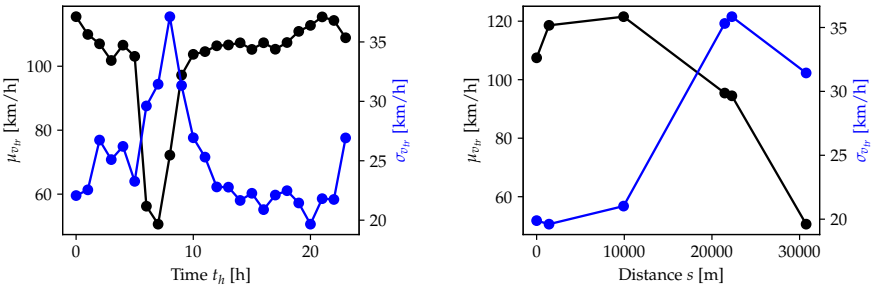


Figure 3.26: Approximation of the traffic speed distribution $p_{V_{tr}}(v_{tr}|s_m, t_h)$ with the GMM for different time intervals t_h at one location s_m

As demonstrated by Figure 3.26, the expected dependence of the traffic speed distribution on the time interval t_h is captured by the FCD and its GMM approximation. In Figure 3.27a, the change of traffic speed mean $\mu_{v_{tr}}$ and standard deviation $\sigma_{v_{tr}}$ across the time intervals t_h on Monday at another location s_m is depicted. The traffic at this location shows a typical commuting pattern with heavy traffic in the morning in the considered direction and heavy traffic in the opposite direction in the afternoon. Therefore, in the morning, the mean velocity drops sharply, coupled with an increase in standard deviation, while in the afternoon, no significant changes to the mean and standard deviation can be observed. Figure 3.27b shows how the mean and standard deviation change along a motorway section at 7 am on Monday, where the section ends with the location analyzed in Figure 3.27a. Along the route, a constant decrease in the mean of the traffic speed and an increase in the standard deviation can be observed. The considered motorway section leads towards an agglomeration (Stuttgart), for which the observed decrease in traffic speed is typical.

3.3.2 Congestion length distributions

The traffic speed distributions derived in Section 3.3.1 describe the distribution $p_{V_{tr}}(v_{tr}|s_m, t_h)$ of the random field $\{V_{tr}(s,t) : s \in \mathcal{S}, t \in \mathcal{T}\}$ at the discrete positions



(a) Traffic speed mean $\mu_{v_{tr}}$ and standard deviation $\sigma_{v_{tr}}$ across the time intervals t_h on Monday at the last position on the motorway section

(b) Traffic speed mean $\mu_{v_{tr}}$ and standard deviation $\sigma_{v_{tr}}$ along the motorway section at 7 am on Monday

Figure 3.27: Dependence of traffic speed distribution $p_{V_{tr}}(v_{tr}|s_m, t_h)$ on time interval t_h and position s_m along motorway section from Heilbronn to Stuttgart

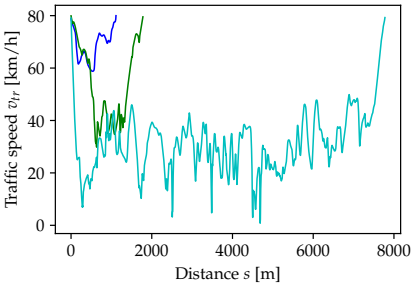
$s_m \in \mathcal{S}_m$ and times $t_h \in \mathcal{T}_h$. Therefore, the distribution of the stochastic process $\{V_{tr}(s)|t_h\}$ conditional on the time t_h at specific locations s_m is also known, but for the reasons given in Section 3.3.1 the correlations of the stochastic process between these locations cannot be derived from the available data without further assumptions. To simplify the problem, we neglect the influence of the traffic speed on the velocity of individual cars on motorways for values $v_{tr} > v_{tr,lim}$. For the correlations of the stochastic process, we assume that when the traffic speed v_{tr} drops below $v_{tr,lim}$, it stays constant for a certain distance l_{tr} before it switches to the next value. This distance, also called the congestion length, defines how long a traffic jam is once $v_{tr} < v_{tr,lim}$ and can be derived from the available data. A traffic jam is defined as an event that starts when the FCD vehicle speed v_{FCD} drops at least 40 km/h below the speed limit v_{lim} and stays below the threshold $v_{tr,lim} = 80$ km/h for 1000 m at a minimum while driving on a motorway. The congestion length is defined as the distance between the moment where $v_{FCD} < 80$ km/h and $v_{FCD} < v_{lim} - 40$ km/h until it gets back up to $v_{FCD} > 80$ km/h. The assumption is that the traffic speed v_{tr} can be approximated by the velocity v_{FCD} of individual vehicles. As already mentioned, we assume for simplification that the traffic speed v_{tr} stays constant once it drops below $v_{tr} < 80$ km/h until it gets back up to $v_{tr} > 80$ km/h. Thereby reducing the correlation structure of $\{V_{tr}(s)|t_h\}$ to velocity changes where the traffic speed v_{tr} stays constant for the congestion length l_{tr}.

In Figure 3.28a, the velocity profiles of three selected traffic jams, as extracted from the FCD with the preceding definition, are depicted. The selected traffic jams range from $l_{tr} \approx 1000$ m to $l_{tr} \approx 8000$ m. From each traffic jam, the mean traffic speed $\bar{v}_{tr}$, defined as the mean velocity between the start and end of the traffic jam, is derived. In Figure

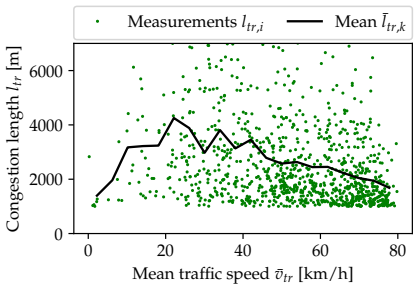
3.28b, the congestion length $l_{tr,i}$ of individual traffic jams i is plotted as a function of the respective mean traffic speed $\bar{v}_{tr,i}$. The Figure also shows the estimated mean of the congestion length

$$\bar{l}_{tr,k} = \frac{1}{N_k} \sum_{i \in \mathcal{I}_k} l_{tr,i} \tag{3.33}$$

derived for 20 mean traffic speed bins k which contain N_k measurements of the congestion length $l_{tr,i}$ and mean traffic speed $\bar{v}_{tr,i}$ with index $i \in \mathcal{I}_k$. For the velocity range $\bar{v}_{tr} > 20\,\mathrm{km/h}$ an approximately linear decrease in the mean congestion length $\bar{l}_{tr}$ can be observed as the mean traffic speed $\bar{v}_{tr}$ increases. For smaller values of $\bar{v}_{tr} < 20\,\mathrm{km/h}$, the mean congestion length drops while the total number of measured traffic jams decreases. Traffic jams with very low mean velocities are naturally rare, and it is plausible that their mean length is small, as otherwise, the time spent in the traffic jam would become unrealistically high.



(a) Traffic speed profiles from three selected traffic jams



(b) Measured congestion length $l_{tr,i}$ of individual traffic jams and mean congestion length $\bar{l}_{tr,k}$ as a function of the mean traffic speed $\bar{v}_{tr}$ ($l_{tr} > 7000\,\mathrm{m}$ not displayed)

Figure 3.28: Analysis of the relation between traffic speed v_{tr} and congestion length l_{tr}

The available data also allows for the analysis of the conditional congestion length distribution $p_{L_{tr}}(l_{tr}|\bar{v}_{tr})$ for each of the 20 mean traffic speed bins. Figure 3.29a depicts the congestion length histogram derived from the measured traffic jams with length $l_{tr,i}$ of one bin k with $40\,\mathrm{km/h} < \bar{v}_{tr,i} < 44\,\mathrm{km/h}$. The data is approximated by an exponential distribution

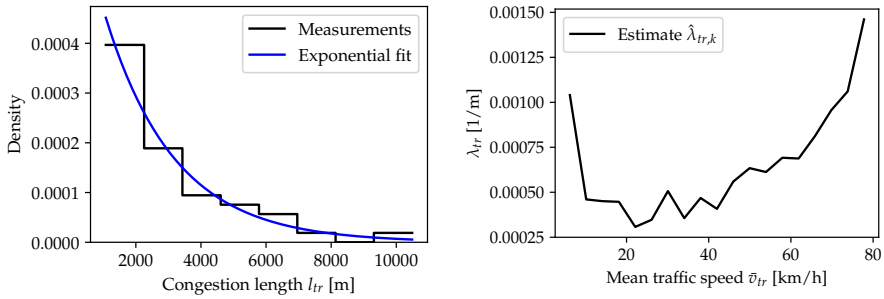
$$p_{L_{tr}}(l_{tr}|\bar{v}_{tr}) = \lambda_{tr} \exp(-\lambda_{tr}(l_{tr} - l_{tr,min})) \tag{3.34}$$

shifted by the minimum congestion length $l_{tr,min} = 1000\,\mathrm{m}$ where the parameter $\lambda_{tr}(\bar{v}_{tr})$ is a function of the mean traffic speed. The distribution parameter $\lambda_{tr,k}$ of a bin k is

related to the mean congestion length $\bar{l}_{tr,k}$ of the same bin trough the maximum likelihood estimator

$$\hat{\lambda}_{tr,k} = \frac{N_k}{\sum_{i \in \mathcal{I}_k} (l_{tr,i} - l_{tr,min})} = \frac{1}{\frac{1}{N_k} \sum_{i \in \mathcal{I}_k} l_{tr,i} - l_{tr,min}} = \frac{1}{\bar{l}_{tr,k} - l_{tr,min}}. \tag{3.35}$$

As depicted in Figure 3.29b, a similar dependence of the parameter $\lambda_{tr,k}$ with respect to the mean traffic speed $\bar{v}_{tr}$ can be observed as for the mean congestion length $\bar{l}_{tr,k}$ in Figure 3.28b. Values of the function $\lambda_{tr}(\bar{v}_{tr})$ between the estimates $\hat{\lambda}_{tr,k}$ of the bins are obtained by linear interpolation, see Figure 3.29b.



(a) Conditional congestion length distribution $p_{L_{tr}}(l_{tr}|\bar{v}_{tr})$ for 40 km/h $< \bar{v}_{tr,i} <$ 44 km/h

(b) Estimated distribution parameters $\hat{\lambda}_{tr,k}$ as a function of the mean traffic speed $\bar{v}_{tr}$

Figure 3.29: Analysis of conditional congestion length distribution $p_{L_{tr}}(l_{tr}|\bar{v}_{tr})$

3.3.3 Traffic dependent velocity profile generation

Based on the derived stationary traffic speed $p_{V_{tr}}(v_{tr}|s_m, t_h)$ and the conditional congestion length $p_{L_{tr}}(l_{tr}|\bar{v}_{tr})$ distributions realistic profiles of the traffic speed $v_{tr}(s)$ on arbitrary motorways in Germany can be generated. These profiles serve as an additional limit on the route velocity $v_{route}(s)$ as described in Section 3.1, see (3.3). To obtain stationary traffic speed distributions at arbitrary positions $s \in \mathcal{S}_k$ on a motorway route segment, the distribution parameters $\mu_{1,v_{tr}}$, $\mu_{2,v_{tr}}$, $\sigma^2_{1,v_{tr}}$, $\sigma^2_{2,v_{tr}}$, and $w_{v_{tr}}$ are linearly interpolated between the supporting points $s_m \in \mathcal{S}_m$ for a given time interval t_h. This yields the stationary traffic speed distribution

$$p_{V_{tr}}(v_{tr}|s, t_h; \mu_{1,v_{tr},s,t_h}, \mu_{2,v_{tr},s,t_h}, \sigma^2_{1,v_{tr},s,t_h}, \sigma^2_{2,v_{tr},s,t_h}, w_{v_{tr},s,t_h}) \tag{3.36}$$

for any position s along a motorway route segment.

The traffic speed profiles for a fixed time $t_{h,j}$ are modeled as a stochastic process $\{V_{tr,s_j}|t_{h,j}\}_{j=1}^{N}$ with the distance increments $s_1, ..., s_N$. The traffic speed values V_{tr,s_j} are drawn along the route at positions s_j and times $t_{h,j}$ and stay constant at the drawn speed $v_{tr,s_j} \sim p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ for the congestion length drawn from $l_{tr,s_j} \sim p_{L_{tr}}(l_{tr,s_j}|v_{tr,s_j})$ where the measured mean traffic speed $\bar{v}_{tr}$ in (3.34) is replaced by the drawn traffic speed v_{tr,s_j}. When the position $s_{j+1} = s_j + l_{tr,s_j}$ is reached the time $t_{h,j+1} = t_{h,j} + l_{tr,s_j}/v_{tr,s_j}$ is updated and a new traffic speed is drawn from $v_{tr,s_{j+1}} \sim p'_{V_{tr}}(v_{tr,s_{j+1}}|t_{h,j+1})$.

We draw from the modified distribution $p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ instead of from the distribution $p_{V_{tr}}(v_{tr}|s_j, t_{h,j})$ interpolated from the FCD, see (3.36), to avoid a bias in the resulting traffic speed distribution $p_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ at location s_j. This bias is due to the traffic speed remaining constant for a distance determined by the congestion length L_{tr}, whose mean conditional on the traffic speed is given by

$$\mathbb{E}[L_{tr}|v_{tr,s_j}] = \frac{1}{\lambda_{tr}(v_{tr,s_j})}. \tag{3.37}$$

Those differences in the mean congestion length depending on the traffic speed lead to different weights being put on the drawn traffic speed samples v_{tr,s_j}. To account for the different congestion lengths, the distribution $p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ from which the samples are drawn, is chosen such that the stationary distribution $p_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ of the traffic speed stochastic process at location s_j equals the FCD traffic speed distribution $p_{V_{tr}}(v_{tr}|s_j, t_{h,j})$.

The weighted stationary distribution of the traffic speed stochastic process is given by

$$p_{V_{tr}}(v_{tr,s_j}|t_{h,j}) = \frac{p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]}{\int_0^\infty p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr}} \tag{3.38}$$

where the integral in the denominator is a normalization constant that can be split at the traffic speed threshold $v_{tr,lim}$

$$\begin{aligned}&\int_0^\infty p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr}\\ &= \int_0^{v_{tr,lim}} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr} + \int_{v_{tr,lim}}^\infty p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,lim}]dv_{tr}.\end{aligned} \tag{3.39}$$

For traffic speeds v_{tr,s_j} below $v_{tr,lim}$ traffic jams are considered with the derived length distribution $p_{L_{tr}}(l_{tr,s_j}|v_{tr,s_j})$ which depends on the drawn traffic speed through the parameter $\lambda_{tr}(v_{tr,s_j})$. Above the threshold the distribution of the length until the next draw, i.e., the analog to the congestion length for $v_{tr,s_j} > v_{tr,lim}$, is assumed to be a shifted Exponential of the same structure as $p_{L_{tr}}(l_{tr,s_j}|v_{tr,s_j})$, see (3.34), with the

unknown parameter $\lambda_{tr,lim}$ which is a constant and not a function of the drawn traffic speed v_{tr,s_j}.

The ratio of the two integrals in (3.39) is given by the ratio of the probability below $v_{tr,lim}$ to the probability above $v_{tr,lim}$

$$\frac{\int_0^{v_{tr,lim}} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr}}{\int_{v_{tr,lim}}^{\infty} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,lim}]dv_{tr}} = \frac{P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}{1 - P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}. \tag{3.40}$$

in the known stochastic process distribution $p_{V_{tr}}(v_{tr,s_j}|t_{h,j})$, see (3.38). Inserting (3.40) back into (3.39) and incorporating (3.38) yields to

$$p_{V_{tr}}(v_{tr,s_j}|t_{h,j}) = \frac{p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]}{\left(1 + \frac{P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}{1 - P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}\right)\int_0^{v_{tr,lim}} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr}} \tag{3.41}$$

which for a given $p_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ and $\mathbb{E}[L_{tr}|v_{tr,s_j}]$ can be solved numerically for the density $p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$. The density $p_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ is given through the interpolated traffic speed distribution obtained from the FCD, see (3.36), to obtain $\mathbb{E}[L_{tr}|v_{tr,s_j}]$, as a function of v_{tr,s_j}, see (3.37).

To solve (3.41) for densities $p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ where $v_{tr,s_j} > v_{tr,lim}$ the expected value $\mathbb{E}[L_{tr}|v_{tr,s_j}]$ of that traffic speed range, i.e., $\mathbb{E}[L_{tr}|v_{tr,lim}]$, is required. To obtain $\mathbb{E}[L_{tr}|v_{tr,lim}]$ Equation (3.40) can be rearranged to

$$\begin{aligned}
&\int_{v_{tr,lim}}^{\infty} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,lim}]dv_{tr} \\
&= \int_0^{v_{tr,lim}} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr}\frac{1 - P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}{P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}, \\
&\mathbb{E}[L_{tr}|v_{tr,lim}] = \frac{\int_0^{v_{tr,lim}} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr}}{\int_{v_{tr,lim}}^{\infty} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})dv_{tr}}\frac{1 - P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}{P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}, \\
&\mathbb{E}[L_{tr}|v_{tr,lim}] = \frac{\int_0^{v_{tr,lim}} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})\mathbb{E}[L_{tr}|v_{tr,s_j}]dv_{tr}}{1 - \int_0^{v_{tr,lim}} p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})dv_{tr}}\frac{1 - P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})}{P(V_{tr,s_j} < v_{tr,lim}|t_{h,j})},
\end{aligned} \tag{3.42}$$

where $\mathbb{E}[L_{tr}|v_{tr,lim}]$ can be pulled out of the integral as it is not a function of the traffic speed v_{tr}. Further, $p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ can be determined for the traffic speed range

$v_{tr,s_j} < v_{tr,lim}$ with (3.41). From the expectation $\mathbb{E}[L_{tr}|v_{tr,lim}]$ we determine the lambda parameter

$$\lambda_{tr,lim} = \frac{1}{\mathbb{E}[L_{tr}|v_{tr,lim}] - l_{tr,min}} \tag{3.43}$$

of the shifted exponential distribution $p_{L_{tr}}(l_{tr,s_j}|v_{tr,s_j})$ for the distance l_{tr,s_j} until the next draw from $p'_{V_{tr}}(v_{tr,s_{j+1}}|t_{h,j+1})$ when $v_{tr,s_j} > v_{tr,lim}$. After determining $\mathbb{E}[L_{tr}|v_{tr,lim}]$, the distribution $p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ is fully defined by (3.38) for a given $p_{V_{tr}}(v_{tr,s_j}|t_{h,j})$.

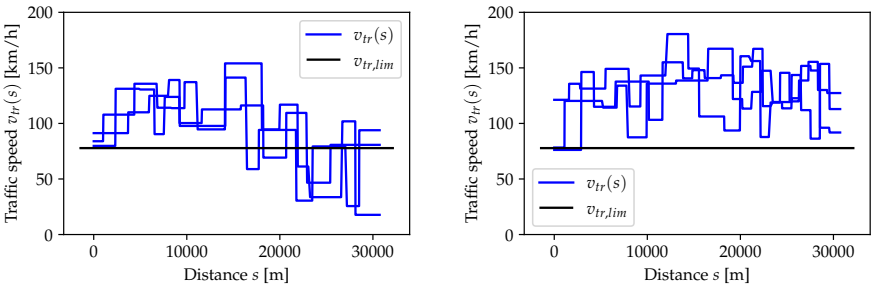
The previously described procedure to generate traffic speed profiles $v_{tr}(s)$ on German motorways with the stochastic process $\{V_{tr,s_j}|t_{h,j}\}_{j=1}^{N}$ is summarized in Algorithm 3.

Algorithm 3 Generation of traffic speed profiles $v_{tr}(s)$

Initialize motorway segment of length l_{seg} with positions $s \in \mathcal{S}_k = [0, l_{seg}]$
Get all supporting points for which the stationary traffic speed distribution is available $s_{mk} \in \mathcal{S}_m \cap \mathcal{S}_k$
$j \leftarrow 1$
$s_j \leftarrow 0$
Set start time $t_{h,j}$ (day of week and hour of day, e.g., Monday 7 am)
while $s_j \leq l_{seg}$ **do**
    Get distribution $p_{V_{tr}}(v_{tr}|s_j, t_{h,j}; \mu_{1,v_{tr},s_j,t_{h,j}}, ..., w_{v_{tr},s_j,t_{h,j}})$ by linear interpolation
    Derive $\lambda_{tr,lim}$ and $p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$ using (3.38) - (3.43)
    Draw $v_{tr,s_j} \sim p'_{V_{tr}}(v_{tr,s_j}|t_{h,j})$
    Draw $l_{tr,s_j} \sim p_{L_{tr}}(l_{tr,s_j}|v_{tr,s_j})$
    $s_{j+1} \leftarrow s_j + l_{tr,s_j}$
    $v_{tr}(s) \leftarrow v_{tr,s_j}$ for $s \in [s_j, s_{j+1})$
    $t_{h,j+1} \leftarrow t_{h,j} + \frac{l_{tr,s_j}}{v_{tr,s_j}}$
    $j \leftarrow j + 1$
end while
$N_j \leftarrow j$

3.3.4 Discussion and results

Figure 3.30 shows multiple traffic speed profiles for two different starting times generated with Algorithm 3 on the same motorway section as in Figure 3.27. The samples in Figure 3.30a for Monday at 7 am show an increasing probability for low traffic speeds $v_{tr} < v_{tr,lim}$ towards the end of the motorway section which is consistent with the trends of the traffic speed mean μ_{tr} and standard deviation σ_{tr} on the same route section at the same time, see Figure 3.27b. The traffic speed profiles for Sunday at 7 am, see Figure 3.30b, only drop slightly below $v_{tr,lim}$, which is consistent with the expected low traffic density at this time.



(a) Monday at 7 am

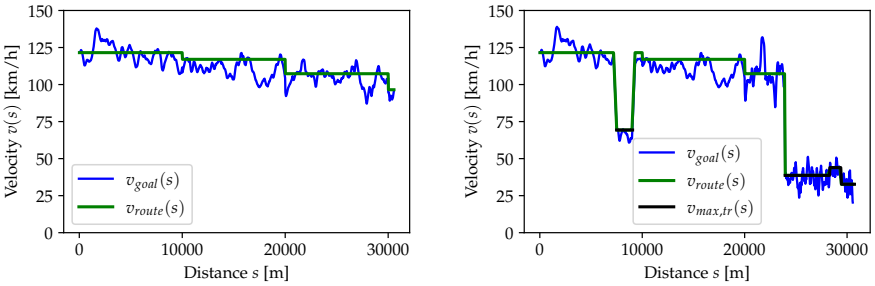
(b) Sunday at 7 am

Figure 3.30: Traffic speed profiles generated with Algorithm 3 on the motorway section from Heilbronn to Stuttgart for two different starting times

In the velocity simulation only traffic speeds below $v_{tr,lim}$ are considered as for $v_{tr}(s) > v_{tr,lim}$ the traffic speed is not expected to limit the desired route velocity $v_{route}(s)$ of the driver, see (3.3). The traffic speed limit $v_{max,tr}(s)$ used in (3.3) is therefore set to

$$v_{max,tr}(s) = \begin{cases} v_{tr}(s) & \text{if } v_{tr}(s) < v_{tr,lim} \\ \infty & \text{if } v_{tr}(s) \geq v_{tr,lim} \end{cases} . \tag{3.44}$$

Figure 3.31b depicts how the traffic speed limit $v_{max,tr}(s)$ can change the route velocity $v_{route}(s)$ and subsequently the goal velocity $v_{goal}(s)$ on a motorway section, compared to the simulation results obtained without taking the traffic speed limit into account, see Figure 3.31a.



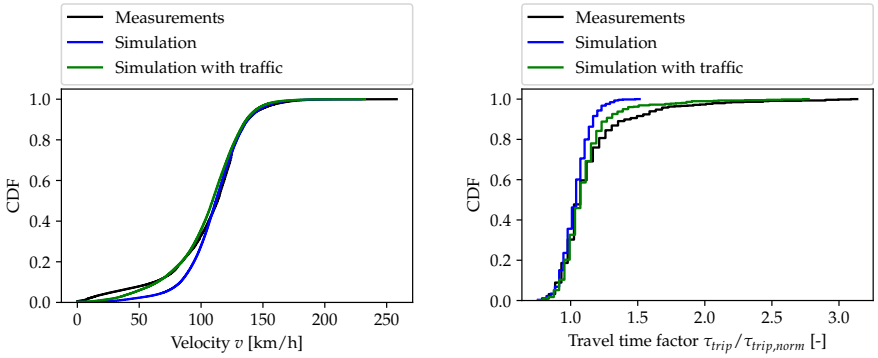
(a) Results without traffic speed limit $v_{max,tr}(s)$

(b) Results with traffic speed limit $v_{max,tr}(s)$

Figure 3.31: Route velocity $v_{route}(s)$ and goal velocity $v_{goal}(s)$ on a motorway section

When considering multiple motorway trips done by different drivers (approximately 32 000 km), the effect of the traffic speed limit on the simulated velocity distribution can

be demonstrated. Figure 3.32a depicts the CDF of the measured velocities on motorways compared to the simulated velocities $v(s)$ with and without the consideration of traffic through the traffic speed limit $v_{max,tr}(s)$. The simulation results with traffic include the traffic-related stop events added as a function of $v_{max,tr}(s)$, see Section 3.4.1. Considering traffic influences in the velocity simulation leads to a significantly better approximation of the measured velocity distribution. Without the traffic speed limit, velocities of $v < 100$ km/h on motorways are underrepresented by the simulation compared to the measurements. The simulation, when including the traffic speed limit, has an increased proportion of low velocities on motorways, where, compared to the measured velocities, the proportion is only slightly underestimated. Another comparison is made based on the travel time factor $\tau_{trip}/\tau_{trip,norm}$ which compares the travel time τ_{trip} spent on motorways for each trip with a hypothetical travel time $\tau_{trip,norm}$ obtained by assuming a mean velocity of 120 km/h on all motorways, see Figure 3.32b. When including the traffic influence, the velocity simulation can better replicate the proportion of high travel time factors occurring due to traffic. The simulation, when including the traffic influence, can reproduce the share of strong increases in the travel time factor $\tau_{trip}/\tau_{trip,norm} > 2$ which without the traffic influence never occurred in the simulation and only slightly underestimates the share of moderate travel time increases $1.2 < \tau_{trip}/\tau_{trip,norm} < 2$. It should be noted that the period when the FCD data used to derive the traffic speed limit $v_{max,tr}(s)$ was collected differs from the collection period of the measurement data used in Figure 3.32. Therefore, any temporal changes in traffic behavior, e.g., due to construction, are not considered in the comparison of measurement and simulation. This difference in the collection period, together with the used modeling assumptions, like the approximation of the traffic speed distributions in Section 3.3.1 with GMMs, contributes to the deviations of the simulation in the velocity and travel time factor distributions in Figure 3.32.



(a) Distribution of velocity on motorways

(b) Distribution of travel time factor for motorway share of trips (length of motorway share > 20 km)

Figure 3.32: Comparison of velocity simulation with and without traffic influence to measured velocities

3.4 Modeling of stop events

Stop events are part of the velocity simulation model, see Section 3.1, and occur at fixed locations such as traffic lights, yield or stop signs, as well as due to traffic jams. The modeling approach is split accordingly into traffic-related stop events, see Section 3.4.1, and ordinary stop events, which generally (but not always) occur at fixed locations, see Section 3.4.2. A stop in the following is defined as a drop of the measured velocity below the threshold $v_{S,th} = 1$ km/h to also account for almost stops and measurement uncertainty.

3.4.1 Traffic-related stop events

With the approach described in the previous Section 3.3, it is possible to generate velocity profiles considering traffic on motorways. These profiles so far only consider a reduction in the route velocity $v_{route}(s)$, see (3.3), but do not include additional stop events caused by high traffic densities on motorways. Traffic-related stops also occur on city and country roads, but as shown in Section 3.4.2, are difficult to classify as such, while on motorways, stops are necessarily caused by high traffic densities. As the traffic model described in Section 3.3 is limited to motorways, we restrict the generation of traffic-related stops to motorways.

We assume that the position of traffic-related stops in a traffic jam is given by a Poisson process with the stop rate $\lambda_{S,tr}$. In [85], an approach is described to model the mean

number of stop events in a fixed distance increment, i.e., the stop rate $\lambda_{S,tr}$, as a function of the traffic density ρ_{tr}. The approach takes the following linear form

$$\lambda_{S,tr}(\rho_{tr}) = \frac{\lambda_{S,tr,jam}}{\rho_{tr,jam} - \rho_{tr,crit}}(\rho_{tr} - \rho_{tr,crit}) \tag{3.45}$$

with the maximum stop rate $\lambda_{S,tr,jam}$ occurring at maximum traffic density $\rho_{tr,jam}$ and the critical traffic density $\rho_{tr,crit}$ above which traffic-related stops can occur. The fundamental diagram of traffic flow gives a relationship between the traffic speed v_{tr} and the traffic density ρ_{tr}. There exist different models for this relationship, with Greenberg's model [38] taking a logarithmic form

$$v_{tr} = v_{tr,q_{max}} \log\left(\frac{\rho_{tr,jam}}{\rho_{tr}}\right) \tag{3.46}$$

where $v_{tr,q_{max}}$ is the traffic speed that occurs at maximum traffic flow q_{max}, i.e., when the number of vehicles passing a location within a given time interval is at its maximum. Solving (3.46) for ρ_{tr} and inserting into (3.45) yields the exponential function

$$\begin{aligned}
\lambda_{S,tr}(v_{tr}) &= \frac{\lambda_{S,tr,jam}}{\rho_{tr,jam} - \frac{\rho_{tr,jam}}{\exp\left(\frac{v_{tr,crit}}{v_{tr,q_{max}}}\right)}}\left(\frac{\rho_{tr,jam}}{\exp\left(\frac{v_{tr}}{v_{tr,q_{max}}}\right)} - \frac{\rho_{tr,jam}}{\exp\left(\frac{v_{tr,crit}}{v_{tr,q_{max}}}\right)}\right), \\
\lambda_{S,tr}(v_{tr}) &= \lambda_{S,tr,jam}\frac{\exp\left(\frac{v_{tr,crit}-v_{tr}}{v_{tr,q_{max}}}\right) - 1}{\exp\left(\frac{v_{tr,crit}}{v_{tr,q_{max}}}\right) - 1},
\end{aligned} \tag{3.47}$$

which relates the traffic stop rate $\lambda_{S,tr}$ to the traffic speed v_{tr}. The parameters to fit the function are the maximum stop rate $\lambda_{S,tr,jam}$ which occurs when $v_{tr} = 0$, the critical velocity $v_{tr,crit}$ below which traffic-related stops are added, i.e., for traffic speeds above $v_{tr,crit}$ the function $\lambda_{S,tr}(v_{tr})$ is set to $\lambda_{S,tr} = 0$, and the traffic speed at maximum flow $v_{tr,q_{max}}$ which controls the shape of the function.

To determine the parameters $\lambda_{S,tr,jam}$, $v_{tr,crit}$, and $v_{tr,q_{max}}$ of the relationship in (3.47), a total of 863 stops on approximately 32 000 km driven on motorways in Germany are analyzed. Motorway trips are split into windows of length $l_{w,S} = 500$ m without overlap. For each window w the number of traffic stops $n_{S,tr,w}$ and the mean velocity $\bar{v}_w$ are recorded. For individual windows, the stop rate can be estimated as

$$\hat{\lambda}_{S,tr,w} = \frac{n_{S,tr,w}}{l_{w,S}}. \tag{3.48}$$

To improve the estimate, windows are binned according to their mean velocity. For one velocity bin k with bin mean velocity $\bar{v}_k$ that contains N_k windows w from the set of possible windows $\mathcal{W}_k$ the stop rate is estimated by

$$\hat{\lambda}_{S,tr,k} = \frac{\sum_{w \in \mathcal{W}_k} n_{S,tr,w}}{N_k l_{w,S}}. \tag{3.49}$$

The bin mean velocity $\bar{v}_k$ is used to replace the traffic speed v_{tr}, which leads to value pairs of $(\bar{v}_k, \hat{\lambda}_{S,tr,k})$, that are used to fit the parameters $\lambda_{S,tr,jam}$, $v_{tr,crit}$, and $v_{tr,q_{max}}$ of the exponential function $\lambda_{S,tr}(v_{tr})$ in (3.47). Figure 3.33a depicts the stop rate estimates for individual windows $\hat{\lambda}_{S,tr,w}$, the estimates derived from the binned windows $\hat{\lambda}_{S,tr,k}$, and the fitted function $\lambda_{S,tr}(v_{tr})$. The exponential function $\lambda_{S,tr}(v_{tr})$ provides a good fit of the stop rate $\hat{\lambda}_{S,tr,k}$ across the entire velocity range.

Based on the derived exponential stop rate function $\lambda_{S,tr}(v_{tr})$, traffic-related stops on motorways can be generated in the velocity simulation using the underlying Poisson process. The number of stops on a given motorway segment with length l_{tr} and traffic speed v_{tr} is given by the Poisson distribution

$$p_{N_{S,tr}}(n_{S,tr}) = \frac{(\lambda_{S,tr} l_{tr})^{n_{S,tr}} \exp(-\lambda_{S,tr} l_{tr})}{n_{S,tr}!}. \tag{3.50}$$

The mean number of stops on a segment is accordingly $\mathbb{E}[N_{S,tr}] = \lambda_{S,tr} l_{tr}$. The positions $s_{S,tr,1}, \ldots, s_{S,tr,N_{S,tr}}$ of the $N_{S,tr}$ stops on a segment from s_a to s_b are drawn from the uniform distribution

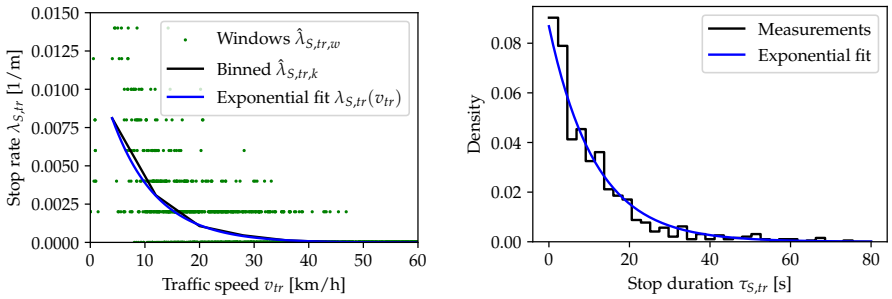
$$s_{S,tr,1}, \ldots, s_{S,tr,N_{S,tr}} \sim \mathcal{U}(s_a, s_b). \tag{3.51}$$

In addition to the positions of the traffic-related stops, their durations also need to be modeled. Figure 3.33b depicts the distribution of the duration $\tau_{S,tr}$ of the measured traffic-related stop events on motorways. The measured durations can be well approximated by an exponential distribution

$$p_{T_{S,tr}}(\tau_{S,tr}) = \lambda_{T_{S,tr}} \exp\left(-\lambda_{T_{S,tr}} \tau_{S,tr}\right) \tag{3.52}$$

with the distribution parameter $\lambda_{T_{S,tr}}$ and the mean stop duration $\bar{\tau}_{S,tr} = 1/\lambda_{T_{S,tr}}$.

Adding stop events to a given velocity profile changes its mean velocity, i.e., the mean $\bar{v} = l_{seg}/\tau_{seg}$ of the simulated velocity $v(t)$ on a given route section with constant route velocity v_{route} doesn't equal v_{route} as the stops reduce the mean velocity. In a section where the route velocity is given by the traffic speed $v_{route} = v_{max,tr}$, see (3.3), the



(a) Estimation of stop rate $\lambda_{S,tr}$ from measurements and fit of an exponential function ($\lambda_{S,tr} > 0.0151/\text{m}$ not displayed)

(b) Traffic-related stop duration distribution $p_{T_{S,tr}}(\tau_{S,tr})$

Figure 3.33: Analysis of the characteristics of traffic-related stop events on motorways

addition of stop events would therefore violate the assumption that the traffic speed v_{tr} equals the mean speed $\bar{v}$ of the section, see Section 3.3. Consequently, the stop rate $\lambda_{S,tr}(v_{tr})$ given as a function of the traffic speed would not match the resulting mean velocity $\bar{v}$ of the section. The traffic speed limit $v_{max,tr}$ thus needs to be adjusted such that $\bar{v} = v_{tr}$ and cannot be set to $v_{max,tr} = v_{tr}$ for $v_{tr} < v_{tr,lim}$ as in (3.44).

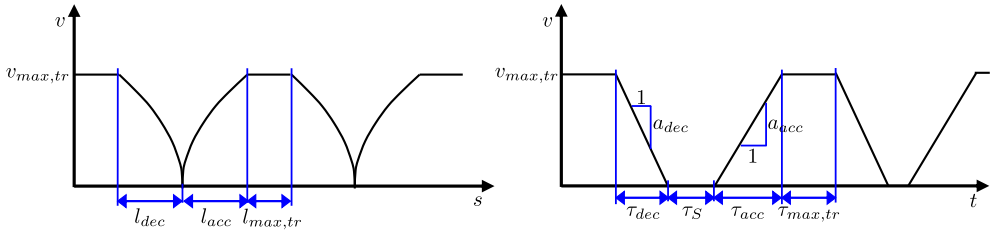


Figure 3.34: Velocity profile with two stop events as a function of distance $v(s)$ and time $v(t)$

Figure 3.34 depicts a velocity profile with two stop events on a route section with constant traffic speed limit $v_{max,tr}$ as a function of distance $v(s)$ and time $v(t)$. Velocity fluctuations are neglected, e.g., $v_{goal}(s) = v_{route}(s)$, and the dynamics are simplified to constant decelerations a_{dec} and accelerations a_{acc} limited by driver behavior and vehicle dynamics. The durations and distances of the acceleration phases are

$$\tau_{dec} = \frac{v_{max,tr}}{a_{dec}}, \tau_{acc} = \frac{v_{max,tr}}{a_{acc}}, l_{dec} = \frac{v_{max,tr}^2}{2a_{dec}}, l_{acc} = \frac{v_{max,tr}^2}{2a_{acc}}. \tag{3.53}$$

The average distance driven at velocity $v_{max,tr}$ on a section with length l_{seg} is

$$\bar{l}_{max,tr} = l_{seg} - \lambda_{S,tr} l_{seg} (l_{dec} + l_{acc}) \tag{3.54}$$

where $\lambda_{S,tr} l_{seg}$ is the mean number of stops on the section. The average time at $v_{max,tr}$ follows as $\bar{\tau}_{max,tr} = \bar{l}_{max,tr}/v_{max,tr}$. The mean duration of a traffic stop is given by $\bar{\tau}_{S,tr} = 1/\lambda_{T_{S,tr}}$. The mean velocity on the section including the stops is

$$\bar{v} = \frac{l_{seg}}{\tau_{seg}} = \frac{l_{seg}}{\lambda_{S,tr} l_{seg} \left(\tau_{dec} + \bar{\tau}_{S,tr} + \tau_{acc}\right) + \bar{\tau}_{max,tr}} \tag{3.55}$$

which, when setting $\bar{v} = v_{tr}$ and plugging in the respective formulas, can be solved for the stop-adjusted traffic speed limit

$$v_{max,tr,S} = \frac{-\lambda_{S,tr}\bar{\tau}_{S,tr} + \frac{1}{v_{tr}} \pm \sqrt{\left(\lambda_{S,tr}\bar{\tau}_{S,tr} - \frac{1}{v_{tr}}\right)^2 - 2\lambda_{S,tr} c}}{\lambda_{S,tr} c} \text{ with } c = \frac{a_{dec} + a_{acc}}{a_{dec} a_{acc}}. \tag{3.56}$$

Using (3.56), the case differentiation for the traffic speed limit $v_{max,tr}$ in (3.44) is modified to take the additional stops into account, which leads to

$$v_{max,tr}(s) = \begin{cases} v_{max,tr,S} & \text{if } v_{tr}(s) < v_{tr,crit} \\ v_{tr}(s) & \text{if } v_{tr}(s) \geq v_{tr,crit} \text{ and } v_{tr}(s) < v_{tr,lim} \\ \infty & \text{if } v_{tr}(s) \geq v_{tr,lim} \end{cases}. \tag{3.57}$$

The first solution of (3.56) with (+) leads to a negative distance $\bar{l}_{max,tr}$, hence only the second solution with (−) is physically plausible. To avoid imaginary solutions, the argument of the root must be positive, which leads to the condition

$$0 < \lambda_{S,tr} < \lambda_{S,tr,max} = -\sqrt{\frac{c(2\bar{\tau}_{S,tr} + c v_{tr})}{\bar{\tau}_{S,tr}^4 v_{tr}}} + \frac{1}{\bar{\tau}_{S,tr} v_{tr}} + \frac{c}{\bar{\tau}_{S,tr}^2} \tag{3.58}$$

for the traffic stop rate $\lambda_{S,tr}$. This condition ensures that for a given mean velocity $\bar{v} = v_{tr}$, mean stop duration $\bar{\tau}_{S,tr}$ and accelerations a_{dec}, a_{acc}, the stop rate $\lambda_{S,tr}$ can also be reached. Higher stop rates than $\lambda_{S,tr,max}$ reduce the mean velocity as the dynamics of the vehicle do not permit the necessary velocity spikes between the stops to reach the targeted mean velocity. The exponential relationship between the stop rate $\lambda_{S,tr}$ and the traffic speed v_{tr} described in (3.47) must thus be limited with the maximum possible stop rate $\lambda_{S,tr,max}$, see Figure 3.35. The limitation usually only affects trucks, as for cars, the vehicle acceleration capability is high enough to reach the targeted mean velocity for a given stop rate $\lambda_{S,tr}(v_{tr})$.

Combining the derived dependence of the stop rate $\lambda_{S,tr}$ on the traffic speed v_{tr}, see (3.47), the traffic stop duration distribution $p_{T_{S,tr}}(\tau_{S,tr})$ and the correction of the

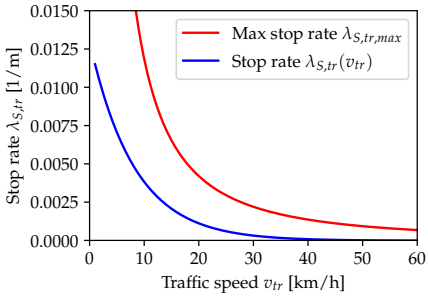


Figure 3.35: Stop rate function $\lambda_{S,tr}(v_{tr})$ and maximum possible stop rate $\lambda_{S,tr,max}$ calculated for $a_{dec} = a_{acc} = 1\,\mathrm{m/s^2}$ and $\bar{\tau}_{S,tr} = 12\,\mathrm{s}$

traffic speed limit $v_{max,tr}$ derived in (3.56), allows to include traffic-related stop events on motorways in the velocity simulation. The traffic-related stops are added in the velocity simulation using Algorithm 4, based on the traffic speed $v_{tr}(s)$ generated with Algorithm 3.

Algorithm 4 Generation of traffic-related stop events on motorways

Initialize motorway segment of length l_{seg} with positions $s \in \mathcal{S}_k = [0, l_{seg}]$
Get positions of traffic speed changes $\{s_j : j = 1, ..., N_j\}$, see Algorithm 3
for $j \leftarrow 1$ to $N_j - 1$ **do**
    if $v_{tr}(s_j) < v_{tr,crit}$ **then**
        $\lambda_{S,tr,j} \leftarrow \lambda_{S,tr,jam} \dfrac{\exp\left(\frac{v_{tr,crit} - v_{tr}(s_j)}{v_{tr,qmax}}\right) - 1}{\exp\left(\frac{v_{tr,crit}}{v_{tr,qmax}}\right) - 1}$
        $\lambda_{S,tr,max,j} \leftarrow -\sqrt{\dfrac{c(2\bar{\tau}_{S,tr} + c v_{tr}(s_j))}{\bar{\tau}_{S,tr}^4 v_{tr}(s_j)}} + \dfrac{1}{\bar{\tau}_{S,tr} v_{tr}(s_j)} + \dfrac{c}{\bar{\tau}_{S,tr}^2}$
        $\lambda_{S,tr,j} \leftarrow \min(\lambda_{S,tr,j}, \lambda_{S,tr,max,j})$
        Draw $N_{S,tr,j} \sim p_{N_{S,tr}}(n_{S,tr})$
        Draw $s_{S,tr,j,1}, ..., s_{S,tr,j,N_{S,tr,j}} \sim \mathcal{U}(s_j, s_{j+1})$
        Draw $\tau_{S,tr,j,1}, ..., \tau_{S,tr,j,N_{S,tr,j}} \sim p_{T_{S,tr}}(\tau_{S,tr})$
        $v_{max,tr}(s) \leftarrow \dfrac{-\lambda_{S,tr,j}\bar{\tau}_{S,tr} + \frac{1}{v_{tr}(s_j)} \pm \sqrt{\left(\lambda_{S,tr,j}\bar{\tau}_{S,tr} - \frac{1}{v_{tr}(s_j)}\right)^2 - 2\lambda_{S,tr,j}c}}{\lambda_{S,tr,j}c}$ for $s \in [s_j, s_{j+1})$
    else if $v_{tr}(s_j) < v_{tr,lim}$ **then**
        $v_{max,tr}(s) \leftarrow v_{tr,s_j}$ for $s \in [s_j, s_{j+1})$
    else
        $v_{max,tr}(s) \leftarrow \infty$ for $s \in [s_j, s_{j+1})$
    end if
end for

3.4.2 Ordinary stop events

The map data used in the velocity simulation gives the position of traffic lights, yield signs, stop signs, and pedestrian crossings along the simulated routes, which are all reasons for potential stop events. The velocity simulation, as described in Section 3.1.1 and in contrast to traffic-related stops, includes these stops by assigning a stop probability to each of those stop reasons, based on which the velocity simulation can add a stop event at the locations given by the map data. When a stop occurs in the simulation, the duration of the stop is drawn from a distribution specific to the stop reason. The aim is to derive the stop probability as well as the stop duration distribution for each of these stop reasons from data. To do so, the first step is to analyze the stops in the measured velocity profiles and match each stop to the most likely stop reason provided by the map data. Based on this, the measured stops can be classified according to their assigned stop reason. The stop probabilities and duration distributions are then derived using the classified measured stops and the information on all potential stop events given by the map data. The analysis is restricted to stops that happen on city and country roads, as stops on motorways must be traffic-related and are therefore handled in Section 3.4.1.

For each measured velocity profile, the positions $s_{S,i}$ of the n_S stops on city and country roads, as well as the positions $s_{Sr,r_S,j}$ of the n_{Sr} individual stop reasons from the classes

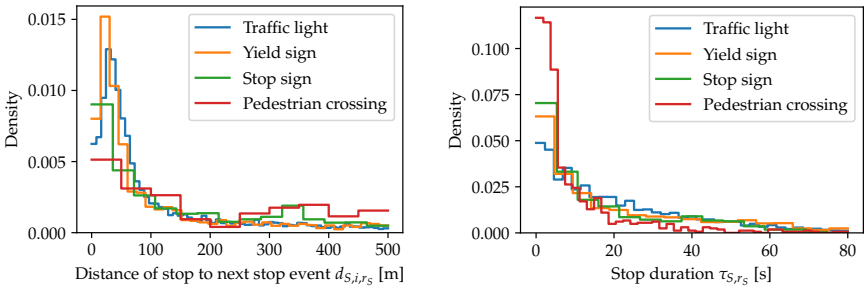
$$r_S \in \{\text{traffic light, yield sign, stop sign, pedestrian crossing}\} \tag{3.59}$$

are recorded. This leads to the distances $d_{S,i,j} = s_{Sr,r_S,j} - s_{S,i}$ between measured stops and stop reason locations in the map data. Each measured stop is assigned to the stop reason with the smallest positive distance

$$d_{S,i,r_S} = \underset{j=1,\dots,n_{Sr}}{\arg\min}\, d_{S,i,j} \text{ for } d_{S,i,j} \geq 0\,\text{m} \tag{3.60}$$

as stops should generally occur before the stop reason, i.e., stops due to traffic lights happen before the respective traffic lights, meaning that the distances d_{S,i,r_S} must be positive. In total, $n_S = 25\,883$ stops by different drivers on approximately 33 000 km driven on city and country roads in Germany are detected in the dataset and assigned to a stop reason. Stops at the start and end of a trip are assigned to a separate class, as they usually cannot be attributed to a stop reason in the map data and instead are caused by parking and maneuvering. The distances d_{S,i,r_S} of stops with the same stop class r_S are collected and plotted in the histograms in Figure 3.36a. All histograms show a decrease in the density towards large distances and converge to a constant value above zero. This pattern hints that after a certain distance, the stop is unlikely to be caused by

the assumed reason but instead is likely caused by other reasons, such as uncontrolled intersections, road narrowings, or traffic jams. After approximately $d_{S,th} = 200\,\text{m}$, the density in the histograms stays constant or only displays random fluctuations. Stops with distances to the next stop reason in the map data below the threshold $d_{S,i,r_S} \leq d_{S,th}$ are therefore classified with this stop reason. Stops where $d_{S,i,r_S} > d_{S,th}$ are categorized as unclassified and are handled separately.



(a) Distance of measured stops to the next stop reason d_{S,i,r_S}

(b) Stop duration distribution $p_{T_{S,r_S}}(\tau_{S,r_S})$

Figure 3.36: Analysis of measured stops on city and country roads

Table 3.2 lists the share $n_{S,r_S}/n_S$ of stops assigned to the different classes relative to the total number of stops. The majority of stops are caused by traffic lights, while a total of 30.3% of stops remain unclassified. The number of stops attributed to one of the stop classes n_{S,r_S} is compared to the total number of stop reasons of the same type n_{Sr,r_S} existing in the map data of the driven routes to estimate the stop probability

$$\hat{p}_S(s|r_S) = \frac{n_{S,r_S}}{n_{Sr,r_S}}. \tag{3.61}$$

Duplicate stops at the same stop reason, e.g., two stops in a queue behind a traffic light, are part of the stop count n_{S,r_S}, but are not taken explicitly into account by the estimated stop probability $\hat{p}_S(s|r_S)$ which is used in the velocity simulation. This means that in the simulation, where only a single stop can happen for each stop reason, the total number of stop reasons where the vehicle stops is higher than in the measurements, which compensates for the lack of duplicate stops. Table 3.2 gives the estimated values of the stop probability for the different stop classes. Some stops in the unclassified category are probably wrongly classified and are instead caused by one of the explicit stop reasons, leading to a systematic underestimation of the stop probabilities of the explicit classes. In [4, 19, 102, 104], stop probabilities at stop signs at different locations in the USA and Canada are reported in the range of $30\% - 55\%$, depending on the location and the used definition of a stop. The estimated stop probability $\hat{p}_S(s|\text{stop sign}) = 45\%$ for Germany,

given in Table 3.2, is therefore in a plausible range. For traffic lights in the USA, the probability of slowing down or stopping is stated at 36.5% in [49], which is above the estimated probability of $\hat{p}_S(s|\text{traffic light}) = 23\%$ for Germany. However, we mention that it also includes slowing down at traffic lights and not just stopping. The estimated stop probabilities are average values that are implicitly weighted by the number of vehicles that pass the respective stop reasons of each class in the measurements and do not reflect the different stop probabilities of individual stop reasons. In the context of most load-generation applications, this simplification is tolerable, as often the total number of deceleration and acceleration cycles caused by stops is relevant and not their specific location.

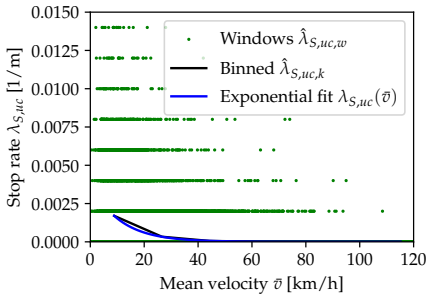
As for the traffic-related stops, see Section 3.4.1, distributions of the stop duration $p_{T_{S,r_S}}(\tau_{S,r_S})$ as a function of the stop class r_S are derived, see Figure 3.36b. For each class, the distribution is approximated by an exponential distribution, see (3.52), with the corresponding estimated distribution parameter $\hat{\lambda}_{T_{S,r_S}}$. In Table 3.2 the mean stop durations $\bar{\tau}_{S,r_S} = 1/\hat{\lambda}_{T_{S,r_S}}$ are listed. The ratios of the different mean stop durations are generally plausible, e.g., the mean stop duration for a pedestrian crossing is shorter than for a traffic light. Compared to the mean stop durations at yield and stop signs, the mean stop duration at traffic lights of 22 s seems to be underestimated but is still in a plausible range relative to mean durations of around 30 s given in [2, 76].

Table 3.2: Estimated properties of different stop classes on city and country roads

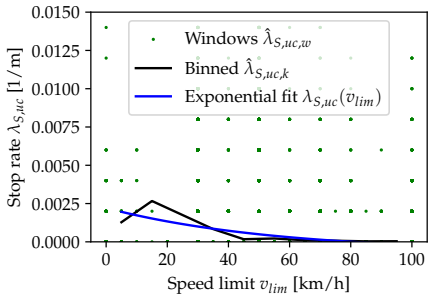
r_S	$n_{S,r_S}/n_S$	$\hat{p}_S(s\|r_S)$	$\bar{\tau}_{S,r_S} = 1/\hat{\lambda}_{T_{S,r_S}}$
Traffic light	38.0%	23%	22 s
Yield sign	6.0%	37%	23 s
Stop sign	1.7%	45%	18 s
Pedestrian crossing	3.6%	12%	12 s
Trip start/end	20.5%	–	12 s
Unclassified	30.3%	–	13 s

As unclassified stops form a significant share of all stops on city and country roads, ignoring them would lead to an underestimation of the total number of stops in the velocity simulation. The causes for these stops are diverse and include, among others, uncontrolled intersections, road narrowings, and traffic jams. The reasons for these stops cannot be modeled explicitly in the velocity simulation, instead, they are included with a Poisson process based on the model for the traffic-related stop events, see Section 3.4.1. The city and country portion of each trip is split into windows of length $l_{w,S} = 500\,\text{m}$ without overlap. For each window w the number of unclassified stops

$n_{S,uc,w}$, the mean velocity $\bar{v}_w$, and the speed limit $v_{lim,w}$ are recorded, whereby the speed limit is taken from map data, see Section 3.1.2. The stop rate of individual windows $\hat{\lambda}_{S,uc,w}$ can then be estimated with (3.48) from which the estimate $\hat{\lambda}_{S,uc,k}$ of binned windows follows with (3.49). As depicted in Figure 3.37, the windows are either binned according to their mean velocity $\bar{v}_w$ or their speed limit $v_{lim,w}$. In both cases the estimates $\hat{\lambda}_{S,uc,k}$ can be reasonably approximated by an exponential function $\lambda_{S,uc}(\bar{v})$, respectively $\lambda_{S,uc}(v_{lim})$, based on (3.45) with the parameters $\lambda_{S,uc,jam}$, $\bar{v}_{crit}$, and $\bar{v}_{q_{max}}$, respectively $v_{lim,crit}$ and $v_{lim,q_{max}}$. Compared to the stop rate of traffic-related stops on motorways, see Figure 3.33, the stop rate outliers of individual windows $\hat{\lambda}_{S,uc,w}$ relative to the binned estimates $\hat{\lambda}_{S,uc,k}$ are much stronger. Meaning that for some windows the stop rate is much higher or lower than captured by the estimates $\hat{\lambda}_{S,uc,k}$ and their exponential fit.



(a) Stop rate $\lambda_{S,uc}$ as a function of the mean velocity $\bar{v}$



(b) Stop rate $\lambda_{S,uc}$ as a function of the speed limit v_{lim}

Figure 3.37: Estimation of the stop rate $\lambda_{S,uc}$ from measurements and fit of an exponential function ($\lambda_{S,uc} > 0.015\,1/\text{m}$ not displayed)

Since there is no model for the traffic speed v_{tr} on city and country roads that could be used to model the mean velocity $\bar{v}$, the exponential function $\lambda_{S,uc}(v_{lim})$ for the stop rate based on the speed limit, which can be taken from map data, is applied in the velocity simulation. Following the model for traffic-related motorway stop events presented in Section 3.4.1, unclassified stops on city and country roads are generated using a Poisson process where the stop rate $\lambda_{S,uc}(v_{lim})$ is an exponential function of the speed limit v_{lim} of the respective city or country route segment. The number of stops on a given segment $N_{S,uc}$ with length l_{seg} and speed limit v_{lim} follows from a Poisson distribution $p_{N_{S,uc}}(n_{S,uc})$ analogue to the one for traffic-related motorway stops in (3.50) with the mean number of stops on a segment being $\mathbb{E}[N_{S,uc}] = \lambda_{S,uc} l_{seg}$. The stop positions $s_{S,uc,1}, \ldots, s_{S,uc,N_{S,uc}}$ of the $N_{S,uc}$ stops are drawn from a uniform distribution as described in (3.51) for the traffic-related stops on motorways.

Based on Algorithm 3 derived for the modeling of traffic-related motorway stops, Algorithm 5 is developed for the modeling of unclassified stops on city and country roads in the velocity simulation. In this case the correction of the route velocity v_{route} is not necessary as the stop rate $\lambda_{S,uc}$ is not a function of the mean velocity, respectively the traffic speed v_{tr}, but the speed limit v_{lim} which is unaffected by the added stop events. The stop rate $\lambda_{S,uc}$ is accordingly not limited by the vehicle and driver dynamics, unlike the traffic stop rate $\lambda_{S,tr}$ which is limited by $\lambda_{S,tr,max}$, see (3.58).

Algorithm 5 Generation of unclassified stop events on city and country roads

1: Initialize city or country segment of length l_{seg} with positions $s \in \mathcal{S}_k = [0, l_{seg}]$
2: Get positions of speed limit changes $\{s_j : j = 1, ..., N_j\}$
3: **for** $j \leftarrow 1$ to $N_j - 1$ **do**
4: **if** $v_{lim}(s_j) < v_{lim,crit}$ **then**
5: $\lambda_{S,uc,j} \leftarrow \lambda_{S,uc,jam} \frac{\exp\left(\frac{v_{lim,crit} - v_{lim}(s_j)}{v_{lim,qmax}}\right) - 1}{\exp\left(\frac{v_{lim,crit}}{v_{lim,qmax}}\right) - 1}$
6: Draw $N_{S,uc,j} \sim p_{N_{S,uc}}(n_{S,uc})$
7: Draw $s_{S,uc,j,1}, ..., s_{S,uc,j,N_{S,uc,j}} \sim \mathcal{U}(s_j, s_{j+1})$
8: Draw $\tau_{S,uc,j,1}, ..., \tau_{S,uc,j,N_{S,uc,j}} \sim p_{T_{S,uc}}(\tau_{S,uc})$
9: **end if**
10: **end for**

3.4.3 Discussion and results

The stop event models derived in Sections 3.4.1 and 3.4.2 are applied to simulate velocity profiles on a set of routes. The routes are those of the measured velocity profiles that are used to parametrize the stop event models. The stops in the simulated velocity profiles are compared to those of the measured profiles to analyze the suitability of the developed stop event models. All simulations are done with traffic influence on motorways based on the methods derived in Section 3.3.

Figure 3.38 depicts a measured and a simulated velocity profile on the same route, and the duration and reason for each stop. The simulation generates a realistic stop pattern, which overall is similar to the measured pattern. To better compare simulated and measured stops and account for their stochasticity, which permits an exact replication of any velocity profile, aggregate metrics are derived. The simplest one is the mean stop rate for each of the stop classes

$$\bar{\lambda}_{S,r_S} = \frac{n_{S,r_S}}{l_{total}} \tag{3.62}$$

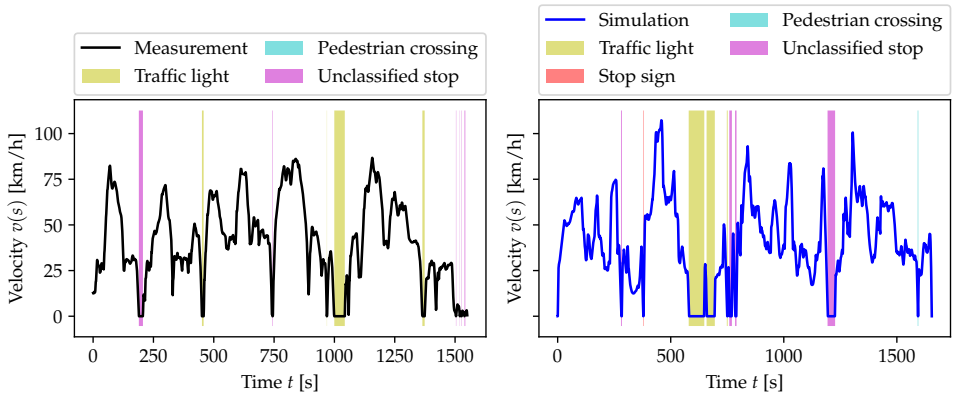


Figure 3.38: Stops in a measured and a simulated velocity profile on the same route

with the total number of stops of each stop class n_{S,r_S} and the total length of the analyzed set of routes $l_{total} \approx 65\,000$ km. Table 3.3 lists the different mean stop rates $\bar{\lambda}_{S,r_S,meas}$ and $\bar{\lambda}_{S,r_S,sim}$ derived from measured and simulated velocity profiles, and the relative error of the simulation

$$\delta_{\bar{\lambda}_{S,r_S}} = \frac{\bar{\lambda}_{S,r_S,sim} - \bar{\lambda}_{S,r_S,meas}}{\bar{\lambda}_{S,r_S,meas}}. \tag{3.63}$$

Stops at the start and end of the trips are excluded from the analysis, as the velocity simulation does not take parking and maneuvering stops into account. Additionally, some measured velocity profiles are missing data at the beginning or end of the trip due to the properties of the measurement system, which also prevents comparisons between simulation and measurement. The mean stop rates of the simulation are all in the same order of magnitude as the ones derived from measured velocities. Besides statistical errors in the estimation of the stop probabilities, some of the differences between simulation and measurement can be attributed to slight differences in the routes used in the simulation and those driven in the measurements. Those differences are caused by inaccuracies in the GPS signal or errors in the map data, which can lead to deviations in the simulated routes derived from the measured GPS traces. The unclassified stops on city and country roads, and the motorway traffic stops are subject to additional errors as their models require more assumptions. Both are based on an exponential fit of the stop rate as a function of the traffic speed or speed limit, see Figures 3.33a and 3.37b, where the quality of the fit influences the resulting number of stops in the simulation and thus the mean stop rate. The strong deviation of the mean stop rate of the traffic-related motorway stops is primarily caused by the underestimation

of the share of small traffic speeds v_{tr} on motorways in the simulation, as depicted in Figure 3.32a. This then leads to an underestimation of the number of motorway stops and thus the motorway mean stop rate due to the dependence of the stop rate $\lambda_{S,tr}$ on the traffic speed v_{tr}.

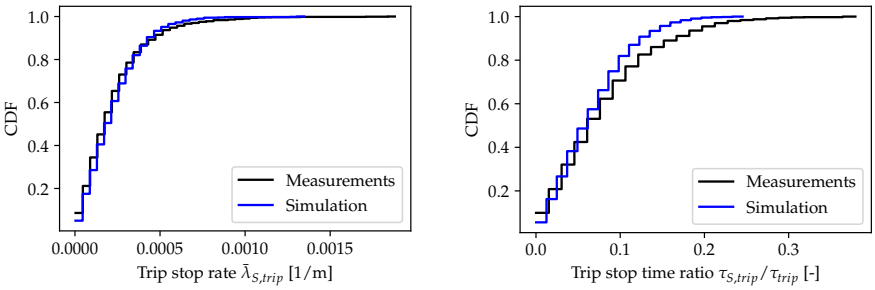
Table 3.3: Comparison of mean stop rate $\bar{\lambda}_{S,r_S}$ for different stop classes r_S estimated from measured and simulated velocities on the same routes

r_S	$\bar{\lambda}_{S,r_S,meas}$	$\bar{\lambda}_{S,r_S,sim}$	$\delta_{\bar{\lambda}_{S,r_S}}$
Traffic light	$1.18 \cdot 10^{-4}\,1/\mathrm{m}$	$1.18 \cdot 10^{-4}\,1/\mathrm{m}$	−0.43%
Yield sign	$1.89 \cdot 10^{-5}\,1/\mathrm{m}$	$2.42 \cdot 10^{-5}\,1/\mathrm{m}$	28.22%
Stop sign	$5.45 \cdot 10^{-6}\,1/\mathrm{m}$	$6.20 \cdot 10^{-6}\,1/\mathrm{m}$	13.76%
Pedestrian crossing	$1.15 \cdot 10^{-5}\,1/\mathrm{m}$	$1.33 \cdot 10^{-5}\,1/\mathrm{m}$	15.3%
Unclassified	$9.43 \cdot 10^{-5}\,1/\mathrm{m}$	$1.25 \cdot 10^{-4}\,1/\mathrm{m}$	32.38%
Motorway traffic	$8.067 \cdot 10^{-6}\,1/\mathrm{m}$	$1.59 \cdot 10^{-6}\,1/\mathrm{m}$	−80.31%

In Figure 3.39a the mean stop rate per trip

$$\bar{\lambda}_{S,trip} = \frac{n_{S,trip}}{l_{trip}} \tag{3.64}$$

with the number of stops per trip $n_{S,trip}$ and the trip length l_{trip}, is compared between simulation and measurement. Overall the CDF of the simulation shows good agreement with the measurements, only the very high mean stop rates per trip of $\bar{\lambda}_{S,trip} > 1.5 \cdot 10^{-3}\,1/\mathrm{m}$ occurring in some measured trips are not reproduced by the simulation. The trip stop time ratio $\tau_{S,trip}/\tau_{trip}$ compares the time spent in stops $\tau_{S,trip}$ with the total travel time of the trip τ_{trip}. As for the trip stop rate $\bar{\lambda}_{S,trip}$ the share of very high ratios $\tau_{S,trip}/\tau_{trip} > 0.2$ is underestimated. The derived stop models are based on mean values, e.g., the estimated stop probabilities at traffic lights only apply to the average traffic light passed, and the stop duration distributions are also not specific to individual locations. This property prevents the models from taking outliers, such as trips on routes with many stop reasons which have high stop probabilities and long mean stop durations, and therefore cause high ratios of $\bar{\lambda}_{S,trip}$ and $\tau_{S,trip}/\tau_{trip}$, into account. Thus, in the context of virtual load generation, where the individual trip in which a stop occurs is generally not relevant, the derived methods and their estimated parameters enable a realistic reproduction of the effects of stop events on the component loads.



(a) Trip stop rate $\bar{\lambda}_{S,trip}$ distribution

(b) Trip stop time ratio $\tau_{S,trip}/\tau_{trip}$ distribution

Figure 3.39: Comparison of stop event properties in measured and simulated velocity profiles (trip length > 20 km)

4 Uncertainty quantification for virtual load generation

To ensure the reliability of automotive components, it is essential to understand the expected operational loads. In the initial phases of development, virtual load generation helps in accurately estimating these loads. Given that the driven velocity significantly impacts many vehicle component loads, such as powertrain torque, vibration excitation, and temperature, velocity simulation is an essential component for creating these virtual loads. The stochastic velocity simulation described in Chapter 3 simulates velocities on a representative set of routes for different drivers and vehicles. Using the simulated velocities and other input quantities, the component loads are derived using system simulations specific to the vehicle, component, and load.

As described in Chapter 3, the velocity simulation [85, 113] is a stochastic simulator $\mathcal{M}_v(s, \boldsymbol{r}(s), \boldsymbol{\theta}_v)$ and considers various random influences to generate random velocity profiles for a given route. The simulation generates a stochastic process

$$\{V_s = \mathcal{M}_v(s, \boldsymbol{r}(s), \boldsymbol{\theta}_v) : s \in S\} \tag{4.1}$$

where s is the position along a route composed of the positions in the set S. Moreover, we consider different properties $\boldsymbol{r}(s)$, such as the legal speed limit, curvature, and slope of the road as a function of the position s along the route as well as the parameter vector $\boldsymbol{\theta}_v$ describing driver behavior and vehicle properties. To assess component loads the simulated velocities are fed into a system simulation (i.e., a transfer function) $\mathcal{M}_g(v(s), \boldsymbol{p}(s), \boldsymbol{\theta}_g)$ which simulates component loads based on the velocity as well as other inputs $\boldsymbol{p}(s)$ and parameters $\boldsymbol{\theta}_g$. Due to the random velocity profiles, the component load is thus also described by a stochastic process

$$\{T_s = \mathcal{M}_g(V_s, \boldsymbol{p}(s), \boldsymbol{\theta}_g) : s \in S\}. \tag{4.2}$$

This load is associated with uncertainties from the velocity and system simulation. The velocity simulation contributes to the load estimation uncertainty due to errors in the simulation model and its parameters, which can be classified as epistemic uncertainty and the inherent randomness of the simulation that contributes to the aleatoric uncertainty. The uncertainties of the system simulation depend on the specific application and must either be estimated based on expert knowledge or inferred from load measurements. As these measurements are often not available for specific applications and expert knowledge can be highly subjective, uncertainties in the system simulation are not considered in the following sections.

Section 4.1 describes a comprehensive UQ framework to handle uncertainties of the velocity simulation in the context of virtual load generation. This framework includes a DoE, surrogate modeling, sensitivity analysis, parameter inference, and evaluation of the results considering the uncertainties. Based on the UQ framework, two applications are demonstrated in Sections 4.2 and 4.3. The first derives uncertainties of the velocity simulation specific to the system simulation used. The second application derives more generic uncertainties of the velocity simulation without explicitly considering a system simulation. Parts of the content of the following sections have already been published in [114].

4.1 Framework

The UQ framework provides the structure and connects the needed modules to identify uncertainties in the velocity simulation. An overview of the framework is given in Figure 4.1. The framework is executed for the considered driver and vehicle combinations with the associated sets of routes to identify the uncertainties in the parameter vector $\boldsymbol{\theta}_v$ of each driver. It is also possible to only execute parts of the framework, e.g., to only derive a surrogate model for use in other applications, or to only identify the most important parameters using sensitivity analysis.

The framework is based on QoIs, which describe aspects of the velocity and load time series. The QoIs are usually derived by splitting the series into windows of a certain length $s_{w,l}$ without overlap and then calculating statistics or application-specific metrics u for each window w leading to the QoI $q_{w,u}$. The reference samples $q_{w,u,ref}$ of the QoIs are either obtained directly from measured velocities or, if they are based on the load time series, by applying the system simulation to the measured velocities. Direct measurements of the load series are not used because the focus is on the uncertainty in the velocity simulation. Additionally, obtaining these measurements in large quantities

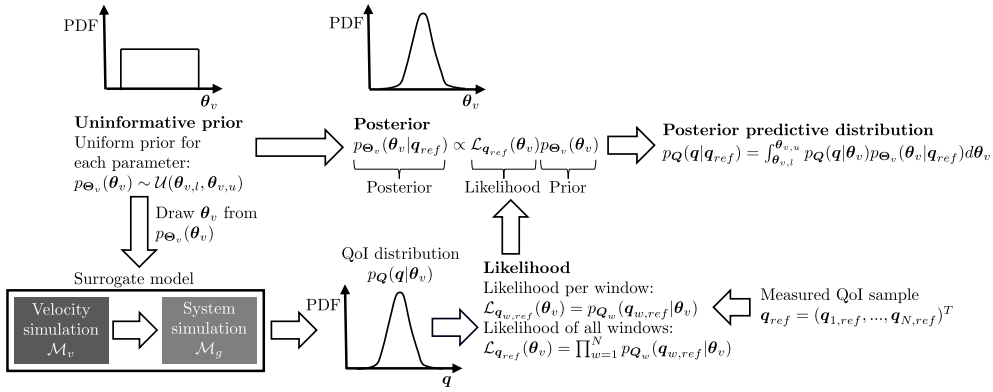


Figure 4.1: Overview of the UQ framework. Figure modified from [114].

is often challenging, as they typically require complex setups for load measurement on the component.

If there were a second set of velocity measurements from the same driver on the same set of routes, the QoIs per window would change as a result of random changes in driving behavior, traffic, etc. It is thus necessary to model the QoIs per window as random variables with associated PDF

$$Q_{w,u} \sim p_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_v) \tag{4.3}$$

which is conditional on the parameter vector $\boldsymbol{\theta}_v$ of the velocity simulation. For multiple windows with multiple QoIs

$$\boldsymbol{Q} = \begin{bmatrix} Q_{1,1} & \dots & Q_{1,P} \\ \vdots & \ddots & \vdots \\ Q_{N,1} & \dots & Q_{N,P} \end{bmatrix} \tag{4.4}$$

we obtain the joint distribution $p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_v)$. The distributions of the different QoIs in the same window $Q_{w,1}, ..., Q_{w,P}$ are generally correlated and cannot be assumed as independent. The properties of the correlated QoI distribution change for each window as the characteristics of the velocity profiles are typically different for each window, e.g., one window may cover a city road while the next one is on a motorway portion of the route. The distributions $\boldsymbol{Q}_w$ for different windows w are therefore not identically distributed but are assumed as independent as the considered windows do not overlap.

The velocity simulation parameters $\boldsymbol{\theta}_v$ are defined as a random variable $\boldsymbol{\Theta}_v$ in a Bayesian interpretation of probability, meaning that we aim to find the posterior distribution

$p_{\boldsymbol{\Theta}_v}(\boldsymbol{\theta}_v|\boldsymbol{q}_{ref})$ using the sample $\boldsymbol{q}_{ref}$ such that the true QoI distribution $\tilde{p}_{\boldsymbol{Q}}(\boldsymbol{q})$ is well approximated and the uncertainty in the simulated distribution $p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_v)$ is expressed through the uncertain parameters $\boldsymbol{\Theta}_v$.

4.1.1 Sampling and design of experiments

The joint QoI distribution $p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_v)$ is conditional on the parameters of the velocity simulation $\boldsymbol{\theta}_v$. To approximate the distribution $p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_v)$ and investigate the effects of the parameters $\boldsymbol{\theta}_v$ a sampling approach is needed.

For the uncertain velocity simulation parameters, a uniform, uninformative prior

$$p_{\boldsymbol{\Theta}_v}(\boldsymbol{\theta}_v) \sim \mathcal{U}(\boldsymbol{\theta}_{v,l}, \boldsymbol{\theta}_{v,u}) \tag{4.5}$$

is defined. The lower $\boldsymbol{\theta}_{v,l}$ and upper $\boldsymbol{\theta}_{v,u}$ bounds are set to threshold values estimated from a database with velocity simulation parameters from numerous drivers to ensure a physically plausible meaning for them. Table A.2 lists all parameters in the vector $\boldsymbol{\theta}_v$ and their bounds. Depending on the considered QoIs, it can be useful to exclude some parameters as their impact on the QoIs can be neglected a priori, which helps to reduce the computational effort in the further steps of the UQ framework.

Different parameter samples $\boldsymbol{\theta}_{v,1}, ..., \boldsymbol{\theta}_{v,M}$ are generated with a Monte Carlo sampling strategy such as LHS, see Section 2.3.1, from the prior distribution $p_{\boldsymbol{\Theta}_v}(\boldsymbol{\theta}_v)$. Using the stochastic velocity simulation $\mathcal{M}_v(s, \boldsymbol{r}(s), \boldsymbol{\theta}_v)$ and, if the QoI is based on the load series, the deterministic system simulation $\mathcal{M}_g(v(s), \boldsymbol{p}(s), \boldsymbol{\theta}_g)$ the parameter samples can be propagated to the QoIs. The parameters of the system simulation $\boldsymbol{\theta}_g$ are assumed to be known exactly and are not considered in the DoE. To approximate the distribution $p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_{v,i})$ for one parameter sample $\boldsymbol{\theta}_{v,i}$ repeated executions of the simulation are necessary to capture the randomness of the simulation. The QoI distribution can then be approximated by the repeated samples $\boldsymbol{Q}_1^{(i)}, ..., \boldsymbol{Q}_R^{(i)} \sim p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_{v,i})$. Based on the repeated samples, moments or other statistics of the distribution can be estimated conditional on the parameter vector $\boldsymbol{\theta}_{v,i}$.

Using repeated samples usually leads to unfeasible computing times, as for all M parameter samples, R repetitions need to be performed, which means the simulation needs to be executed $R \cdot M$ times. As described in Section 2.3.4, this is not necessary when using GLaM surrogate models. To fit those surrogates, only one repetition is necessary, meaning that for each parameter sample $\boldsymbol{\theta}_{v,i}$ only one QoI sample $\boldsymbol{Q}_i \sim p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_{v,i})$ is drawn, requiring only a single execution of the simulation. The resulting dataset of a DoE with M samples is thus given by the parameter samples $\{\boldsymbol{\theta}_{v,1}, ..., \boldsymbol{\theta}_{v,M}\}$ and the

associated QoI samples $\{q_{w,u,1}, ..., q_{w,u,M}\}$ for each window w and QoI u. Since the parameter samples are drawn from the prior distribution, the resulting QoI samples are draws from the prior predictive distribution and can therefore be used to approximate it.

4.1.2 Likelihood function

The likelihood is calculated with the measured sample

$$\boldsymbol{q}_{ref} = \begin{bmatrix} q_{1,1,ref} & \cdots & q_{1,P,ref} \\ \vdots & \ddots & \vdots \\ q_{N,1,ref} & \cdots & q_{N,P,ref} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_{1,ref} \\ \vdots \\ \boldsymbol{q}_{N,ref} \end{bmatrix} \tag{4.6}$$

which contains the measurements $q_{w,u,ref}$ for each QoI u and window w which are samples from the true, unknown distribution $\tilde{p}_Q(\boldsymbol{q})$. Based on the simulated QoI distribution per window $p_{Q_w}(\boldsymbol{q}_w|\boldsymbol{\theta}_v)$ we define the likelihood of the measurements per window $\boldsymbol{q}_{w,ref}$ by

$$\mathcal{L}_{\boldsymbol{q}_{w,ref}}(\boldsymbol{\theta}_v) = p_{Q_w}(\boldsymbol{q}_{w,ref}|\boldsymbol{\theta}_v). \tag{4.7}$$

Due to the assumed independence of the windows, the joint likelihood of all windows is given by their product

$$\mathcal{L}_{\boldsymbol{q}_{ref}}(\boldsymbol{\theta}_v) = \prod_{w=1}^{N} p_{Q_w}(\boldsymbol{q}_{w,ref}|\boldsymbol{\theta}_v), \tag{4.8}$$

and the log-likelihood follows in the same manner as the sum of their logarithms

$$\ell_{\boldsymbol{q}_{ref}}(\boldsymbol{\theta}_v) = \sum_{w=1}^{N} \log\Big(p_{Q_w}(\boldsymbol{q}_{w,ref}|\boldsymbol{\theta}_v)\Big). \tag{4.9}$$

4.1.3 Surrogate model generation

A surrogate model approximates a simulation model to reduce the number of costly evaluations of the full simulation. As the used velocity simulation in combination with the system simulations is a stochastic simulator, this requires a stochastic surrogate model which allows for faster evaluation of the distribution $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_v)$ by the approximate surrogate distribution $\hat{p}_Q(\boldsymbol{q}|\boldsymbol{\theta}_v)$ [109]. Different stochastic surrogate models are

described in [62, 110, 112]. For our problem, we choose the GLaM [110] which is built on the GLD [30], as described in Section 2.3.4.

As the GLaM in its used form can only be used to approximate distributions in one dimension, a separate surrogate model is built for the distribution $\hat{p}_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_v)$ of each window and QoI. The lambda parameters of the GLD are given as a function of the velocity simulation parameters $\boldsymbol{\theta}_v$ specific to each QoI and window, and thus the GLaM for each QoI and window gives the approximate distribution as a function of $\boldsymbol{\theta}_v$

$$\hat{Q}_{w,u}(\boldsymbol{\theta}_v) \sim \hat{p}_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_v) = \mathrm{GLD}(\lambda_{1,w,u}(\boldsymbol{\theta}_v), \lambda_{2,w,u}(\boldsymbol{\theta}_v), \lambda_{3,w,u}(\boldsymbol{\theta}_v), \lambda_{4,w,u}(\boldsymbol{\theta}_v)). \quad (4.10)$$

The GLaMs are build on the experimental design described in Section 4.1.1 without replications, i.e., for each sample from the parameter distribution $p_{\boldsymbol{\Theta}_v}(\boldsymbol{\theta}_v)$ one sample from the conditional QoI distribution $p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_v)$ is drawn. Using the MLE procedure described in [110], the PCE coefficients $c_{l,\boldsymbol{\alpha}}$ for the lambda parameters are fitted. The polynomial degrees for the four lambdas and the truncation scheme are chosen depending on the QoI but are constant for all windows of that QoI. It can be beneficial for some applications to fix the shape parameters of the GLD, e.g., $\lambda_3 = 0.13$ and $\lambda_4 = 0.13$ for the approximate shape of a normal distribution, if the shape of the QoI distribution is already known a priori or to regularize the model and make it less prone to overfitting [16].

In case of multiple QoIs per window, their correlations need to be considered. The different distributions $\hat{p}_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_v)$ approximated by the GLaMs are independent and do not consider the existing correlations. To consider the correlations, a multidimensional normal distribution

$$\hat{\boldsymbol{Q}}_w = (\hat{Q}_{w,1}, \ldots, \hat{Q}_{w,P}) \sim \mathcal{N}(\boldsymbol{\mu}_w|\boldsymbol{\theta}_v, \boldsymbol{\Sigma}_w|\boldsymbol{\theta}_v). \quad (4.11)$$

of the QoIs is assumed. The conditional mean vector $\boldsymbol{\mu}_w|\boldsymbol{\theta}_v$ and diagonal of the covariance matrix $\boldsymbol{\Sigma}_w|\boldsymbol{\theta}_v$ are given by the individual means and variances of the GLaMs

$$\begin{aligned} \boldsymbol{\mu}_w|\boldsymbol{\theta}_v &= (\mathbb{E}[\hat{Q}_{w,1}|\boldsymbol{\theta}_v], \ldots, \mathbb{E}[\hat{Q}_{w,P}|\boldsymbol{\theta}_v])^T, \\ \mathrm{diag}(\boldsymbol{\Sigma}_w|\boldsymbol{\theta}_v) &= (\sigma^2(\hat{Q}_{w,1}|\boldsymbol{\theta}_v), \ldots, \sigma^2(\hat{Q}_{w,P}|\boldsymbol{\theta}_v))^T. \end{aligned} \quad (4.12)$$

To complete the covariance matrix the correlations $\rho_{Q_{w,u},Q_{w,h}}$ between two QoIs $Q_{w,u}$ and $Q_{w,h}$ must be determined. To simplify the model we assume that the correlations

$\rho_{Q_{w,u},Q_{w,h}}$ are constant and independent from the parameters $\boldsymbol{\theta}_v$ which allows to estimate them from the DoE dataset using the sample correlation coefficient

$$\hat{\rho}_{Q_{w,u},Q_{w,h}} = \frac{\sum_{i=1}^{M}(q_{w,u,i} - \bar{q}_{w,u})(q_{w,h,i} - \bar{q}_{w,h})}{\sqrt{\sum_{i=1}^{M}(q_{w,u,i} - \bar{q}_{w,u})^2 \sum_{i=1}^{M}(q_{w,h,i} - \bar{q}_{w,h})^2}} \tag{4.13}$$

with the samples $q_{w,u,i}$ and $q_{w,h,i}$ from the prior predictive distribution. From the correlations and standard deviations, follow the off-diagonal entries of the covariance matrix

$$\mathrm{cov}(\hat{Q}_{w,u}, \hat{Q}_{w,h}) = \hat{\rho}_{Q_{w,u},Q_{w,h}} \sigma(\hat{Q}_{w,u}|\boldsymbol{\theta}_v)\sigma(\hat{Q}_{w,h}|\boldsymbol{\theta}_v) \tag{4.14}$$

thereby completely defining the entries of the covariance matrix $\boldsymbol{\Sigma}_w|\boldsymbol{\theta}_v$ conditional on the parameter vector $\boldsymbol{\theta}_v$.

4.1.4 Surrogate model validation

For validation, a second DoE with M_{val} samples is done where for each parameter sample $\boldsymbol{\theta}_{v,i}$ we run R repetitions to obtain the QoI samples $\boldsymbol{q}_1^{(i)}, ..., \boldsymbol{q}_R^{(i)}$ to approximate the true distribution $p_{\boldsymbol{Q}}(\boldsymbol{q}|\boldsymbol{\theta}_v)$.

Between the GLaM distribution $\hat{p}_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_{v,i})$ and the distribution of the full simulation $p_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_{v,i})$ the empirical Wasserstein distance $d_{WS}(Q_{w,u}(\boldsymbol{\theta}_{v,i}), \hat{Q}_{w,u}(\boldsymbol{\theta}_{v,i}))$ (e.g. [51]), is calculated for each parameter sample i, QoI u, and window w. The empirical Wasserstein distance is computed with the R samples obtained from $p_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_{v,i})$ through repeated sampling and an equal amount of samples drawn from the GLaM distribution $\hat{p}_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_{v,i})$. Each computed Wasserstein distance is normalized by the standard deviation of the full simulation distribution $\sigma(Q_{w,u}(\boldsymbol{\theta}_{v,i}))$, which gives the normalized Wasserstein distance

$$d(Q_{w,u}(\boldsymbol{\theta}_{v,i}), \hat{Q}_{w,u}(\boldsymbol{\theta}_{v,i})) = \frac{d_{WS}(Q_{w,u}(\boldsymbol{\theta}_{v,i}), \hat{Q}_{w,u}(\boldsymbol{\theta}_{v,i}))}{\sigma(Q_{w,u}(\boldsymbol{\theta}_{v,i}))}. \tag{4.15}$$

The standard deviation $\sigma(Q_{w,u}(\boldsymbol{\theta}_{v,i}))$ is identical to the Wasserstein distance between the distribution $p_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_{v,i})$ and the mean $\mathbb{E}[Q_{w,u}(\boldsymbol{\theta}_{v,i})]$, see [110] for more details. The normalized Wasserstein distance can thus be interpreted as the ratio between the Wasserstein distances $d_{WS}(Q_{w,u}(\boldsymbol{\theta}_{v,i}), \hat{Q}_{w,u}(\boldsymbol{\theta}_{v,i}))$ and $d_{WS}(Q_{w,u}(\boldsymbol{\theta}_{v,i}), \mathbb{E}[Q_{w,u}(\boldsymbol{\theta}_{v,i})])$, i.e., how much improvement is obtained by using the stochastic surrogate distribution $\hat{p}_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_{v,i})$ compared to a degenerate distribution at the true mean of the full simulation.

For each window and QoI, the mean of the normalized Wasserstein distance for all parameter samples $\boldsymbol{\theta}_{v,i} \sim p_{\boldsymbol{\Theta}_v}(\boldsymbol{\theta}_v)$ in the validation dataset is taken to derive the error measure per window and QoI

$$\epsilon_{w,u} = \mathbb{E}[d(Q_{w,u}(\boldsymbol{\Theta}_v), \hat{Q}_{w,u}(\boldsymbol{\Theta}_v))] \approx \frac{1}{M_{val}} \sum_{i=1}^{M_{val}} d(Q_{w,u}(\boldsymbol{\theta}_{v,i}), \hat{Q}_{w,u}(\boldsymbol{\theta}_{v,i})). \tag{4.16}$$

A second metric for surrogate model validation is the accuracy with which the log-likelihood of a measurement (see Section 4.1.2) can be predicted with the surrogate model. To calculate the log-likelihood, a reference sample is defined as the mean of the prior predictive distribution (see also (4.21) for the posterior predictive distribution)

$$\boldsymbol{q}_{ref,Pr} = \mathbb{E}[\boldsymbol{Q}(\boldsymbol{\Theta}_v)] \text{ with } p_{\boldsymbol{\Theta}_v}(\boldsymbol{\theta}_v) \sim \mathcal{U}(\boldsymbol{\theta}_{v,l}, \boldsymbol{\theta}_{v,u}) \tag{4.17}$$

which serves as a representative measurement for which the log-likelihood can be calculated and compared between the GLaM and reference distributions. The log-likelihood of the reference sample using the GLaM distribution $\hat{p}_Q(\boldsymbol{q}|\boldsymbol{\theta}_v)$ is thus calculated according to (4.9). To calculate the log-likelihood with the reference distribution $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_v)$ of the full simulation, a Kernel Density Estimation (KDE) is applied to the repeated samples from $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_v)$ to approximate the PDF of the distribution. The log-likelihoods of the GLaM $\hat{\ell}_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_{v,i})$ and of the reference $\ell_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_{v,i})$ are evaluated for all parameter samples $\boldsymbol{\theta}_{v,i}$ based on which the coefficient of determination R^2 is calculated. This metric allows for evaluating the quality of the surrogate with a single number that combines all QoIs, windows, and parameter samples. As the metric is based on the log-likelihood, it is particularly meaningful to evaluate the surrogate in the context of likelihood based parameter inference, e.g., to calculate the posterior distribution in a Bayesian framework.

4.1.5 Sensitivity analysis

Based on the derived GLaM surrogates, a global sensitivity analysis using Sobol' indices [90], see Section 2.3.3, of the input parameters is conducted. In [111], different approaches to global sensitivity analysis of stochastic simulators are described. We focus on sensitivity analysis based on deterministic QoIs derived from the QoI distribution $\hat{p}_Q(\boldsymbol{q}|\boldsymbol{\theta}_v)$.

For each QoI u and window w the Sobol' indices of the mean $\mathbb{E}[\hat{Q}_{w,u}|\boldsymbol{\theta}_v]$ and standard deviation $\sigma(\hat{Q}_{w,u}|\boldsymbol{\theta}_v)$ are estimated by Monte Carlo simulations [82, 83] using the respective GLaM. Mean and standard deviation are directly obtained from the description of

the GLD in the GLaM. The parameter space for the Monte Carlo simulations is defined by the prior $p_{\boldsymbol{\Theta}_v}(\boldsymbol{\theta}_v)$ from (4.5), where $M_{Sensi} = 10^4$ samples are drawn. To analyze the overall sensitivity of the QoIs to the parameters $\theta_{v,j}$ in $\boldsymbol{\theta}_v = (\theta_{v,1}, ..., \theta_{v,K})^T$ the mean of the first and total order Sobol' indices

$$\begin{aligned} \bar{S}_{j,u} &= \frac{1}{N} \sum_{w=1}^{N} S_{j,w,u}, \\ \bar{S}_{Tj,u} &= \frac{1}{N} \sum_{w=1}^{N} S_{Tj,w,u}, \end{aligned} \tag{4.18}$$

with respect to the N windows are defined. Mean Sobol' indices of other orders follow analogously.

To globally analyze the importance of the parameters, taking into account all windows and QoIs, the sensitivity of the GLaM log-likelihood $\hat{\ell}_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_v)$ to the velocity simulation parameters in $\boldsymbol{\theta}_v$ is calculated. The log-likelihood is computed with the reference sample $\boldsymbol{q}_{ref,Pr}$ defined as the mean of the prior predictive distribution, see (4.17). The Sobol' indices thus indicate how sensitive the log-likelihood of the reference sample is to changes in the parameters.

4.1.6 Bayesian parameter inference

Using the results of the sensitivity analysis, see Section 4.1.5, the velocity simulation parameter vector is reduced to the parameters in $\mathcal{A}_v$ which have a significant influence on the QoIs and the derived quantities such as the log-likelihood $\hat{\ell}_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_v)$. These parameters $\boldsymbol{\theta}_v^* = (\theta_{v,j} : j \in \mathrm{A}_v)^T$ are modeled as a random variable $\boldsymbol{\Theta}_v^*$, while all other parameters $(\theta_{v,j} : j \notin \mathrm{A}_v)^T$ are set to their default values in the driver-specific default parameter vector $\boldsymbol{\theta}_{v,0}$. The default values are identified with statistics from the measured velocities of the drivers, e.g., the parameters of the velocity fluctuations, see Section 3.2, or taken as a guess value from experience. The aim of Bayesian parameter inference, see also Section 2.3.5, is to obtain the posterior distribution of $\boldsymbol{\Theta}_v^*$ through Bayes' rule given by

$$p_{\boldsymbol{\Theta}_v^*}(\boldsymbol{\theta}_v^* | \boldsymbol{q}_{ref}) \propto \hat{\mathcal{L}}_{\boldsymbol{q}_{ref}}(\boldsymbol{\theta}_v^*) p_{\boldsymbol{\Theta}_v^*}(\boldsymbol{\theta}_v^*). \tag{4.19}$$

with the measured QoI sample $\boldsymbol{q}_{ref}$ and the likelihood $\hat{\mathcal{L}}_{\boldsymbol{q}_{ref}}(\boldsymbol{\theta}_v^*)$ given by the GLaM. To sample from the posterior distribution in (4.19), the MCMC algorithm with slice sampling (e.g. [74]), see Section 2.3.2, is employed. By drawing a sufficiently large

number of samples, the posterior distribution $p_{\Theta_v^*}(\boldsymbol{\theta}_v^* | \boldsymbol{q}_{ref})$ can be approximated by a KDE.

The posterior distribution contains all available information on the estimated velocity simulation parameters, including the uncertainty of the estimate, and can be used as an input to the velocity simulation to propagate the uncertainty to the QoIs at the output of the simulation. In some cases, it can be useful to reduce the parameter estimate to a single number, i.e., a point estimate, which is easier to handle. In a Bayesian framework, typical estimators are the MAP estimator

$$\hat{\boldsymbol{\theta}}_v^* = \arg\max_{\boldsymbol{\theta}_v^*} p_{\Theta_v^*}(\boldsymbol{\theta}_v^* | \boldsymbol{q}_{ref}) \tag{4.20}$$

or the posterior expectation $\hat{\boldsymbol{\theta}}_v^* = \mathbb{E}[\boldsymbol{\Theta}_v^* | \boldsymbol{q}_{ref}]$, which both can be plugged into the velocity simulation without any modifications to the simulation chain.

4.1.7 Uncertainty propagation

The identified posterior distribution is propagated to the QoIs by evaluating the posterior predictive distribution, which, according to Section 2.3.5, is given by

$$p_Q(\boldsymbol{q} | \boldsymbol{q}_{ref}) = \int_{\boldsymbol{\theta}_{v,l}^*}^{\boldsymbol{\theta}_{v,u}^*} p_Q(\boldsymbol{q} | \boldsymbol{\theta}_v^*) p_{\Theta_v^*}(\boldsymbol{\theta}_v^* | \boldsymbol{q}_{ref}) d\boldsymbol{\theta}_v^*. \tag{4.21}$$

Through the posterior predictive distribution, the identified uncertainty of the velocity simulation parameters $p_{\Theta_v^*}(\boldsymbol{\theta}_v^* | \boldsymbol{q}_{ref})$ is combined with the aleatoric uncertainty of the QoIs $p_Q(\boldsymbol{q} | \boldsymbol{\theta}_v^*)$ caused by the stochastic velocity simulation. The integral in (4.21) is evaluated through Monte Carlo sampling which leads to a KDE approximation of the joint posterior predictive distribution $p_Q(\boldsymbol{q} | \boldsymbol{q}_{ref})$ and its marginal distributions $p_{Q_{w,u}}(q_{w,u} | \boldsymbol{q}_{ref})$ per window w and QoI u. All simulation parameters not included in $\boldsymbol{\theta}_v^*$ are set to their default values in $\boldsymbol{\theta}_{v,0}$. Depending on the use-case, the computational effort of the respective system simulations, and the available runtime, the integral in (4.21) can be evaluated by using the surrogate model with the approximate distribution $\hat{p}_Q(\boldsymbol{q} | \boldsymbol{\theta}_v^*)$ or using the distribution $p_Q(\boldsymbol{q} | \boldsymbol{\theta}_v^*)$ obtained by Monte Carlo sampling with the full simulation. Using the full simulation has the advantage of not introducing additional uncertainty through the surrogate model while having the disadvantage of requiring more runtime.

If the evaluation of the posterior predictive distribution in (4.21) is not feasible or the information about the parameter uncertainty is not required, one of the point estimates $\hat{\boldsymbol{\theta}}_v^*$ described in Section 4.1.6, is plugged into the GLaM to directly obtain the QoI

distribution $\hat{p}_Q(\boldsymbol{q}|\hat{\boldsymbol{\theta}}_v^*)$, or into the full simulation to obtain the distribution $p_Q(\boldsymbol{q}|\hat{\boldsymbol{\theta}}_v^*)$ through Monte Carlo sampling and KDE approximation.

Based on the posterior predictive distribution, further analyses can be conducted, which include the calculation of credible intervals, means, standard deviations, and other statistics from the posterior or posterior predictive distributions. The credible intervals in the following are always defined as median credible intervals given by the quantiles $[q_{(1-\gamma)/2}, q_{(1+\gamma)/2}]$ such that γ is the probability of a value being in the interval.

Other analyses are based on the distribution of a QoI u across multiple windows within a sample $\boldsymbol{q}_{u,i}$ drawn from the posterior predictive distribution $p_{Q_u}(\boldsymbol{q}_u|\boldsymbol{q}_{ref})$ or if uncertainties of the parameters are not considered from the distribution $p_{Q_u}(\boldsymbol{q}_u|\boldsymbol{\theta}_v)$ where, e.g., the default parameter vector $\boldsymbol{\theta}_{v,0}$ or the point estimate $\hat{\boldsymbol{\theta}}_v^*$ are plugged in for $\boldsymbol{\theta}_v$. A sample from the posterior predictive distribution is denoted as $\boldsymbol{q}_{u,i}(\boldsymbol{q}_{ref})$, while samples from the distribution $p_{Q_u}(\boldsymbol{q}_u|\boldsymbol{\theta}_v)$ are denoted as $\boldsymbol{q}_{u,i}(\boldsymbol{\theta}_v)$, respectively $\boldsymbol{q}_{u,i}(\boldsymbol{\theta}_{v,0})$ and $\boldsymbol{q}_{u,i}(\hat{\boldsymbol{\theta}}_v^*)$.

From the sample $\boldsymbol{q}_{u,i}$ random windows w are drawn which defines the random variable $\tilde{Q}_u$ with distribution $p_{\tilde{Q}_u}(\tilde{q}_u|\boldsymbol{q}_{u,i})$ conditional on $\boldsymbol{q}_{u,i}$. The samples $\boldsymbol{q}_{u,i}$ are realizations of the QoI $\boldsymbol{Q}_u$ for each window when the driver drives once across all considered routes and the derived windows. The distribution $p_{\tilde{Q}_u}(\tilde{q}_u|\boldsymbol{q}_{u,i})$ thus gives the probability of the QoI reaching a certain value in any of the windows, i.e., in any part of the considered routes, which can be compared to the distribution of the measured QoI values $p_{\tilde{Q}_u}(\tilde{q}_u|\boldsymbol{q}_{u,ref})$ on the same routes. By drawing multiple samples $\boldsymbol{q}_{u,i}$, to consider the parametric uncertainty and the randomness of the velocity simulation, multiple distributions $p_{\tilde{Q}_u}(\tilde{q}_u|\boldsymbol{q}_{u,i})$ can be derived. Thereby allowing the construction of Probability-boxes (p-boxes) which give upper and lower bounds on the probability distribution of $\tilde{Q}_u$. The available windows $w \in \{1, ..., N\}$ are a sample from the set of all windows of all routes driven during the lifespan of a vehicle. The uncertainty regarding the measured sample on the available window set with the distribution $p_{\tilde{Q}_u}(\tilde{q}_u|\boldsymbol{q}_{u,ref})$ is thus reducible through more measurements, i.e., it is epistemic. This uncertainty is captured by the width of the p-box which not only captures the epistemic uncertainty of the parameters but also the epistemic uncertainty caused by the size of the route sample which leads to larger differences in the samples $\boldsymbol{q}_{u,i}$ for shorter route sets with fewer windows.

In many virtual load generation use-cases, cumulative damage sums are the relevant QoI. Therefore the mean

$$\bar{\tilde{q}}_{u,i} = \mathbb{E}[\tilde{q}_u|\boldsymbol{q}_{u,i}] = \frac{1}{N}\sum_{w=1}^{N} q_{w,u,i} \tag{4.22}$$

or the sum of the QoI values per window, derived from the samples $\boldsymbol{q}_{u,i}$, is of particular importance. Considering multiple samples of $\boldsymbol{q}_{u,i}$ leads to an approximation of the mean distribution $p_{\bar{\tilde{Q}}_u}(\bar{\tilde{q}}_u)$ which considers the parametric uncertainty and the randomness of the velocity simulation in combination with the uncertainty caused by the length of the chosen route set and can be compared to the measured mean $\bar{\tilde{q}}_{u,ref}$. Here, mean distributions derived from the posterior predictive are denoted as $p_{\bar{\tilde{Q}}_u}(\bar{\tilde{q}}_u|\boldsymbol{q}_{ref})$, while distributions derived from $p_{Q_u}(\boldsymbol{q}_u|\boldsymbol{\theta}_v)$ are denoted as $p_{\bar{\tilde{Q}}_u}(\bar{\tilde{q}}_u|\boldsymbol{\theta}_v)$.

4.2 System simulation specific Uncertainty Quantification

The UQ framework described in Section 4.1 is applied to a specific use-case with a fixed system simulation and the associated QoI. The derived surrogates, sensitivities, and parameter uncertainties are therefore specific to the chosen QoI. This procedure leads to results tailored to the specific application, i.e., the identified parameters and their uncertainties are adapted to the used QoI, but the results are not transferable to other QoIs.

4.2.1 Simulation setup

The considered QoI is given by the torque load on the drivetrain of an electric vehicle. However, we note that the developed framework can be applied to any other load case, and as described in Section 4.1, also to multiple loads and their QoIs at once. The system simulation $\mathcal{M}_g(v(s), \boldsymbol{p}(s), \boldsymbol{\theta}_g)$ is a model of the vehicle and its drivetrain, which, under consideration of the vehicle dynamics, outputs the respective torque in the drivetrain component. The inputs $\boldsymbol{p}(s)$ include the slope of the road given by the map data, while the parameter vector $\boldsymbol{\theta}_g$ defines the properties of the vehicle. Unlike the parameters of the velocity simulation $\boldsymbol{\theta}_v$ the parameters of the system simulation are modeled without uncertainty. If their uncertainty is known from other analyses, it can be included in addition to the velocity simulation parameter uncertainty in the uncertainty propagation step, see Section 4.1.7. The QoI is based on the normalized gear damage sum $d'_{G,w}$ per window described in Section 3.2.7, with window length $s_{w,l} = 5000\,\mathrm{m}$ and shift $s_{w,s} = s_{w,l}$. Due to the exponential form of the damage sum, its logarithm

$$\begin{aligned} q_{w,d_G} &= \log\left(d'_{G,w}\right), \\ \boldsymbol{q}_{d_G} &= (q_1, \ldots, q_N)^T, \end{aligned} \tag{4.23}$$

is considered as the QoI, respectively, the derived random variable $\boldsymbol{Q}_{d_G}$ to consider the stochasticity of the velocity profiles. In the following we drop the index d_G as only a single QoI is considered, i.e., $\boldsymbol{Q} = \boldsymbol{Q}_{d_G}$. The reference measurement $\boldsymbol{q}_{ref}$ is obtained by applying the torque simulation $\mathcal{M}_g(v(s), \boldsymbol{p}(s), \boldsymbol{\theta}_g)$ to the measured velocity profiles while using the same road properties $\boldsymbol{p}(s)$ and parameters $\boldsymbol{\theta}_g$ as for the simulated velocities. The schematic simulation chains to derive the QoI using simulated and measured velocities are depicted in Figure 4.2.

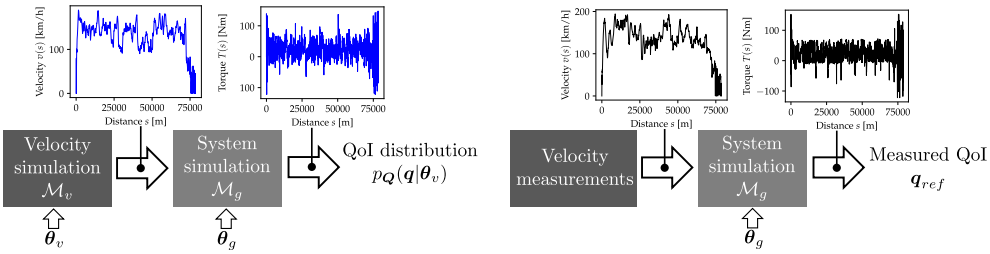


Figure 4.2: Setup of the torque simulation with simulated and measured velocities

4.2.2 Numerical results

The simulation setup described in Section 4.2.1 is applied to 18 different drivers, which all drive on the same 216 km long route set, where for each driver a set of velocity measurements on the route set is available for analysis. The subsequent steps of the UQ framework are applied to all drivers, leading to a different posterior distribution of the velocity simulation parameters for each driver. The procedure and its results are presented below for one of the drivers as an example, although the results for the surrogate model and the sensitivity analysis are identical for the different drivers, as the used prior distribution and route set are the same for all drivers. Finally, for all drivers, the aggregate results of the identified posterior distributions and the resulting mean distributions $p_{\tilde{\tilde{Q}}}(\tilde{\tilde{q}}|\boldsymbol{q}_{ref})$ are compared to the default parameters and the mean distributions $p_{\tilde{\tilde{Q}}}(\tilde{\tilde{q}}|\boldsymbol{\theta}_{v,0})$ obtained with them.

Surrogate model generation

A DoE with $M = 1\,000$ samples without replications is run with the velocity and torque simulation described in Section 4.2.1. The velocity simulation parameter samples $\boldsymbol{\theta}_{v,1}, ..., \boldsymbol{\theta}_{v,M}$ are drawn from a uniform prior distribution, see (4.5), and Table A.2 for the bounds on the distribution. Considering all 36 velocity simulation parameters in

Table A.2 would be unfeasible due to the very high number of required parameter samples for a sufficient surrogate model fit. Using expert knowledge, the parameter space can be reduced to 9 parameters which have a potential influence on the torque profiles and thus the gear damage sum based QoI, see (4.23), while all other parameters are assumed to be irrelevant with respect to the considered QoI. The considered parameter vector $\boldsymbol{\theta}_v$ is therefore reduced to those 9 parameters

$$\boldsymbol{\theta}_v = (\mu_{a_{x,max,dr}}, \mu_{a_{x,min,dr}}, o_{v,ci}, o_{v,co}, o_{v,mo}, \mu_{\sigma_{\zeta,log},ci}, \mu_{\sigma_{\zeta,log},co}, \mu_{\sigma_{\zeta,log},mo}, \mu_{v_{max,dr}})^T \quad (4.24)$$

which are modeled as a random variable $\boldsymbol{\Theta}_v$ while all other parameters are set to their default values in $\boldsymbol{\theta}_{v,0}$. The reduction to 9 parameters helps to limit the required number of samples to fit a surrogate model with sufficient accuracy, and therefore brings the computational effort to an acceptable level.

The velocity simulation, and subsequently the torque simulation, is run on the route set for each of the DoE parameter samples. The resulting torque profiles are split into $N = 41$ windows, where for each window the QoI in (4.23) is calculated, leading to the QoI vector $\boldsymbol{q} = (q_1, ..., q_N)^T$. The resulting parameter sample set $\{\boldsymbol{\theta}_{v,1}, ..., \boldsymbol{\theta}_{v,M}\}$ and the associated QoI samples $\{\boldsymbol{q}_1, ..., \boldsymbol{q}_M\}$ are used to fit a GLaM for each of the QoI distributions per window $\hat{p}_{Q_w}(q_w|\boldsymbol{\theta}_v)$ with the methods described in Sections 2.3.4 and 4.1.3. The polynomial degrees of the PCE of the four lambdas of the GLaM are set to $p_1 = 2, p_2 = p_3 = p_4 = 1$ while $q_1 = q_2 = q_3 = q_4 = 1$ for the full polynomial basis in the truncation sets $\{\mathcal{A}_l : l = 1, ..., 4\}$.

Surrogate model validation

The second DoE for validation, see Section 4.1.4, is run with $M_{val} = 100$ parameter samples and $R = 1\,000$ repetitions for each parameter sample. Figure 4.3 compares how the mean μ_w and the 50%, 75%, and 90% credible intervals of the distributions $p_{Q_w}(q_w|\boldsymbol{\theta}_v)$ and $\hat{p}_{Q_w}(q_w|\boldsymbol{\theta}_v)$ differ for a fixed value of the parameter vector $\boldsymbol{\theta}_v$ along the different windows of one of the routes. The GLaM distributions show some deviations from the reference distributions, but overall, the differences in mean and variance between the windows are well reproduced.

Between the GLaM $\hat{p}_{Q_w}(q_w|\boldsymbol{\theta}_v)$ and reference $p_{Q_w}(q_w|\boldsymbol{\theta}_v)$ distributions the normalized Wasserstein distance according to (4.15) and subsequently the error measure per window ϵ_w, see (4.16), is calculated. Figure 4.4 depicts the distribution of the error measure ϵ_w

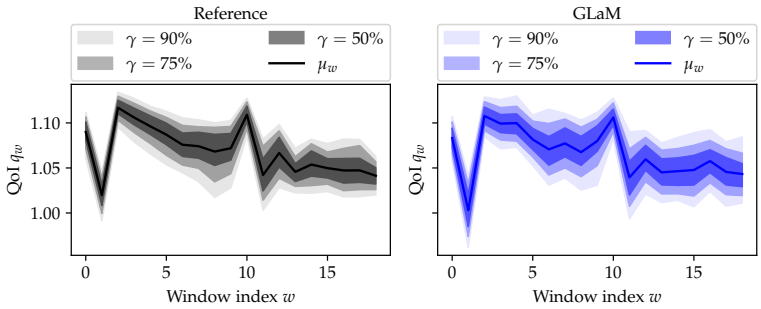


Figure 4.3: Comparison of reference $p_{Q_w}(q_w|\boldsymbol{\theta}_v)$ and GLaM $\hat{p}_{Q_w}(q_w|\boldsymbol{\theta}_v)$ distributions with a fixed parameter vector $\boldsymbol{\theta}_v$ for different windows along a route. The center line is the mean μ_w and the different color shades indicate the 50%, 75%, and 90% credible intervals of the distributions. Figure modified from [114].

across all windows. For most windows, an error $\epsilon_w < 0.5$ is observed, where the mean error of all windows is

$$\bar{\epsilon} = \frac{1}{N}\sum_{w=1}^{N} \epsilon_w = 0.55. \tag{4.25}$$

When the GLaM is slightly modified by assuming normality of the distributions $\hat{p}_{Q_w}(q_w|\boldsymbol{\theta}_v)$, i.e., $\lambda_3 = \lambda_4 = 0.13$, the mean error measure increases slightly to $\bar{\epsilon} = 0.58$.

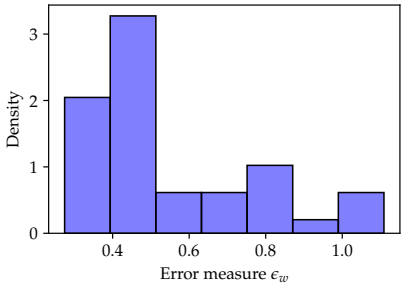
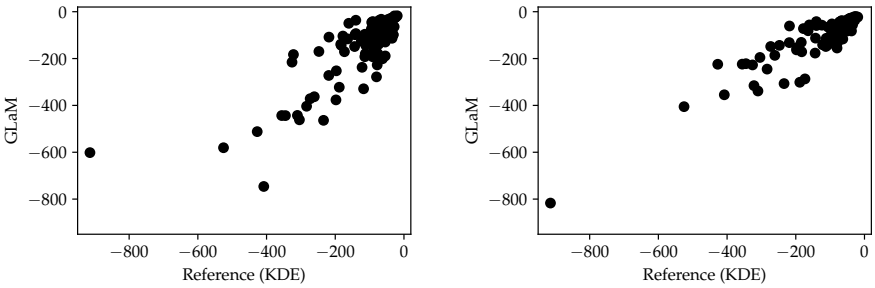


Figure 4.4: Distribution of the error measure ϵ_w. Figure modified from [114].

Using the reference measurement derived from the prior, see (4.17), the log-likelihood is calculated according to (4.9) with the reference and GLaM distributions for each of the $M_{val} = 100$ parameter samples $\boldsymbol{\theta}_{v,i}$ in the validation dataset. The resulting coefficient of determination for the comparison of the likelihoods equals $R^2 = 0.51$. When we simplify the GLaM by using the normality assumption and setting $\lambda_3 = \lambda_4 = 0.13$ the prediction is improved to $R^2 = 0.81$, see Figure 4.5b. As shown in the analysis of the

normalized Wasserstein distance, there is only a slightly better approximation of the distributions $p_{Q_w}(q_w|\boldsymbol{\theta}_v)$ when using the full GLD. Additionally, through the normality assumption, the model is regularized and less prone to overfitting, contributing to the improved prediction of the log-likelihood. Due to these reasons, all further analyses are carried out with the simplified GLaM which assumes a normal distribution of the surrogates per window $\hat{p}_{Q_w}(q_w|\boldsymbol{\theta}_v)$.



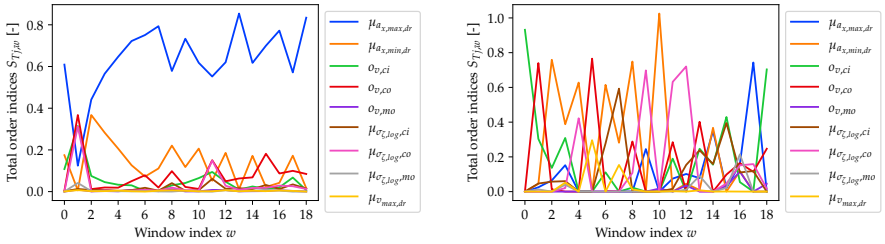
(a) Full GLaM ($R^2 = 0.51$) **(b)** GLaM with $\lambda_3 = \lambda_4 = 0.13$ ($R^2 = 0.81$)

Figure 4.5: Comparison of log-likelihood $\ell_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_{v,i})$ for different validation samples $\boldsymbol{\theta}_{v,i}$ between the reference solution obtained by a KDE approximation of repeated samples and the GLaM surrogate model [114]

Sensitivity analysis

Using the simplified GLaM, the Sobol' indices based sensitivity analysis is conducted to investigate the influence of the velocity simulation parameters on the gear damage sum based QoI. Figure 4.6 depicts the total order Sobol' indices $S_{Tj,w}$ along a route calculated for the mean $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ and the standard deviation $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$, see Table A.2 for a description of the different parameters $\theta_{v,j}$. For this route, the mean $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ of most windows is strongly influenced by $\mu_{a_{x,max,dr}}$. The standard deviation $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$, on the other hand, is influenced by various parameters depending on the window.

To analyze the aggregate sensitivity of the QoI to the parameters, the mean first and total order Sobol' indices for $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ and $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$ are calculated according to (4.18). Overall, the mean $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ is mainly sensitive to $\mu_{a_{x,max,dr}}$ while the standard deviation $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$ is primarily sensitive to $\mu_{a_{x,min,dr}}$ and $\mu_{v_{max,dr}}$, see Figure 4.7. For both mean and standard deviation, there is little difference between the mean first $\bar{S}_j$ and total order $\bar{S}_{Tj}$ indices, indicating no significant couplings of the velocity simulation parameters $\boldsymbol{\theta}_v$.

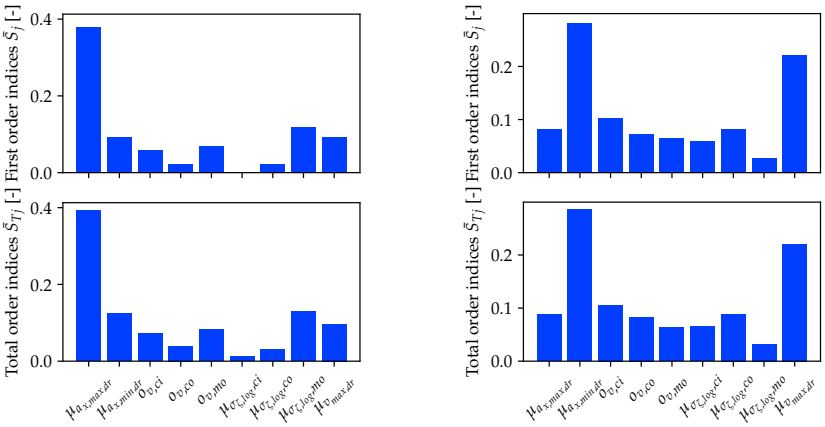


(a) $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ **(b)** $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$

Figure 4.6: Total order Sobol' indices $S_{Tj,w}$ along a route for $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ and $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$. Figure modified from [114].

The parameters $\mu_{a_{x,max,dr}}$ and $\mu_{a_{x,min,dr}}$ influence the acceleration and deceleration behavior of the driver, which is strongly related to the torque in the drivetrain and thus the investigated gear damage sum based QoI. Both parameters are therefore expected to strongly influence the mean $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ and standard deviation $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$ of the QoI distributions per window. The mean $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ is mainly influenced by the mean of the maximum driver acceleration $\mu_{a_{x,max,dr}}$, which has a direct impact on the torque signal and thus the gear damage sum based QoI when accelerating the vehicle. The mean of the maximum driver deceleration $\mu_{a_{x,min,dr}}$ on the other hand only slightly changes the torque signal as the recuperation of the electric motor when braking is limited but it still impacts the standard deviation $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$ strongly as it influences the number of potential acceleration events in a window which is limited by how often the vehicle can accelerate again after it first needs to decelerate to a certain velocity. The mean of the maximum driver speed $\mu_{v_{max,dr}}$ also doesn't directly influence the torque signal but through the influence on the mean velocity of the driver (especially on motorways) it impacts the duration and frequency of the driver accelerating and decelerating, and thus similar to the parameter $\mu_{a_{x,min,dr}}$ the standard deviation $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$.

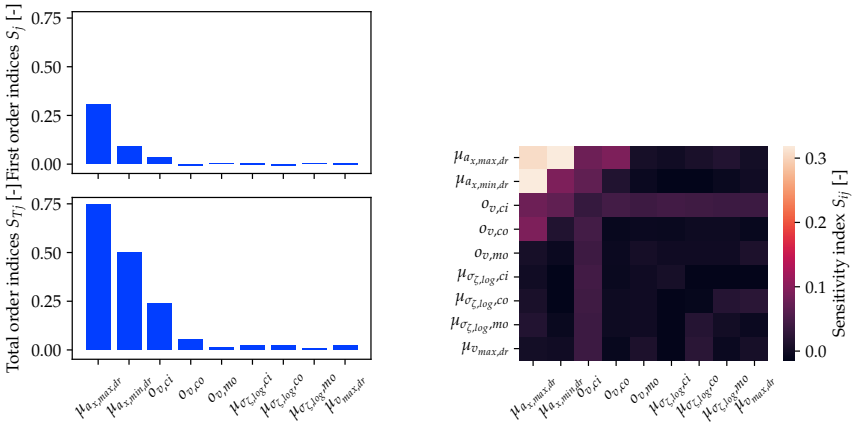
In the sensitivity analysis of the log-likelihood $\hat{\ell}_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_v)$, the first S_j and total order S_{Tj} Sobol' indices indicate a strong influence of the three parameters $\mu_{a_{x,max,dr}}$, $\mu_{a_{x,min,dr}}$ and $o_{v,ci}$, see Figure 4.8a, the first two of which were also the strongest influence factors in the sensitivity analysis of mean $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ and standard deviation $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$. The offset $o_{v,ci}$ influences the mean velocity driven in the city and has similar effects on the gear damage sum as the parameter $\mu_{v_{max,dr}}$ mentioned before, but with a focus on city streets. The values of the first and second-order indices are significantly different, indicating higher-order coupling effects. Figure 4.8b depicts the second order Sobol' indices S_{ij} with a strong coupling effect between $\mu_{a_{x,max,dr}}$ and $\mu_{a_{x,min,dr}}$. The log-likelihood is usually a non-linear function in which the interactions between the parameters can lead



(a) $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ **(b)** $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$

Figure 4.7: Mean first $\bar{S}_j$ and total order $\bar{S}_{Tj}$ indices for $\mathbb{E}[\hat{Q}_w|\boldsymbol{\theta}_v]$ and $\sigma(\hat{Q}_w|\boldsymbol{\theta}_v)$. Figure modified from [114].

to stronger changes in the log-likelihood than the parameters in isolation, which leads to the observed second and higher order coupling effects in the Sobol' indices.



(a) First S_j and total order S_{Tj} Sobol' indices **(b)** Second order Sobol' indices S_{ij}

Figure 4.8: Sobol' indices for the log-likelihood $\hat{\ell}_{q_{ref},Pr}(\boldsymbol{\theta}_v)$. Figure modified from [114].

Bayesian parameter inference

The reduced parameter set for which the posterior distribution $p_{\Theta_v^*}(\boldsymbol{\theta}_v^*|\boldsymbol{q}_{ref})$ is identified contains the parameters in

$$\mathrm{A}_v = \{\mu_{a_{x,max,dr}}, \mu_{a_{x,min,dr}}, o_{v,ci}\} \tag{4.26}$$

which have the most influence on the log-likelihood based on the Sobol' indices. The reduction is performed based on the sensitivities of the log-likelihood as it naturally takes all windows into account and is itself a central component of Bayesian inference. The posterior distribution is identified for each of the 18 drivers where for each driver a different set of reference measurements $\boldsymbol{q}_{ref} = (q_{1,ref}, ..., q_{N,ref})^T$ for the $N = 41$ windows of the considered route set exists. A KDE of the posterior distribution for one of the drivers, marginalized to one and two dimensions, is depicted in Figure 4.9. The black diamonds and lines indicate the default values of the parameter vector $\boldsymbol{\theta}_{v,0}$ of the respective driver, compared to the red diamonds and lines for the mean of the posterior predictive $\hat{\boldsymbol{\theta}}_v^*$. For $\mu_{a_{x,min,dr}}$ and $o_{v,ci}$, the default values are within the posterior distributions. The posterior distribution of $\mu_{a_{x,max,dr}}$ is bounded to the left due to the lower bound of the prior distribution $\boldsymbol{\theta}_{v,l}$ to keep the parameter values within a physically plausible range.

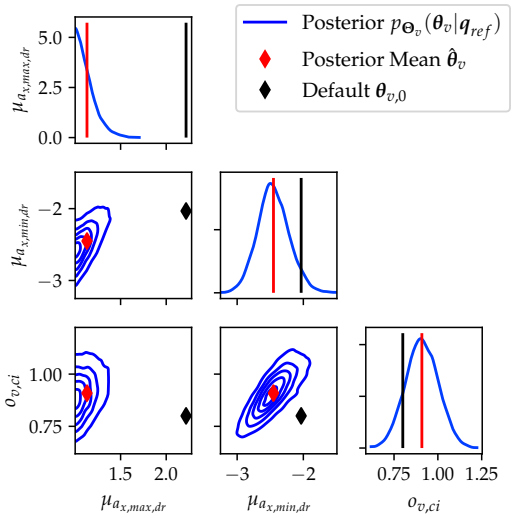


Figure 4.9: Joint posterior distribution $p_{\Theta_v^*}(\boldsymbol{\theta}_v^*|\boldsymbol{q}_{ref})$ marginalized to one and two dimensions compared to the default values $\boldsymbol{\theta}_{v,0}$ and the posterior mean $\hat{\boldsymbol{\theta}}_v^*$ of the parameter vector of the respective driver. Figure modified from [114].

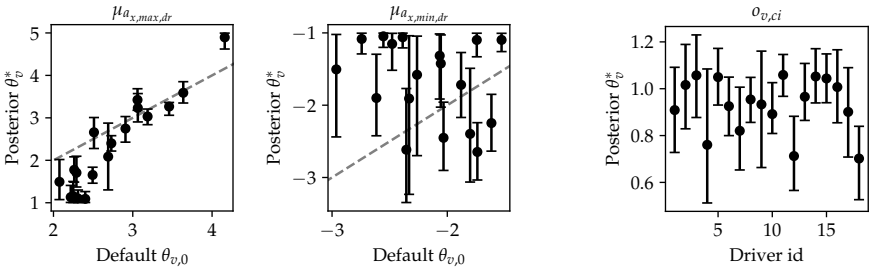
In Figure 4.10a, the posterior mean and the 95% credible interval for the parameters $\mu_{a_{x,max,dr}}$ and $\mu_{a_{x,min,dr}}$ are compared to their default values for all investigated drivers. The default values are estimated directly from the velocity profiles of the individual drivers with the aid of heuristics and, in contrast to the posterior distributions determined for the parameters, do not allow for a description of the estimation uncertainty. The identified posterior distributions of the mean of the maximum driver acceleration $\mu_{a_{x,max,dr}}$ are in a range similar to the default parameters of the respective drivers, indicating that both, the method used to derive the default parameters, and the Bayesian framework, lead to suitable values with a similar ranking of the drivers. Additionally, the posterior distribution provides the advantage of incorporating uncertainty in the parameter estimation results. As shown in the sensitivity analysis, see Figures 4.7 and 4.8, $\mu_{a_{x,max,dr}}$ is the most influential parameter on the investigated damage based QoI, making the identified values of this parameter especially relevant.

For the mean of the maximum driver deceleration $\mu_{a_{x,min,dr}}$ the posterior distributions mostly deviate from the default parameters and the uncertainty as expressed by the credible intervals is larger than for the parameter $\mu_{a_{x,max,dr}}$. The deviation of the posterior distributions of $\mu_{a_{x,min,dr}}$ indicates a possible error in the default parameter estimates or the simulation model itself, which is compensated in the posterior estimates. The higher uncertainty in the estimates of $\mu_{a_{x,min,dr}}$ compared to $\mu_{a_{x,max,dr}}$ is likely due to the lower sensitivity of the QoI to $\mu_{a_{x,min,dr}}$, especially for the mean, see Figure 4.7a, leading to less clear parameter identification results.

The velocity offset on city roads $o_{v,ci}$ is a separate case, as the default parameter values are not identified from data since the parameter cannot be directly measured from the velocity profiles. Instead, a rough estimate is used for the default parameters, which is identical for all drivers, hence, in Figure 4.10b, the identified posterior distributions are compared between the different drivers and not to the default values. In this case, the UQ framework provides the advantage of identifying unique parameter estimates for the different drivers, which can replace the generic default parameter estimates that are not based on data.

Uncertainty propagation

The posterior predictive distribution in (4.21) is evaluated for each driver by running the full simulation chain with 1 000 samples from the respective posterior distribution of each driver. The posterior predictive $p_Q(\boldsymbol{q}|\boldsymbol{q}_{ref})$ is compared to the distribution $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_{v,0})$ obtained by repeating the full simulation 1 000 times with the default parameter vector $\boldsymbol{\theta}_{v,0}$, as well as the reference measurement $\boldsymbol{q}_{ref}$ of the respective driver.



(a) Comparison of the posterior mean $\hat{\boldsymbol{\theta}}_v^*$ and the 95% credible intervals of the parameters $\mu_{a_{x,max,dr}}$ and $\mu_{a_{x,min,dr}}$ to the default parameter vector $\boldsymbol{\theta}_{v,0}$ for the different drivers. The dashed line indicates $\hat{\boldsymbol{\theta}}_v^* = \boldsymbol{\theta}_{v,0}$.

(b) Comparison of the posterior mean $\hat{\boldsymbol{\theta}}_v^*$ and the 95% credible intervals of the parameter $o_{v,ci}$ between the different drivers

Figure 4.10: Analysis of the identified posterior distribution $p_{\boldsymbol{\Theta}_v^*}(\boldsymbol{\theta}_v^* | \boldsymbol{q}_{ref})$ for the different drivers

Figure 4.11 depicts the 50%, 75%, 90% and 99% credible intervals as well as the mean μ_w of the distributions $p_{Q_w}(q_w|\boldsymbol{\theta}_{v,0})$ and $p_{Q_w}(q_w|\boldsymbol{q}_{ref})$ compared with the measured values $q_{w,ref}$ for the windows along a route for one of the drivers. Overall the intervals are shifted towards the measured reference solution $\boldsymbol{q}_{ref}$ when using the posterior parameter distribution $p_{\boldsymbol{\Theta}_v^*}(\boldsymbol{\theta}_v^*|\boldsymbol{q}_{ref})$ compared to the default parameters $\boldsymbol{\theta}_{v,0}$. When evaluating the 99% credible intervals for all windows of all routes of the investigated driver, the posterior predictive distribution covers 85% of all measured windows in $\boldsymbol{q}_{ref}$ compared to 71% of the default distribution.

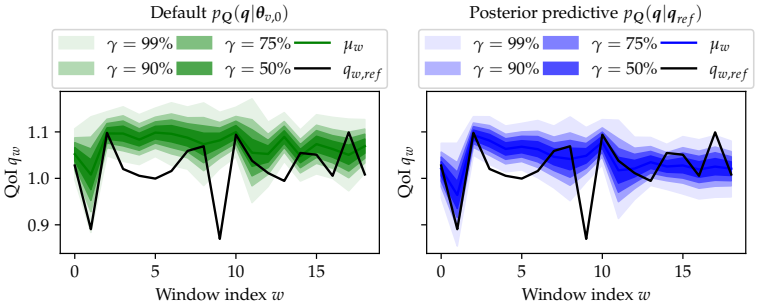
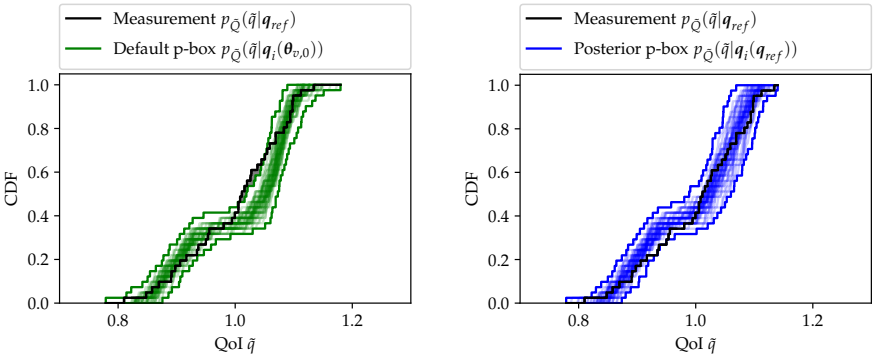


Figure 4.11: Comparison of the posterior predictive distribution $p_Q(\boldsymbol{q}|\boldsymbol{q}_{ref})$ and the distribution obtained with default parameters $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_{v,0})$ for multiple windows along a route. The center line is the mean μ_w of the distributions, the different color shades indicate the 50%, 75%, 90%, and 99% credible intervals. Figure modified from [114].

As described in Section 4.1.7 a p-box can be derived from the distributions $p_{\tilde{Q}}(\tilde{q}|\boldsymbol{q}_i)$ conditional on 1 000 different samples $\boldsymbol{q}_i$ drawn from the posterior predictive distribu-

tion $p_Q(\boldsymbol{q}|\boldsymbol{q}_{ref})$ or the distribution $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_{v,0})$ using default parameters. In Figure 4.12 these p-boxes are compared to the measured distribution $p_{\tilde{Q}}(\tilde{q}|\boldsymbol{q}_{ref})$ of the investigated driver. When using the posterior predictive distribution, the simulated p-box fully encloses the measured distribution $p_{\tilde{Q}}(\tilde{q}|\boldsymbol{q}_{ref})$, while for the default parameters, there is a significant deviation from the measurement.



(a) p-box obtained with $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_{v,0})$ **(b)** p-box obtained with $p_Q(\boldsymbol{q}|\boldsymbol{q}_{ref})$

Figure 4.12: Simulated p-boxes of the distribution $p_{\tilde{Q}}(\tilde{q})$ compared to the measured distribution $p_{\tilde{Q}}(\tilde{q}|\boldsymbol{q}_{ref})$. Figure modified from [114].

As q_w is the logarithm of the normalized gear damage sum, see (4.23), the mean $\bar{\tilde{q}}$ of it is linearly related to the accumulated log damage sum across all windows $d_{log,sum} = \sum_{w=1}^{N} \log(d_{G,w})$ as shown by

$$\begin{aligned}
\bar{\tilde{q}} &= \frac{1}{N}\sum_{w=1}^{N}\log\left(d'_{G,w}\right) = \frac{1}{N}\sum_{w=1}^{N}\log\left(\frac{d_{G,w}}{s_{w,l}\bar{d}_{G,meas}}\right), \\
\bar{\tilde{q}} &= \frac{1}{N}\left(\sum_{w=1}^{N}\log(d_{G,w}) - N(s_{w,l}\bar{d}_{G,meas})\right), \text{ with } \sum_{w=1}^{N}\log(d_{G,w}) = d_{log,sum}, \\
d_{log,sum} &= N\bar{\tilde{q}} + N(s_{w,l}\bar{d}_{G,meas}),
\end{aligned} \tag{4.27}$$

with the window length $s_{w,l}$, and the measured mean damage of all windows $\bar{d}_{G,meas}$, making it especially relevant for the presented gear dimensioning use-case. To compare the normalized damage sum $d'_{G,w}$ and the derived QoI q_w between different drivers, the measured mean damage of all windows $\bar{d}_{G,meas}$ of the driver with the highest mean damage is used for the normalization of the damage sums of all drivers.

Using (4.22) the mean distribution $p_{\bar{\tilde{Q}}}(\bar{\tilde{q}})$ is derived from the samples $\boldsymbol{q}_i$ taken from the posterior predictive distribution $p_Q(\boldsymbol{q}|\boldsymbol{q}_{ref})$ or the distribution $p_Q(\boldsymbol{q}|\boldsymbol{\theta}_{v,0})$ using

the default parameters. Figure 4.13 compares the mean distributions to the measured mean $\bar{\tilde{q}}_{ref}$ for all considered drivers. Overall, the stochastic velocity simulation, using both the default parameters and the posterior distribution, leads to results close to the measured mean while reproducing the differences between the drivers well. By using the posterior distribution of the parameters the 99% credible interval of the resulting mean distribution $p_{\bar{\tilde{Q}}}(\bar{\tilde{q}}|\boldsymbol{q}_{ref})$ includes the measured mean $\bar{\tilde{q}}_{ref}$ for 15 of 18 drivers compared to 9 of 18 drivers for the mean distribution $p_{\bar{\tilde{Q}}}(\bar{\tilde{q}}|\boldsymbol{\theta}_{v,0})$ resulting from the default parameters.

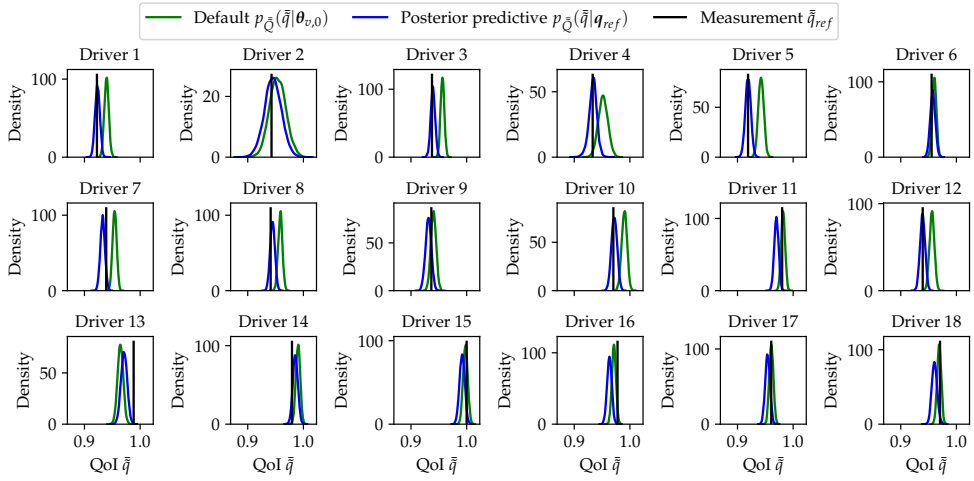


Figure 4.13: Comparison of mean distributions $p_{\bar{\tilde{Q}}}(\bar{\tilde{q}}|\boldsymbol{\theta}_{v,0})$ and $p_{\bar{\tilde{Q}}}(\bar{\tilde{q}}|\boldsymbol{q}_{ref})$ to the measured mean $\bar{\tilde{q}}_{ref}$ for all considered drivers

4.2.3 Discussion

The results presented in Section 4.2.2 show that the UQ framework can be applied to a system simulation and load-specific use-case to identify uncertainties in the velocity simulation parameters specific to that use-case. The GLaM surrogates can capture the random results of the simulation chain with sufficient accuracy when the parameter space is reduced to 9 parameters. For higher dimensions, the model deteriorates with the available number of samples, which is limited by the required computational effort. For an increasing number of samples, fitting of the GLaM also requires significantly more time, which at least partly is caused by the inverse definition of the PDF, see (2.70), that requires a numerical solution. The global sensitivity analysis helps to identify the parameters with the most influence on the QoI, considering the randomness of the QoI itself. Based on the surrogate model, the posterior distributions of the most important

parameters are identified in a Bayesian framework. The posterior distributions allow for quantification of the parametric uncertainties of the simulation, and generally are in a realistic range, reflecting the observed differences in the driving behavior between the drivers. Using the default parameters while considering the randomness of the simulation chain already leads to a close match of the various simulated QoI distributions to the measured values of the QoI. By additionally propagating the identified uncertainties of the posterior distribution through the simulation chain to obtain the posterior predictive distribution, the coverage of the measured QoI values is further improved.

4.3 Generic velocity simulation Uncertainty Quantification

To showcase the ability of the UQ framework to handle multiple QoIs, a use-case without a specific system simulation is considered. The QoIs are derived directly from the velocity and acceleration profiles, aiming to achieve a more generic description of the velocity simulation uncertainty compared to using a single system simulation specific QoI as in Section 4.2. Using velocity and acceleration based QoIs is beneficial if the system simulation is computationally expensive, which in practice is often the case, as the system simulation does not have to be run as part of the DoE needed to derive the surrogate model. To still obtain uncertainties meaningful with respect to the system simulation QoIs, the velocity and acceleration based QoIs are chosen such that they are relevant to the system simulation QoIs using expert knowledge, heuristics, or correlation analysis. In the following use-case, velocity and acceleration based QoIs relevant to the gear damage sum based QoI used in Section 4.2 are derived and used to quantify uncertainties in the velocity simulation. The results are then compared to those directly obtained with the system simulation based QoI.

4.3.1 Simulation setup

A variety of statistics can be calculated from the velocity and acceleration profiles to describe their properties. These include mean, standard deviation, various quantiles, and others, see Table 4.1 for all statistics considered. These statistics u are calculated per window w, with window length $s_{w,l} = 5000\,\text{m}$ and shift $s_{w,s} = s_{w,l}$, to derive the QoIs $q_{w,u}$. For the mean, standard deviation, and quantiles, the QoI per window follows directly from the statistic, e.g., for the mean velocity $q_{w,\mu(v)} = \mu(v)_w$ where $\mu(v)_w$ is the mean velocity of window w. As for the gear damage sum used in Section 4.2,

the rainflow and retention pseudo damage sums are normalized and logarithmized to derive their QoIs. The QoI of the rainflow damage sum of the velocity is, for example, given by

$$q_{w,d_{RF}(v)} = \log\left(d'_{RF,w}(v)\right), \text{ with } d'_{RF,w}(v) = \frac{d_{RF,w}(v)}{s_{w,l}\bar{d}_{RF,meas}(v)} \tag{4.28}$$

with the QoIs of the other damage sums being calculated in the same manner. The retention pseudo damage sum is only calculated for the acceleration time series, but due to its dependence on the driven distance between time steps Δs_i which is a function of the driven velocity, it also depends on the velocity time series. Therefore, the QoI is denoted as $q_{w,d_{ret}(v,a)}$, although the dependence on v and a is dropped in the following for a more compact notation given by $q_{w,d_{ret}}$. The normalization constant, e.g., $\bar{d}_{RF,meas}(v)$ for the rainflow damage sum, is the mean damage of all measured windows of the respective driver. From the combination of all windows and statistics follows the QoI matrix $\boldsymbol{q}$ and the derived random variable $\boldsymbol{Q}$ to consider the stochasticity of the velocity profiles. The reference measurement $\boldsymbol{q}_{ref}$ is directly obtained from the measured velocities and accelerations of each driver. The schematic procedure to derive multiple QoIs from simulated and measured velocities is depicted in Figure 4.14.

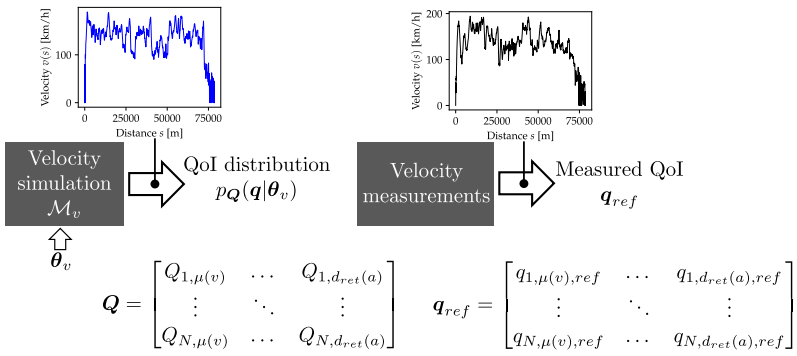


Figure 4.14: Derivation of QoIs directly from the simulated and measured velocities

4.3.2 Numerical results

The simulation setup described in Section 4.3.1 is applied to the same 18 drivers and their 216 km long route set with the corresponding velocity measurements as the setup in Section 4.2 specific to the system simulation. Therefore, the UQ framework is again applied to all drivers to derive a posterior distribution of the velocity simulation parameters for each driver. The procedure is demonstrated with one of the drivers as

Table 4.1: Statistics to describe time series $\boldsymbol{y} = (y_1, ..., y_N)^T$

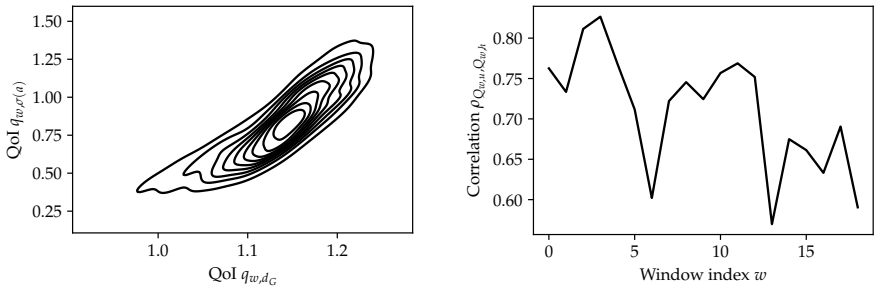
Statistic	Description
μ, σ, q_p	Mean, standard deviation, and p-th quantile with $p = 10\%, 90\%$
d_{RF}	Rainflow pseudo damage sum. Counting of cycles in time series $\boldsymbol{y}$ with rainflow algorithm [3], then damage accumulation with Miner's rule [70] assuming slope $k = 6$.
d_{ret}	Retention pseudo damage sum. Calculated as $d_{ret} = \sum_{i=1}^{N} \lvert y_i^k \Delta s_i \rvert$, with the driven distance between time steps Δs_i depending on the velocity and the damage exponent set to $k = 6$.

an example, where again, as in Section 4.2, the results for the surrogate model and the sensitivity analysis are identical for all drivers. The resulting posterior distributions and parts of the posterior predictive distributions are compared between all drivers as well as to the default parameters and the results obtained with them.

Quantity of Interest correlation analysis

In a preliminary analysis, the correlation of the velocity and acceleration based QoIs to the gear damage sum Q_{w,d_G} is investigated to identify those velocity and acceleration based QoIs with the closest resemblance to Q_{w,d_G}. A DoE with $M = 1\,000$ samples without replications, similar to the one in Section 4.2, is run with the velocity simulation to generate QoI samples from the prior predictive distribution. The prior from which the velocity simulation parameter samples $\boldsymbol{\theta}_{v,1}, ..., \boldsymbol{\theta}_{v,M}$ are drawn is a uniform distribution, see (4.5), and Table A.2 for the bounds on the distribution. As previously in Section 4.2, we reduce the parameter vector to the 9 parameters given by (4.24) because these parameters can influence the gear damage sum based QoI Q_{w,d_G} and therefore, this reduction is still valid since we aim to identify velocity and acceleration based QoIs that are similar to Q_{w,d_G} and are thus influenced by the same parameters.

For each window w using (4.13) the correlation $\hat{\rho}_{Q_{w,d_G},Q_{w,u}}$ between the prior predictive samples of Q_{w,d_G} and the prior predictive samples of all the other QoIs $Q_{w,u}$ is estimated. Figure 4.15a depicts the joint distribution of the prior predictive of Q_{w,d_G} and $Q_{w,\sigma(a)}$ for one of the windows where the correlation is $\hat{\rho}_{Q_{w,d_G},Q_{w,\sigma(a)}} = 0.87$. The correlations between the QoIs vary for each window, as the properties of the associated distributions change depending on the road and driving situation in the window. In Figure 4.15b this change of $\hat{\rho}_{Q_{w,d_G},Q_{w,\sigma(a)}}$ for the different windows along a route is depicted.



(a) Joint distribution of the prior predictive of Q_{w,d_G} and $Q_{w,\sigma(a)}$ for one window with the correlation $\hat{\rho}_{Q_{w,d_G},Q_{w,\sigma(a)}} = 0.87$

(b) Change of prior predictive correlation $\hat{\rho}_{Q_{w,d_G},Q_{w,\sigma(a)}}$ between Q_{w,d_G} and $Q_{w,\sigma(a)}$ for the windows along a route

Figure 4.15: Analysis of prior predictive distribution correlations between Q_{w,d_G} and $Q_{w,\sigma(a)}$

The correlations per window are used to derive the surrogate model for the QoIs based on a multidimensional normal distribution as described in Section (4.11). To assess the relation between the velocity and acceleration based QoIs and the gear damage sum QoI Q_{w,d_G}, the mean correlation

$$\bar{\rho}_{\boldsymbol{Q}_u,\boldsymbol{Q}_h} = \frac{1}{N}\sum_{w=1}^{N} \hat{\rho}_{Q_{w,u},Q_{w,h}} \tag{4.29}$$

between two QoIs $\boldsymbol{Q}_u$ and $\boldsymbol{Q}_h$ is defined. Table 4.2 lists the mean correlations of all velocity and acceleration based QoIs to the gear damage sum based QoI $\boldsymbol{Q}_{d_G}$ in descending order.

Table 4.2: Mean correlations $\bar{\rho}_{\boldsymbol{Q}_{d_G},\boldsymbol{Q}_u}$ of different velocity and acceleration based QoIs to the gear damage sum based QoI $\boldsymbol{Q}_{d_G}$

QoI	$\boldsymbol{Q}_{\sigma(a)}$	$\boldsymbol{Q}_{d_{ret}}$	$\boldsymbol{Q}_{d_{RF}(a)}$	$\boldsymbol{Q}_{\sigma(v)}$	$\boldsymbol{Q}_{d_{RF}(v)}$	$\boldsymbol{Q}_{q_{0.1}(a)}$
$\bar{\rho}_{\boldsymbol{Q}_{d_G},\boldsymbol{Q}_u}$	0.77	0.74	0.6	0.49	0.41	0.4
QoI	$\boldsymbol{Q}_{q_{0.9}(a)}$	$\boldsymbol{Q}_{q_{0.9}(v)}$	$\boldsymbol{Q}_{q_{0.1}(v)}$	$\boldsymbol{Q}_{\mu(v)}$	$\boldsymbol{Q}_{\mu(a)}$	
$\bar{\rho}_{\boldsymbol{Q}_{d_G},\boldsymbol{Q}_u}$	0.34	0.29	0.15	0.09	0.08	

The highest correlations are observed for QoIs which take changes in the acceleration series into account such as $\boldsymbol{Q}_{\sigma(a)}$, $\boldsymbol{Q}_{d_{ret}}$, and $\boldsymbol{Q}_{d_{RF}(a)}$, while the mean values $\boldsymbol{Q}_{\mu(v)}$ and $\boldsymbol{Q}_{\mu(a)}$ have a negligible mean correlation with $\boldsymbol{Q}_{d_G}$. That is to be expected as the gear damage sum based QoI $\boldsymbol{Q}_{d_G}$ is calculated from the torque series, which are sensitive

to vehicle accelerations. In the subsequent steps we focus on the two QoIs with the strongest correlations $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$, the QoI matrix is therefore reduced to

$$\boldsymbol{Q} = \begin{bmatrix} Q_{1,\sigma(a)} & Q_{1,d_{ret}} \\ \vdots & \vdots \\ Q_{N,\sigma(a)} & Q_{N,d_{ret}} \end{bmatrix}. \tag{4.30}$$

This choice of QoIs is an example to demonstrate how a set of velocity and acceleration based QoIs relevant to a specific component load can be derived. In other applications, these QoIs might already be known, or they can be derived with alternative methods.

Surrogate model generation

Based on the DoE of the previous section, a surrogate model for the two considered QoIs is derived. First a separate GLaM is build for both QoIs and each of their $N = 41$ windows, based on the parameter sample set $\{\boldsymbol{\theta}_{v,1}, ..., \boldsymbol{\theta}_{v,M}\}$ and the associated samples $\{\boldsymbol{q}_{\sigma(a),1}, ..., \boldsymbol{q}_{\sigma(a),M}\}$ and $\{\boldsymbol{q}_{d_{ret},1}, ..., \boldsymbol{q}_{d_{ret},M}\}$ for each QoI. Each sample vector contains a value for each window in the simulated route set, e.g., $\boldsymbol{q}_{\sigma(a),i} = (q_{1,\sigma(a)}, ..., q_{N,\sigma(a)})_i^T$. The GLaMs and their distributions $\hat{p}_{Q_{w,u}}(q_{w,u}|\boldsymbol{\theta}_v)$ for each QoI u and window w are fitted with the methods described in Sections 2.3.4 and 4.1.3. In this case of a multidimensional QoI distribution per window, the GLaM distributions are fixed to the shape of a normal distribution, i.e., $\lambda_3 = \lambda_4 = 0.13$, see Section 4.1.3. The polynomial degrees of the remaining two lambda parameters λ_1 and λ_2 are set to $p_1 = 2, p_2 = 1$ while $q_1 = q_2 = 1$ for the full polynomial basis in the truncation sets $\{\mathcal{A}_l : l = 1, 2\}$.

Based on the GLaMs and the QoI correlations derived from the DoE samples using (4.13) the surrogate model of the joint distribution of each window $\hat{p}_{\boldsymbol{Q}_w}(\boldsymbol{q}_w|\boldsymbol{\theta}_v)$ conditional on the velocity simulation parameters $\boldsymbol{\theta}_v$ is given by the multidimensional normal distribution

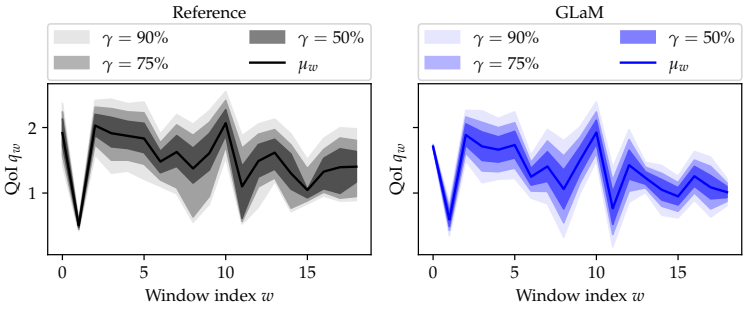
$$\boldsymbol{Q}_w = (Q_{w,\sigma(a)}, Q_{w,d_{ret}}) \sim \mathcal{N}(\boldsymbol{\mu}_w|\boldsymbol{\theta}_v, \boldsymbol{\Sigma}_w|\boldsymbol{\theta}_v) \tag{4.31}$$

previously defined in (4.11).

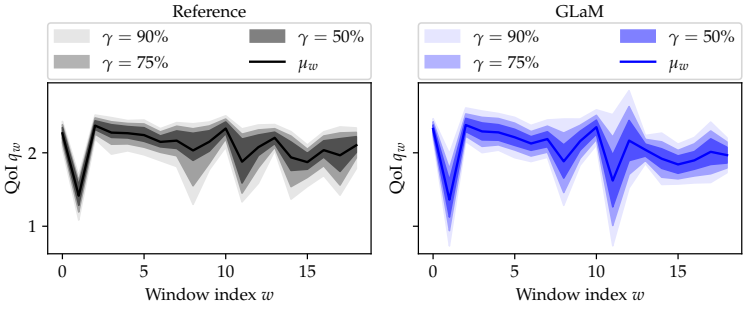
Surrogate model validation

For validation, as described in Section 4.1.4, we run a second DoE using the full simulation with $M_{val} = 100$ parameter samples and $R = 1\,000$ repetitions for each parameter sample. Figure 4.16 compares how the means $\mu_{w,\sigma(a)}$ and $\mu_{w,d_{ret}}$ as well

as the 50%, 75%, and 90% credible intervals of the distributions $\hat{p}_{Q_{w,\sigma(a)}}(q_{w,\sigma(a)}|\boldsymbol{\theta}_v)$ and $\hat{p}_{Q_{w,d_{ret}}}(q_{w,d_{ret}}|\boldsymbol{\theta}_v)$ differ to their respective values derived from the reference distributions of the two QoIs for a fixed value of the parameter vector $\boldsymbol{\theta}_v$ along the different windows of one of the routes. The GLaM distributions show some deviations from the reference distributions, especially for the credible intervals, but overall, the differences in mean and variance between the windows are still sufficiently reproduced.



(a) Comparison of $p_{Q_{w,\sigma(a)}}(q_{w,\sigma(a)}|\boldsymbol{\theta}_v)$ with $\hat{p}_{Q_{w,\sigma(a)}}(q_{w,\sigma(a)}|\boldsymbol{\theta}_v)$

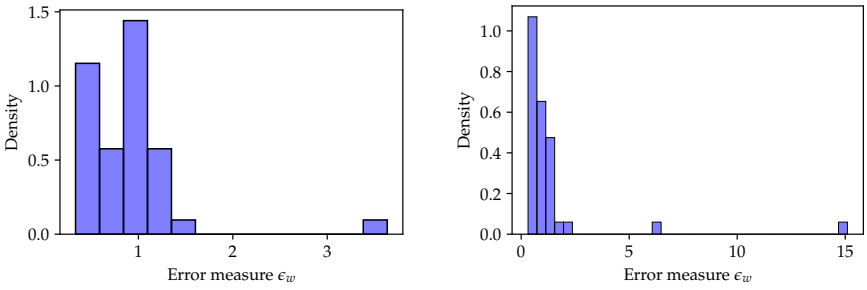


(b) Comparison of $p_{Q_{w,d_{ret}}}(q_{w,d_{ret}}|\boldsymbol{\theta}_v)$ with $\hat{p}_{Q_{w,d_{ret}}}(q_{w,d_{ret}}|\boldsymbol{\theta}_v)$

Figure 4.16: Comparison of reference and GLaM distributions of two QoIs $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ with a fixed parameter vector $\boldsymbol{\theta}_v$ for different windows along a route. The center line is the mean $\mu_{w,\sigma(a)}$, or $\mu_{w,d_{ret}}$ respectively, and the different color shades indicate the 50%, 75%, and 90% credible intervals of the distributions.

Between the GLaMs and the reference distributions of the two QoIs, the normalized Wasserstein distance according to (4.15) and subsequently the error measures per window $\epsilon_{w,\sigma(a)}$ and $\epsilon_{w,d_{ret}}$, see (4.16), are calculated. In Figure 4.17 the distributions of the error measures $\epsilon_{w,\sigma(a)}$ and $\epsilon_{w,d_{ret}}$ across all windows are depicted. For both QoIs outliers of the error measures are observed for some windows, i.e., depending on the window, the quality of the GLaM surrogate model can change considerably. The mean

errors given by (4.25) are $\bar{\epsilon}_{\sigma(a)} = 0.89$ and $\bar{\epsilon}_{d_{ret}} = 1.29$, highlighting the difficulty of building a surrogate model which works consistently across all windows and QoIs.



(a) $\epsilon_{w,\sigma(a)}$ **(b)** $\epsilon_{w,d_{ret}}$

Figure 4.17: Distributions of the error measures $\epsilon_{w,\sigma(a)}$ and $\epsilon_{w,d_{ret}}$

Using the prior mean, see (4.17), a reference measurement is derived, which is used to calculate the log-likelihood according to (4.9) with both the reference $p_{Q_w}(\boldsymbol{q}_w|\boldsymbol{\theta}_{v,i})$ and surrogate model $\hat{p}_{Q_w}(\boldsymbol{q}_w|\boldsymbol{\theta}_{v,i})$ joint distributions of both QoIs per window for each of the $M_{val} = 100$ parameter samples $\boldsymbol{\theta}_{v,i}$ in the validation dataset. Comparing the log-likelihoods naturally allows for evaluating the overall quality of the surrogate model considering both QoIs and their correlated joint distributions of all windows. As depicted in Figure 4.18, the general trend of the log-likelihood is reproduced correctly, but considerable fluctuations are leading to a coefficient of determination of only $R^2 = 0.56$, indicating the challenges associated with building a surrogate model for a correlated joint distribution of multiple QoIs.

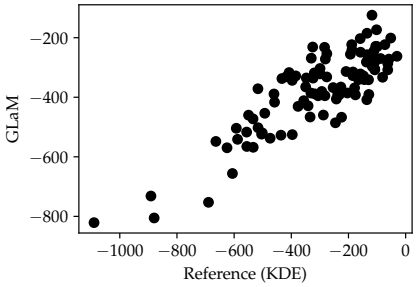
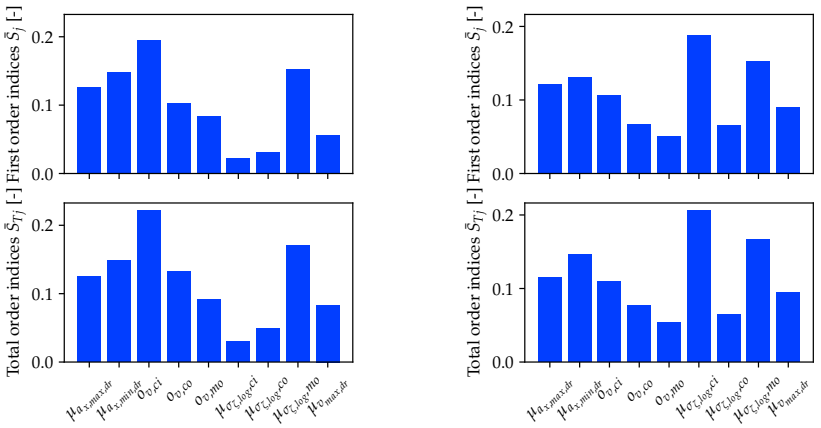


Figure 4.18: Comparison of the joint log-likelihood $\ell_{q_{ref,Pr}}(\boldsymbol{\theta}_{v,i})$ for different validation samples $\boldsymbol{\theta}_{v,i}$ between the reference solution obtained by a KDE approximation of repeated samples and the multivariate normal surrogate model ($R^2 = 0.56$)

Sensitivity analysis

Based on the multivariate normal surrogate model, see (4.11), a Sobol' indices based sensitivity analysis is conducted to investigate the influence of the velocity simulation parameters on the QoIs $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$. Analyzing the mean Sobol' indices, see (4.18), for the mean $\mathbb{E}[\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v]$ and the standard deviation $\sigma(\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v)$ in Figure 4.19 does not indicate a dominant sensitivity to one or a few of the parameters, see Table A.2 for a description of them, meaning that at least for some windows almost all parameters have a considerable influence on the mean and standard deviation of $\boldsymbol{Q}_{\sigma(a)}$. This is expected as the standard deviation of the acceleration, from which $\boldsymbol{Q}_{\sigma(a)}$ is derived, is a comparatively generic quantity and thus sensitive to many parameters. For both analyzed quantities $\mathbb{E}[\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v]$ and $\sigma(\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v)$ there is no significant difference between the mean first $\bar{S}_j$ and total order $\bar{S}_{Tj}$ indices and therefore only limited interaction between the velocity simulation parameters $\boldsymbol{\theta}_v$ concerning those quantities.

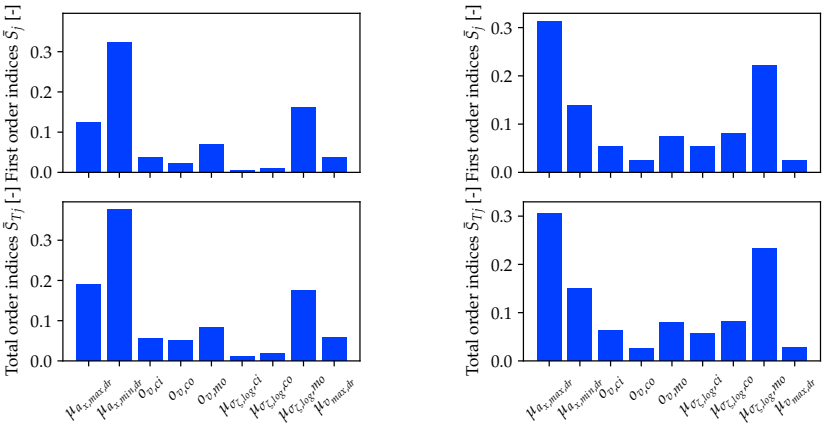


(a) $\mathbb{E}[\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v]$ **(b)** $\sigma(\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v)$

Figure 4.19: Mean first $\bar{S}_j$ and total order $\bar{S}_{Tj}$ indices for $\mathbb{E}[\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v]$ and $\sigma(\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v)$

The analysis of the mean Sobol' indices for the mean $\mathbb{E}[\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v]$ and the standard deviation $\sigma(\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v)$ in Figure 4.20 reveals a different pattern with stronger influences of the three parameters $\mu_{a_{x,max,dr}}$, $\mu_{a_{x,min,dr}}$, and $\mu_{\sigma_{\zeta,log},mo}$ on $\mathbb{E}[\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v]$ and $\sigma(\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v)$ while still displaying little interactions between the parameters. The mean of the maximum driver acceleration $\mu_{a_{x,max,dr}}$ and the mean of the maximum driver deceleration $\mu_{a_{x,min,dr}}$ directly influence the acceleration profile and therefore the retention pseudo damage sum and derived QoI $\boldsymbol{Q}_{d_{ret}}$. The mean of the log-standard deviation of velocity fluctuations on motorways $\mu_{\sigma_{\zeta,log},mo}$ impacts the amplitude and frequency of velocity

fluctuations on motorways and thus through the resulting changes in the velocity and acceleration profiles the QoI $Q_{d_{ret}}$.



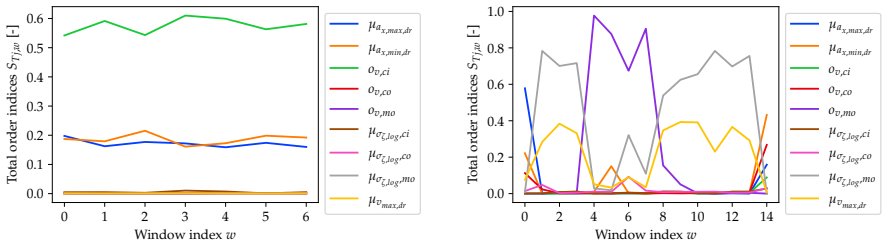
(a) $\mathbb{E}[\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v]$ **(b)** $\sigma(\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v)$

Figure 4.20: Mean first $\bar{S}_j$ and total order $\bar{S}_{Tj}$ indices for $\mathbb{E}[\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v]$ and $\sigma(\hat{Q}_{w,d_{ret}}|\boldsymbol{\theta}_v)$

Figure 4.21 depicts how the total order Sobol' indices $S_{Tj,w}$ of $\mathbb{E}[\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v]$ change depending on the window on two different routes, where one route mainly includes city roads and one route mainly includes motorway sections. The parameters with the highest Sobol' indices differ largely for the two routes, on the city route $o_{v,ci}$, $\mu_{a_{x,min,dr}}$, and $\mu_{a_{x,max,dr}}$ have a strong influence while on the motorway $\mu_{\sigma_{\zeta,log},mo}$, $o_{v,mo}$, and $\mu_{v_{max,dr}}$ have a high impact which is sensible as some of these parameters are specific to the respective road type where they show a strong influence and are expected to influence the accelerations and their standard deviation as described in Table A.2. This analysis shows exemplarily for the other Sobol' indices and their mean values in Figure 4.19 why there are no dominant parameters, as depending on the route and driving situation, the importance of the parameters changes significantly.

The sensitivity analysis of the joint log-likelihood $\hat{\ell}_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_v)$ of both QoIs $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ leads to similar results as for the QoI Q_{d_G} in Section 4.2. Meaning that the first S_j and total order S_{Tj} Sobol' indices indicate a strong influence of the same three parameters $\mu_{a_{x,max,dr}}$, $\mu_{a_{x,min,dr}}$ and $o_{v,ci}$, see Figure 4.22a, which is natural as the distributions per window of the QoIs $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ are correlated to the distributions of Q_{d_G}, as described within this Section 4.3.2, and those distributions form the basis for the calculation of the log-likelihood.

Compared to separately analyzing the sensitivity of $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ to the parameters, see Figures 4.19 and 4.20, analyzing the sensitivity of $\hat{\ell}_{\boldsymbol{q}_{ref,Pr}}(\boldsymbol{\theta}_v)$ which combines both



(a) City route **(b)** Motorway route

Figure 4.21: Total order Sobol' indices $S_{Tj,w}$ along two different routes for $\mathbb{E}[\hat{Q}_{w,\sigma(a)}|\boldsymbol{\theta}_v]$

QoIs, leads to the same set of important parameters as the system simulation specific sensitivity analysis in Section 4.2, which is the result of the combination leading to a quantity more specific to $\boldsymbol{Q}_{d_G}$. This demonstrates one of the advantages of using a correlated multidimensional surrogate model compared to modeling each QoI separately, which provides less clear information about the most important parameters regarding the combination of both QoIs and thus consequently also the system simulation specific QoI $\boldsymbol{Q}_{d_G}$.

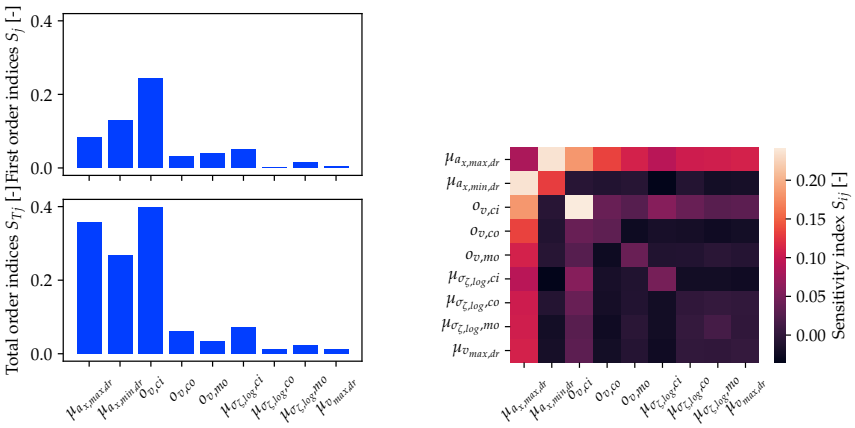
As in Section 4.2, the values of the first and second order indices are significantly different, indicating higher-order coupling effects, which are directly visible in the second order Sobol' indices S_{ij} in Figure 4.22b. The second order Sobol' indices show strong coupling effects between $\mu_{a_{x,max,dr}}$ and $\mu_{a_{x,min,dr}}$ as well as weaker couplings between $\mu_{a_{x,max,dr}}$ and most other parameters. As described in Section 4.2, these couplings can be attributed to the non-linearity of the log-likelihood function, leading to strong interactions between the parameters with respect to the log-likelihood.

Bayesian parameter inference

As in Section 4.2 we reduce the parameter set for which the posterior distribution $p_{\boldsymbol{\Theta}_v^*}(\boldsymbol{\theta}_v^*|\boldsymbol{q}_{ref})$ is identified to the parameters in

$$\mathrm{A}_v = \{\mu_{a_{x,max,dr}}, \mu_{a_{x,min,dr}}, o_{v,ci}\} \tag{4.32}$$

which have the most influence on the log-likelihood of $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$ based on the Sobol' indices. Since the log-likelihood considers all windows and both QoIs and is a central component in the context of Bayesian inference, its sensitivities are particularly informative and are therefore used for the decision on the reduction of the parameter set. The parameters in A_v are the same as those identified in Section 4.2 for the gear



(a) First S_j and total order S_{Tj} Sobol' indices

(b) Second order Sobol' indices S_{ij}

Figure 4.22: Sobol' indices for the joint log-likelihood $\hat{\ell}_{\boldsymbol{q}_{ref},Pr}(\boldsymbol{\theta}_v)$ of the two QoIs $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$

damage sum based QoI. The posterior distribution is identified for each of the 18 drivers, where for each driver, there exists a different set of reference measurements

$$\boldsymbol{q}_{ref} = \begin{bmatrix} q_{1,\sigma(a),ref} & q_{1,d_{ret},ref} \\ \vdots & \vdots \\ q_{N,\sigma(a),ref} & q_{N,d_{ret},ref} \end{bmatrix} \tag{4.33}$$

for the $N = 41$ windows in the considered route set and the two QoIs $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$. Figure 4.23 illustrates a KDE of the identified posterior distribution for one of the drivers, marginalized to one and two dimensions. The black diamonds and lines represent the default values of the parameter vector $\boldsymbol{\theta}_{v,0}$ for the respective driver, while the red diamonds and lines denote the mean of the posterior predictive $\hat{\boldsymbol{\theta}}_v^*$. For $\mu_{a_{x,min,dr}}$, the default value falls within the posterior distribution, while for the other parameters, there are significant differences between the default values and the posterior distribution and its mean. The posterior distribution of $\mu_{a_{x,max,dr}}$ is constrained on the left by the lower bound of the prior distribution $\boldsymbol{\theta}_{v,l}$, which defines the physically plausible parameter range, leading to a very narrow distribution near the lower bound that indicates a systematic error in the simulation compensated by extreme parameter values.

In Figure 4.24a, the posterior mean and 95% credible interval for the parameters $\mu_{a_{x,max,dr}}$ and $\mu_{a_{x,min,dr}}$ are compared to their default values for all investigated drivers. The default values are estimated with heuristics directly from the velocity profiles of each driver, and unlike the identified posterior distributions for the parameters, don't

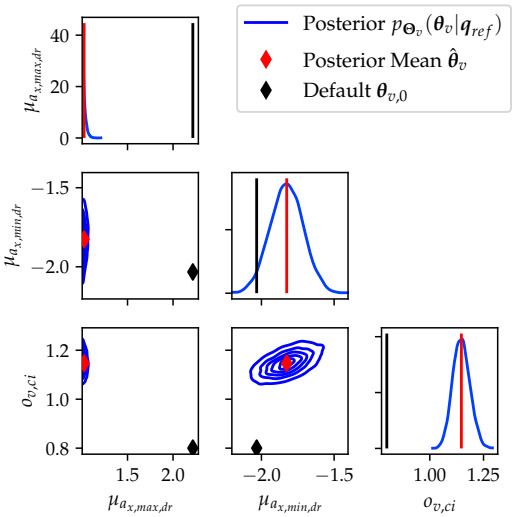
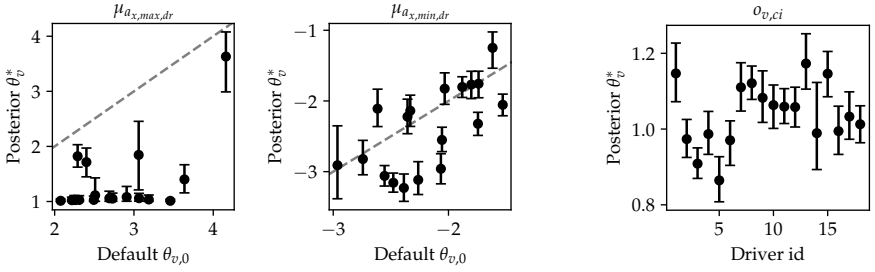


Figure 4.23: Joint posterior distribution $p_{\Theta_v^*}(\theta_v^*|q_{ref})$ marginalized to one and two dimensions compared to the default values $\theta_{v,0}$ and the posterior mean $\hat{\theta}_v^*$ of the parameter vector of the respective driver

allow for a description of the estimation uncertainty. Contrary to the system simulation specific analysis of the QoI Q_{d_G} in Section 4.2, see Figure 4.10a, the identified posterior distributions of the mean of the maximum driver acceleration $\mu_{a_{x,max,dr}}$ are not in a range similar to the default parameters of the respective drivers but instead are often limited by their lower bound given by $\theta_{v,l}$ as previously observed in Figure 4.23 for a single driver and indicating an error in the simulation model corrected through the posterior estimates. On the other hand, the posteriors of the mean of the maximum driver deceleration $\mu_{a_{x,min,dr}}$ are more similar to the default parameters of the various drivers and have a smaller 95% credible interval compared to the results obtained with the QoI Q_{d_G}. The differences in the posterior distributions indicate that the errors in the default parameter estimates and simulation model are different with regard to $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ compared to Q_{d_G} which is due to $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ being related to Q_{d_G} but not identical.

The changes of the 95% credible interval sizes compared to the system simulation specific analysis are reflective of the different sensitivities of the log-likelihood visible by comparing Figures 4.8 and 4.22 where the log-likelihood of the QoIs $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ is more uniformly sensitive to the parameters $\mu_{a_{x,max,dr}}$, $\mu_{a_{x,min,dr}}$, and $o_{v,ci}$ leading to similar uncertainties in their posterior estimates compared to the QoI Q_{d_G} where $\mu_{a_{x,max,dr}}$ has a more dominant influence and thus can be identified with less uncertainty than the other parameters.

Using velocity and acceleration based QoIs in the UQ framework allows for deriving estimates of parameters which otherwise are not estimated from data, but are set to a rough estimate which is the same for all drivers, without requiring a specific system simulation. One example of these parameters is the velocity offset on city roads $o_{v,ci}$, where the default parameter values are not identified from data, as the parameter cannot be estimated directly from the velocity profiles. Using the identified posterior distributions, these generic default parameter values can be replaced by unique parameter estimates for the different drivers. In Figure 4.24b, the mean and 95% credible intervals of the identified posterior distributions of $o_{v,ci}$ are depicted for the different drivers. When comparing the posterior results of $o_{v,ci}$ to the system simulation specific results in Figure 4.10b, their 95% credible intervals are significantly smaller which is reflective of the increased importance of $o_{v,ci}$ in the sensitivity analysis of the log-likelihood, see Figure 4.22. Overall, the investigated velocity and acceleration based QoIs lead to different posterior distributions of the velocity simulation parameters as the gear damage sum based QoI in Section 4.2, which, due to their different characteristics, is expected, while providing the added advantage of not requiring potentially costly runs of the system simulation.



(a) Comparison of the posterior mean $\hat{\boldsymbol{\theta}}_v^*$ and the 95% credible intervals of the parameters $\mu_{a_{x,max,dr}}$ and $\mu_{a_{x,min,dr}}$ to the default parameter vector $\boldsymbol{\theta}_{v,0}$ for the different drivers. The dashed line indicates $\hat{\boldsymbol{\theta}}_v^* = \boldsymbol{\theta}_{v,0}$.

(b) Comparison of the posterior mean $\hat{\boldsymbol{\theta}}_v^*$ and the 95% credible intervals of the parameter $o_{v,ci}$ between the different drivers

Figure 4.24: Analysis of the identified posterior distribution $p_{\Theta_v^*}(\boldsymbol{\theta}_v^* | \boldsymbol{q}_{ref})$ for the different drivers

Uncertainty propagation

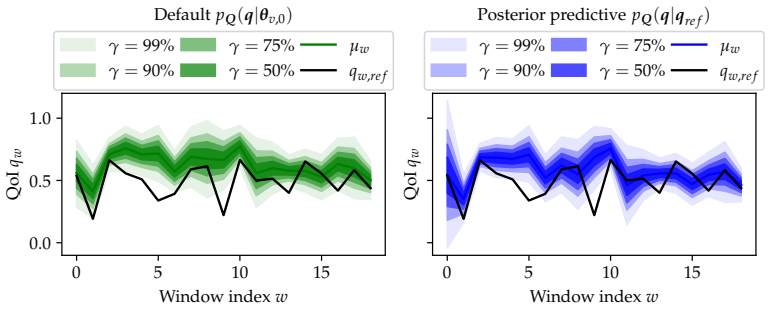
In the uncertainty propagation step, not only the influence of the uncertainties in the posterior distribution of each driver on the velocity and acceleration based QoIs $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$ is investigated, but also on the gear damage sum based QoI $\boldsymbol{Q}_{d_G}$ which as shown in Table 4.2 is correlated to $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$. The aim is to investigate

if the identified posterior distributions based on the QoIs $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$ are not only meaningful concerning those QoIs but also for the QoI $\boldsymbol{Q}_{d_G}$. Using the UQ framework based on velocity and acceleration based QoIs could thus potentially provide the advantage of a reduced computational effort compared to the system simulation specific approach in Section 4.2, as no runs of the system simulation are required for the identification of the posterior distributions.

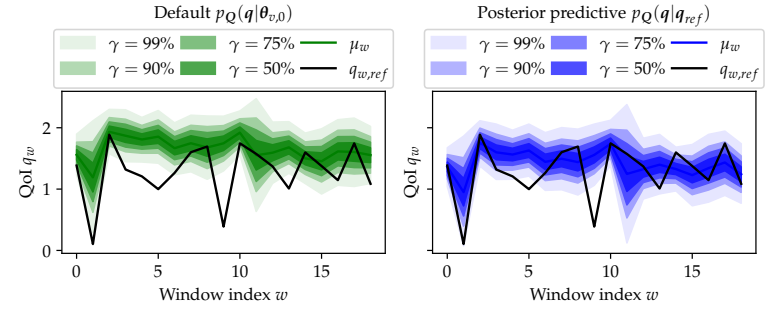
The posterior predictive distribution in (4.21) is evaluated for each driver by running the full simulation chain with 1 000 samples from the respective posterior distribution of each driver. The marginal distributions for each QoI derived from the posterior predictive, e.g., $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{q}_{ref})$ for the QoI $\boldsymbol{Q}_{\sigma(a)}$, are compared to the respective default distributions, e.g., $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{\theta}_{v,0})$, obtained by repeating the full simulation 1 000 times with the default parameter vector $\boldsymbol{\theta}_{v,0}$, and the reference measurement of the respective driver and QoI, e.g., $\boldsymbol{q}_{\sigma(a),ref}$.

Figure 4.25 compares the 50%, 75%, 90% and 99% credible intervals as well as the means $\mu_{w,\sigma(a)}$ and $\mu_{w,d_{ret}}$ of the distributions $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{\theta}_{v,0})$ and $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{q}_{ref})$, respective $p_{\boldsymbol{Q}_{d_{ret}}}(\boldsymbol{q}_{d_{ret}}|\boldsymbol{\theta}_{v,0})$ and $p_{\boldsymbol{Q}_{d_{ret}}}(\boldsymbol{q}_{d_{ret}}|\boldsymbol{q}_{ref})$, with the measured values $q_{w,\sigma(a),ref}$ and $q_{w,d_{ret},ref}$ for the windows along a route for one of the drivers. For both QoIs the width of the intervals for some windows changes significantly and the mean of the distributions is generally shifted towards the measured reference solutions $\boldsymbol{q}_{\sigma(a),ref}$ and $\boldsymbol{q}_{d_{ret},ref}$ when using the posterior parameter distribution $p_{\boldsymbol{\Theta}_v^*}(\boldsymbol{\theta}_v^*|\boldsymbol{q}_{ref})$ compared to the default parameters $\boldsymbol{\theta}_{v,0}$. Evaluating the 99% credible intervals for all windows of all routes of the investigated driver, the posterior predictive distribution of $\boldsymbol{Q}_{\sigma(a)}$ covers 93% of all measured windows in $\boldsymbol{q}_{\sigma(a),ref}$ compared to 78% for the default distribution, while for $\boldsymbol{Q}_{d_{ret}}$ the posterior predictive distribution covers 85% compared to 73% for the default distribution.

From each, the posterior predictive distribution $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{q}_{ref})$ and the distribution $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{\theta}_{v,0})$ using default parameters, 1 000 samples $\boldsymbol{q}_{\sigma(a),i}$ are drawn, respectively 1 000 samples $\boldsymbol{q}_{d_{ret},i}$ are drawn from $p_{\boldsymbol{Q}_{d_{ret}}}(\boldsymbol{q}_{d_{ret}}|\boldsymbol{q}_{ref})$ and $p_{\boldsymbol{Q}_{d_{ret}}}(\boldsymbol{q}_{d_{ret}}|\boldsymbol{\theta}_{v,0})$. Using those samples and the procedure described in Section 4.1.7, p-boxes can be derived from the distributions $p_{\tilde{\boldsymbol{Q}}_{\sigma(a)}}(\tilde{q}_{\sigma(a)}|\boldsymbol{q}_{\sigma(a),i})$ and $p_{\tilde{\boldsymbol{Q}}_{d_{ret}}}(\tilde{q}_{d_{ret}}|\boldsymbol{q}_{d_{ret},i})$ conditional on $\boldsymbol{q}_{\sigma(a),i}$ and $\boldsymbol{q}_{d_{ret},i}$. In Figure 4.26 these p-boxes are compared to the measured distributions $p_{\tilde{\boldsymbol{Q}}_{\sigma(a)}}(\tilde{q}_{\sigma(a)}|\boldsymbol{q}_{\sigma(a),ref})$ and $p_{\tilde{\boldsymbol{Q}}_{d_{ret}}}(\tilde{q}_{d_{ret}}|\boldsymbol{q}_{d_{ret},ref})$ of the investigated driver. In both cases, the simulated p-box derived using the posterior predictive distribution almost fully encloses the measured distribution, while for the simulated p-box of $p_{\tilde{\boldsymbol{Q}}_{\sigma(a)}}(\tilde{q}_{\sigma(a)})$ using default parameters, there are significant deviations from the measured distribution and whilst for the p-box of $p_{\tilde{\boldsymbol{Q}}_{d_{ret}}}(\tilde{q}_{d_{ret}})$ there is little difference between the default parameters and the posterior predictive distribution.



(a) $p_{Q_{\sigma(a)}}(q_{\sigma(a)}|\theta_{v,0})$ and $p_{Q_{\sigma(a)}}(q_{\sigma(a)}|q_{ref})$

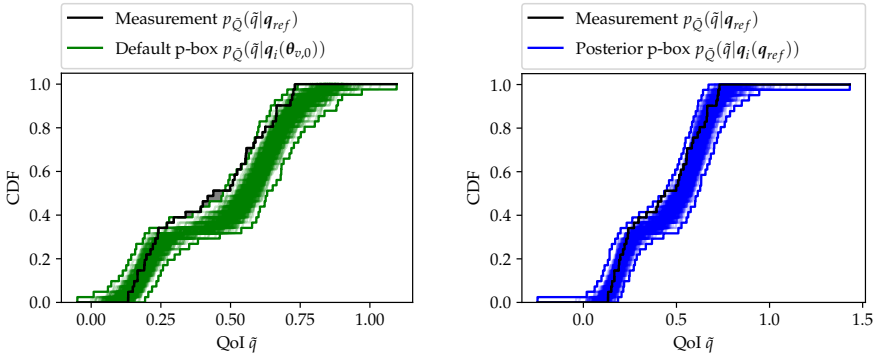


(b) $p_{Q_{d_{ret}}}(q_{d_{ret}}|\theta_{v,0})$ and $p_{Q_{d_{ret}}}(q_{d_{ret}}|q_{ref})$

Figure 4.25: Comparison of the posterior predictive distributions and the distributions obtained with default parameters for two QoIs $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ for multiple windows along a route. The center line is the mean $\mu_{w,\sigma(a)}$, or $\mu_{w,d_{ret}}$ respectively, and the different color shades indicate the 50%, 75%, 90%, and 99% credible intervals of the distributions.

To evaluate the usability of the posterior distributions identified using the velocity and acceleration based QoIs $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ for predicting the system simulation based QoI Q_{d_G}, the p-box of $p_{\tilde{Q}_{d_G}}(\tilde{q}_{d_G})$ is also evaluated and compared to the system simulation specific p-box in Section 4.2. Figure 4.27 depicts the comparison of the p-boxes derived using the posterior distribution based on Q_{d_G} and the posterior distribution based on $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ for the investigated driver. Both approaches fully enclose the measured distribution $p_{\tilde{Q}_{d_G}}(\tilde{q}_{d_G}|q_{d_G,ref})$, the posterior distribution derived using only the velocity simulation based QoIs $Q_{\sigma(a)}$ and $Q_{d_{ret}}$ is therefore a suitable substitute for the system simulation specific posterior distribution based on the QoI Q_{d_G}.

To compare the posterior predictive results between the different drivers we use the mean distributions $p_{\bar{\tilde{Q}}_{\sigma(a)}}(\bar{\tilde{q}}_{\sigma(a)})$ and $p_{\bar{\tilde{Q}}_{d_{ret}}}(\bar{\tilde{q}}_{d_{ret}})$ derived using (4.22) from the samples $q_{\sigma(a),i}$ and $q_{d_{ret},i}$ taken from the respective posterior predictive and default distributions. To make the normalized retention pseudo damage sum $d'_{ret,w}(v,a)$ and the derived



(a) p-box obtained with $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{\theta}_{v,0})$

(b) p-box obtained with $p_{\boldsymbol{Q}_{\sigma(a)}}(\boldsymbol{q}_{\sigma(a)}|\boldsymbol{q}_{ref})$

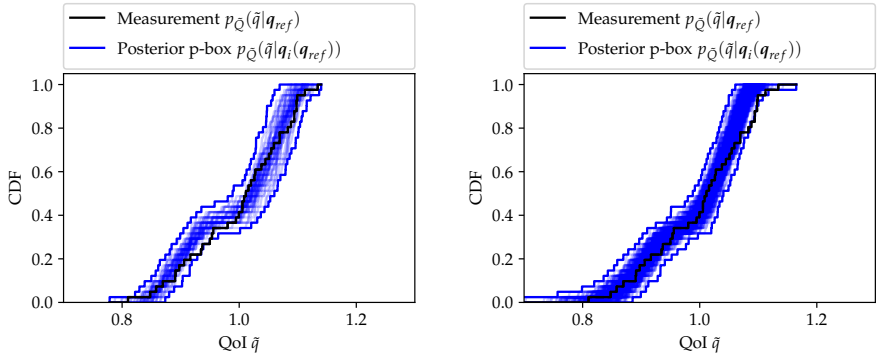
(c) p-box obtained with $p_{\boldsymbol{Q}_{d_{ret}}}(\boldsymbol{q}_{d_{ret}}|\boldsymbol{\theta}_{v,0})$

(d) p-box obtained with $p_{\boldsymbol{Q}_{d_{ret}}}(\boldsymbol{q}_{d_{ret}}|\boldsymbol{q}_{ref})$

Figure 4.26: Simulated p-boxes of the distributions $p_{\tilde{Q}_{\sigma(a)}}(\tilde{q}_{\sigma(a)})$ and $p_{\tilde{Q}_{d_{ret}}}(\tilde{q}_{d_{ret}})$ compared to the measured distributions $p_{\tilde{Q}_{\sigma(a)}}(\tilde{q}_{\sigma(a)}|\boldsymbol{q}_{\sigma(a),ref})$ and $p_{\tilde{Q}_{d_{ret}}}(\tilde{q}_{d_{ret}}|\boldsymbol{q}_{d_{ret},ref})$

QoI $q_{w,d_{ret}}$ comparable between drivers the measured mean damage of all windows $\bar{d}_{ret,meas}(v,a)$ of the driver with the highest mean damage is used for the normalization of all drivers.

Figure 4.28 compares the mean distributions $p_{\bar{\tilde{Q}}_{\sigma(a)}}(\bar{\tilde{q}}_{\sigma(a)})$ of the QoI $\boldsymbol{Q}_{\sigma(a)}$ to the measured mean $\bar{\tilde{q}}_{\sigma(a),ref}$ for all considered drivers. The 99% credible interval of the mean distribution $p_{\bar{\tilde{Q}}_{\sigma(a)}}(\bar{\tilde{q}}_{\sigma(a)}|\boldsymbol{\theta}_{v,0})$ of the stochastic velocity simulation derived using the default parameters covers the measured mean for 7 of 18 drivers. When using the mean distribution $p_{\bar{\tilde{Q}}_{\sigma(a)}}(\bar{\tilde{q}}_{\sigma(a)}|\boldsymbol{q}_{ref})$ based on the posterior predictive the mean distributions are generally closer to the measured mean and the 99% credible interval leads to a coverage of the measured mean for 14 of 18 drivers.



(a) Posterior distribution based on $\boldsymbol{Q}_{d_G}$

(b) Posterior distribution based on $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$

Figure 4.27: Comparison of p-boxes of the distribution $p_{\tilde{Q}_{d_G}}(\tilde{q}_{d_G})$ derived using posterior distributions based on $\boldsymbol{Q}_{d_G}$ or $\boldsymbol{Q}_{\sigma(a)}$ and $\boldsymbol{Q}_{d_{ret}}$

The mean distributions $p_{\bar{\tilde{Q}}_{d_{ret}}}(\bar{\tilde{q}}_{d_{ret}})$ of the QoI $\boldsymbol{Q}_{d_{ret}}$ depicted in Figure 4.29 show considerable differences in their variance depending on the driver, these differences are reflected by both the distributions derived using the default parameters and the posterior predictive. The 99% credible interval of the mean distribution $p_{\bar{\tilde{Q}}_{d_{ret}}}(\bar{\tilde{q}}_{d_{ret}}|\boldsymbol{q}_{ref})$ based on the posterior predictive covers the measured mean $\bar{\tilde{q}}_{d_{ret},ref}$ of all 18 drivers while the distribution $p_{\bar{\tilde{Q}}_{\sigma(a)}}(\bar{\tilde{q}}_{\sigma(a)}|\boldsymbol{\theta}_{v,0})$ derived using default parameters only leads to a coverage of the measured mean for 16 of 18 drivers.

4.3.3 Discussion

The results presented in this Section 4.3.2 demonstrate the capability of the UQ framework to handle multiple QoIs derived directly from velocity and acceleration profiles. This approach provides a more generic description of velocity simulation uncertainty compared to using a system simulation specific QoI as in Section 4.2. Two velocity and acceleration based QoIs are chosen for the analysis due to their high correlation with the system simulation specific QoI of Section 4.2. Thereby allowing the derivation of uncertainties with respect to the system simulation specific QoI without having to run the system simulation, which in practice is often computationally expensive. The surrogate model based on a multivariate normal distribution can capture the general trends of the joint QoI distributions but shows deficiencies in its accuracy, highlighting the need for more advanced surrogates for correlated multidimensional distributions or alternative approaches such as combining multiple QoIs into a single

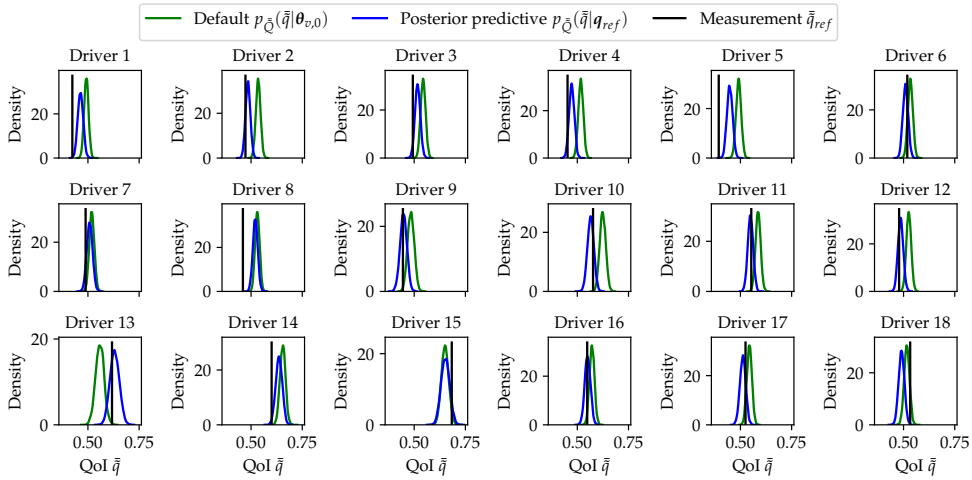


Figure 4.28: Comparison of the measured means $\tilde{\bar{q}}_{\sigma(a),ref}$ to the mean distributions $p_{\tilde{\bar{Q}}_{\sigma(a)}}(\tilde{\bar{q}}_{\sigma(a)}|\boldsymbol{\theta}_{v,0})$ and $p_{\tilde{\bar{Q}}_{\sigma(a)}}(\tilde{\bar{q}}_{\sigma(a)}|\boldsymbol{q}_{ref})$ for all considered drivers

quantity, which can be approximated by simpler surrogate models. The global sensitivity analysis shows that the parameters which have the greatest influence on the two velocity and acceleration based QoIs are similar to those that have the greatest influence on the system simulation specific QoI, which is due to the close relation of the different QoIs as shown by their high correlation. Based on the surrogate model, the posterior distributions of the most important parameters are identified within a Bayesian framework considering both QoIs. Due to the velocity and acceleration based QoIs being related but still different from the system simulation specific QoI, the identified posterior distributions differ from those identified with the system simulation specific QoI in Section 4.2 but are still in a realistic range and reflect the generally observed differences between the drivers. Propagating the identified posterior distributions to the QoIs leads to a similar or better match of the various simulated distributions to the measured values of both QoIs compared to using the default parameters, i.e., an improvement in both QoI dimensions is achieved. A straightforward variation of the presented use-case would be to consider multiple QoIs based on various loads and system simulations. This approach would allow for the derivation of velocity simulation uncertainties considering various components and their critical loads, leading to more robust and reliable predictions of system behavior under various loading conditions.

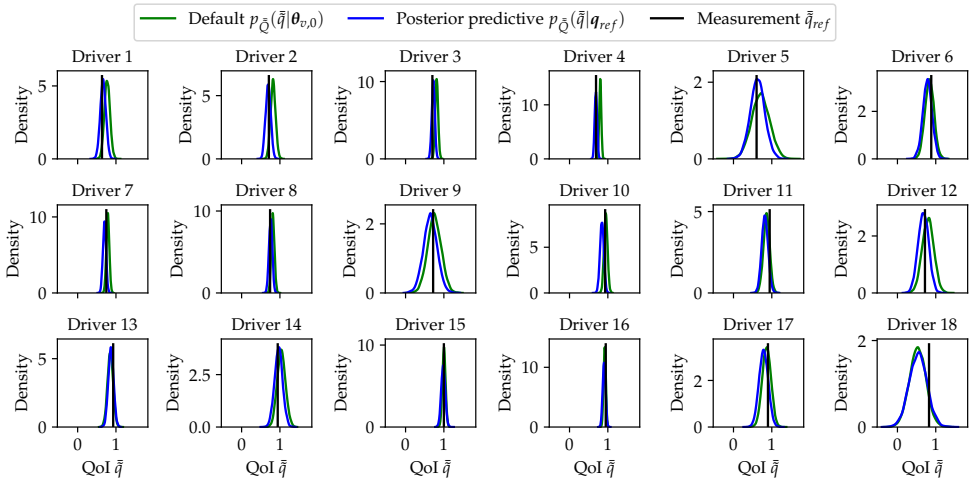


Figure 4.29: Comparison of the measured means $\bar{\tilde{q}}_{d_{ret},ref}$ to the mean distributions $p_{\bar{\tilde{Q}}_{d_{ret}}}(\bar{\tilde{q}}_{d_{ret}}|\boldsymbol{\theta}_{v,0})$ and $p_{\bar{\tilde{Q}}_{d_{ret}}}(\bar{\tilde{q}}_{d_{ret}}|\boldsymbol{q}_{ref})$ for all considered drivers

5 Conclusions

This chapter begins with a summary of the main findings of the thesis in Section 5.1. It then outlines potential directions for future research in Section 5.2, focusing on enhancing velocity simulation and UQ methods for virtual load generation.

5.1 Summary

This thesis addresses two related subjects. The first subject involves the enhancement and addition of methods for stochastic velocity simulation within the context of virtual load generation. The second subject focuses on the application of UQ methods to the velocity simulation discussed in the first subject, in order to account for uncertainties in the simulated velocities. Virtual load generation is an approach to derive automotive component loads fully virtually, making them available at an early stage of development, which enables the design of reliable components in a resource-efficient manner. The velocity simulation generates velocity profiles for different vehicle types and driving behaviors on arbitrary routes, which, through component-specific system simulations, are propagated to the component loads. The focus on the velocity simulation within this thesis is thus motivated by the dependence of many loads on the driven velocities, making the velocity simulation a central component in many virtual load generation use-cases. The application of UQ methods to the velocity simulation is particularly demanding due to the velocities and subsequent loads being random, requiring the use of UQ methods specific to stochastic simulators such as the stochastic velocity simulation [109].

By analyzing velocity fluctuations in measured velocity profiles, a red-noise AR(1) and an ARMA(2,1) model are identified as suitable for representing these fluctuations as a piecewise stationary process, see Section 3.2.2. Compared to the state-of-the-art AR(1) model with white noise [85], both identified models can better reproduce the correlation structure of the measured fluctuations, as demonstrated by the distinctly

smaller correlations in their residuum. Velocity fluctuations are shown to be bounded as they cannot exceed certain bounds, either due to physical limits of the vehicle or due to the driver's behavior. To address this in the simulated fluctuations, a method to bound the identified stochastic processes based on resampling is developed in Section 3.2.3.

The red-noise AR(1) and the ARMA(2,1) model are fitted to windows taken from the velocity profiles, resulting in a set of parameters for each window and model. The length of the windows is chosen such that the fluctuations within most windows can be assumed to be stationary, while windows that clearly violate this assumption are excluded, resulting in a piecewise stationary process. From the identified parameters and the mean velocities of the windows, a model is derived in Section 3.2.4 that describes the parameters as a multidimensional normal distribution conditional on the mean velocity of the fluctuations. By treating the fluctuation parameters as a distribution, the variability of the fluctuations is accounted for while the dependence on the mean velocity allows to distinguish between the typically stronger fluctuations at lower speeds, i.e., in the city, compared to the smaller fluctuations at higher speeds on the motorway, leading to a realistic velocity fluctuation model for a broad range of drivers and vehicles.

Dense traffic influences driven velocities and, consequently, the related component loads. Previous velocity simulation models [85, 93] do not sufficiently account for the impact of traffic on simulated velocities, leading to an underestimation of the proportion of low velocities, particularly on motorways where, without considering congestion, velocities are consistently high and do not reflect the lower speeds observed in reality. The traffic model proposed in Section 3.3 addresses this issue by generating realistic velocity drops on motorways due to traffic. This is achieved by combining location and time-dependent traffic speed distributions with congestion length distributions, i.e., distributions of the length of a velocity drop caused by traffic. A large database with FCD from Germany is analyzed to derive the required distributions for German motorways, where the model shows a significant improvement in the match of simulated and measured velocities through the consideration of slow velocities caused by traffic.

Motivated by the traffic influence model, which considers changes in the mean velocity caused by traffic, a model for traffic-related stop events is built in Section 3.4.1 to generate stop events resulting from traffic jams. It is shown that the stop density of traffic-related stops depends on the mean traffic speed. By considering this relationship in the velocity simulation, a good agreement of the simulation with measurements of traffic-related stops is achieved. Stop events on city and country roads, caused by stop signs, traffic lights, and other stop reasons, are modeled by assigning a stop probability and a distribution of the stop duration to each stop reason. In Section 3.4.2, stops in measured velocity profiles are systematically analyzed to derive new, more realistic

parameters for the stop probabilities and stop duration distributions, resulting in a more accurate representation of measured stops on city and country roads in the simulation.

Addressing the need to quantify the uncertainties of the simulated velocities with respect to the virtually generated loads, a UQ framework for this application is built in Section 4.1. Based on the simulated stochastic velocity profiles or the application-specific load profiles derived from them, QoIs are determined. These QoIs are modeled as probability distributions to consider their randomness caused by the stochastic velocity simulation. To reduce the runtime of the simulation chain while considering the randomness of the QoIs, stochastic surrogate models are derived and validated. The surrogates used are GLaMs, which combine the GLD with a PCE of the distribution parameters and can approximate a wide range of unimodal distributions. The surrogate models enable computationally efficient global sensitivity analysis using Sobol' indices and Bayesian parameter inference with MCMC methods to consider the uncertainty in the velocity simulation parameters. The identified posterior distributions describing the parameter uncertainties are propagated to the QoIs leading to the posterior predictive distribution describing the virtual loads including their uncertainty.

Two applications of the UQ framework demonstrate its ability to identify the most important parameters of the velocity simulation concerning the loads of the applications and to derive the posterior distribution of the identified parameters. In the first application presented in Section 4.2, the gear damage sum used for gear dimensioning, based on the drivetrain torque load, is utilized as a single QoI, leading to uncertainties specific to that load. By propagating the identified posterior distribution to the torque load, a significantly increased coverage of measured loads is achieved. The second application presented in Section 4.3 investigates the possibility of considering multiple QoIs within the UQ framework. By using multiple QoIs for different loads, e.g., a QoI for a thermal load and a QoI for a mechanical load, the effects of uncertain velocities on multiple loads can be considered simultaneously. Alternatively, multiple QoIs can be derived from velocity and acceleration profiles to substitute for a single system simulation based QoI by having related characteristics, thus enabling UQ analysis of that single QoI without running the potentially computationally intensive system simulation.

Overall, the significant improvements of the velocity simulation in combination with the UQ methods will further increase the usage of virtual loads in the development of new, innovative components in the automotive industry by increasing the accuracy and trustworthiness of the generated loads.

5.2 Future work

The velocity simulation model, as well as the developed UQ framework, still show deficiencies and potential for improvement. Adding new features could further improve the accuracy and usability of the virtually generated loads. This section addresses some of these deficiencies and potential improvements, outlining directions for future research.

By acquiring and analyzing FCD for more regions, the traffic model for motorways presented in Section 3.3 could be directly expanded to those regions, enabling the consideration of traffic influences in more usage scenarios. Future work could also investigate the possibility of extending the model to account for traffic influences on city and country roads. Traffic on those roads is more complex and often strongly influenced by intersections and traffic lights, making modeling more challenging. Addressing these challenges likely requires developing a method to better differentiate between slow traffic directly caused by stop reasons, such as traffic lights, and slow traffic caused by the traffic flow exceeding the road's capacity, which can also be limited by, for example, a traffic light. An additional difficulty is the drastically increasing length of the road network to consider compared to motorways, leading to significantly larger amounts of FCD that need to be analyzed, which can make data handling difficult.

The example used in Section 4.3, with two velocity and acceleration based QoIs substituting for the gear damage sum QoI of the first application, reveals some challenges associated with creating surrogate models for multiple, correlated QoIs. In the current approach, different surrogate models are trained for each QoI. By training the surrogate models jointly with the correlated QoIs, predictions could be improved. This could be achieved by extending the used GLaM to a multivariate GLaM based on an extension of the GLD to the multivariate case or by constructing another flexible multivariate distribution and approximating its parameters with PCEs. Alternatively, other surrogate models, such as the Gaussian process regression (GPR) extension cokriging [96], which can already handle multiple stochastic outputs, or the Stochastic polynomial chaos expansion (SPCE) method [112], which needs to be extended to the multidimensional setting, could be investigated. Addressing these further research needs on surrogates for stochastic simulators could significantly enhance the application of UQ methods to virtual load generation use-cases involving multiple loads and QoIs.

Besides the further research needs concerning the velocity simulation and UQ methods mentioned previously, other uncertainties along the virtual load generation simulation chain also need to be addressed in the future. These include uncertainties regarding the assumed usage scenarios, map data, and system simulations. Statistical analysis of the

usage scenarios, i.e., the routes driven by the drivers and their driving behavior, could be employed to check for the representativeness of the assumed usage scenarios compared to the ground truth of all real-world usage scenarios. By conducting a sensitivity analysis of the QoIs with respect to the map data, the importance of road properties such as slope, curvature, and speed limit regarding the simulated loads could be analyzed. Further investigations into the quality of the map data regarding the relevant road properties could then be conducted to quantify the effect of uncertainties in the map data on the QoIs. Finally, uncertainties in the system simulations, such as parameter uncertainties, could be addressed similarly to the uncertainties in the velocity simulation, except that system simulations are typically deterministic and thus do not require UQ methods specific to stochastic simulators, which usually simplifies the analysis. By quantifying these uncertainties and combining them with the velocity simulation uncertainties through methods such as Monte Carlo sampling, confidence in virtual loads can be increased. This facilitates further reductions in costly physical testing during product development while simultaneously increasing product reliability.

Appendix

A Velocity simulation parameters

Table A.1: List of velocity simulation vehicle parameters

Parameter θ_g	Description
A	Frontal area in m^2
c_w	Coefficient of drag
P_{max}	Engine power in kW
$T_{eng,max}$	Engine Torque in N m
$n_{dr,ax}$	Number of driven axles
n_{ax}	Total number of axles
f_{roll}	Rolling resistance coefficient
m_{veh}	Vehicle mass in kg
μ_k	Friction coefficient
r_{wheel}	Wheel diameter in m
n_g	Gear ratio
$v_{max,veh}$	Maximum vehicle velocity in km/h
v_{max}	Drivetrain efficiency parameter in km/h
λ_{max}	Drivetrain efficiency parameter

Table A.2: List of velocity simulation driver parameters and their bounds

Parameter θ_v	$\theta_{v,l}$	$\theta_{v,u}$	Description
$\mu_{a_{x,max,dr}}$	1	5	Mean of maximum driver acceleration in m/s^2
$\sigma_{a_{x,max,dr}}$	0.01	0.5	Standard deviation of maximum driver acceleration in m/s^2
$\mu_{a_{x,min,dr}}$	−5	−1	Mean of maximum driver deceleration in m/s^2
$\sigma_{a_{x,min,dr}}$	0.1	0.7	Standard deviation of maximum driver deceleration in m/s^2
$a_{y,max,dr}$	3	8	Maximum lateral driver acceleration in m/s^2
$\mu_{v_{max,dr}}$	110	150	Mean of maximum driver speed in km/h
$\sigma_{v_{max,dr}}$	0.05	15	Standard deviation of maximum driver speed in km/h
$a_{v_{max,dr}}$	60	80	Minimum of maximum driver speed in km/h
$b_{v_{max,dr}}$	120	160	Maximum of maximum driver speed in km/h
$o_{v,\{ci,co,mo\}}$	0.5	1.3	Velocity offset on city/country/motorway roads
$\mu_{\eta,\{ci,co,mo\}}$	0.7	1	Mean of correlation η of velocity fluctuations on city/country/motorway roads
$\sigma_{\eta,\{ci,co,mo\}}$	n.s.	n.s.	Standard deviation of correlation η of velocity fluctuations on city/country/motorway roads
$\mu_{\theta,\{ci,co,mo\}}$	0.4	1	Mean of correlation θ of velocity fluctuations on city/country/motorway roads
$\sigma_{\theta,\{ci,co,mo\}}$	n.s.	n.s.	Standard deviation of correlation θ of velocity fluctuations on city/country/motorway roads
$\mu_{\sigma_{\zeta,log},\{ci,co,mo\}}$	−5 −6 −7	−2 −2 −3	Mean of log-standard deviation of velocity fluctuations on city/country/motorway roads
$\sigma_{\sigma_{\zeta,log},\{ci,co,mo\}}$	0.05 0.05 0.05	0.7 1 1	Standard deviation of log-standard deviation of velocity fluctuations on city/country/motorway roads
$\underline{z}_{\{ci,co,mo\}}$	−4 −4 −3	−1 −0.5 −0.1	Lower bound of velocity fluctuations on city/country/motorway roads
$\bar{z}_{\{ci,co,mo\}}$	0.3 0.1 0.05	1.5 1 0.7	Upper bound of velocity fluctuations on city/country/motorway roads

List of Abbreviations

ACF	Autocorrelation function
AIC	Akaike information criterion
AR	Autoregressive
ARMA	Autoregressive–Moving-Average
BASt	Bundesanstalt für Straßen und Verkehrswesen (Federal Highway and Traffic Research Institute)
BIC	Bayesian information criterion
CDF	Cumulative Distribution Function
CI	Confidence interval
DoE	Design of Experiments
EM	Expectation–maximization
FCD	Floating car data
FGLS	Feasible generalized least-squares
GLaM	Generalized Lambda Model
GLD	Generalized Lambda Distribution
GMM	Gaussian mixture model
GPR	Gaussian process regression
GPS	Global Positioning System
IQR	Interquartile range
KDE	Kernel Density Estimation
LHS	Latin Hypercube Sampling
MA	Moving-Average
MAP	Maximum a posteriori
MBS	Multibody simulation
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation

MLMC	Multilevel Monte Carlo
MSE	Mean squared error
NISP	Non-Intrusive Spectral Projection
p-box	Probability-box
PACF	Partial autocorrelation function
PCE	Polynomial Chaos Expansion
PDF	Probability Density Function
QoI	Quantity of Interest
SPCE	Stochastic polynomial chaos expansion
UQ	Uncertainty Quantification

List of Figures

List of Tables

Bibliography

[1] Akaike, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6). Number: 6, pp. 716–723. doi: 10.1109/TAC.1974.1100705. url: http://ieeexplore.ieee.org/document/1100705/ (visited on 02/15/2025).

[2] Alkhatib, A. A., K. A. Maria, S. AlZu'bi, and E. A. Maria (2022): Smart Traffic Scheduling for Crowded Cities Road Networks. *Egyptian Informatics Journal* **23**(4). Number: 4, pp. 163–176. doi: 10.1016/j.eij.2022.10.002. url: https://linkinghub.elsevier.com/retrieve/pii/S1110866522000688 (visited on 04/10/2025).

[3] Amzallag, C., J. Gerey, J. Robert, and J. Bahuaud (1994): Standardization of the rainflow counting method for fatigue analysis. *International Journal of Fatigue* **16**(4). Number: 4, pp. 287–293. doi: 10.1016/0142-1123(94)90343-3. url: https://linkinghub.elsevier.com/retrieve/pii/0142112394903433 (visited on 12/05/2023).

[4] Arhin, S. A., A. Gatiba, M. Anderson, and M. Ribbisso (2019): Predicting STOP-Sign Compliance at All-Way Stop Intersections in Close Proximity to Signalized Intersections. *International Journal of Engineering Research and* **V8**(7). Number: 07. doi: 10.17577/IJERTV8IS070231. url: https://www.ijert.org/predicting-stop-sign-compliance-at-all-way-stop-intersections-in-close-proximity-to-signalized-intersections (visited on 04/10/2025).

[5] BASt (2019): *Automatische Zählstellen 2019*. url: https://www.bast.de/.

[6] Bertsche, B. and M. Dazer (2022): *Zuverlässigkeit im Fahrzeug- und Maschinenbau: Ermittlung von Bauteil- und System-Zuverlässigkeiten*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-65024-0. url: https://link.springer.com/10.1007/978-3-662-65024-0 (visited on 05/30/2025).

[7] Biedinger, C., T. Weyh, A. Opalinski, and M. Wagner (2016): Simulation of customer-specific vehicle usage. In: *Commercial vehicle technology 2016*. Commercial Vehicle Technology Symposium (CVT).

[8] Bishop, C. M. (1995): *Neural networks for pattern recognition*. Oxford : New York: Clarendon Press ; Oxford University Press. 482 pp.

[9] Blatman, G. and B. Sudret (2010): An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics* **25**(2). Number: 2, pp. 183–197. doi: 10.1016/j.probengmech.2009.10.003. url: https://linkinghub.elsevier.com/retrieve/pii/S0266892009000666 (visited on 03/07/2025).

[10] Blatman, G. and B. Sudret (2011): Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics* **230**(6). Number: 6, pp. 2345–2367. doi: 10.1016/j.jcp.2010.12.021. url: https://linkinghub.elsevier.com/retrieve/pii/S0021999110006856 (visited on 03/06/2025).

[11] Boettner, C. and N. Boers (2022): Critical slowing down in dynamical systems driven by nonstationary correlated noise. *Physical Review Research* **4**(1). Number: 1, p. 013230. doi: 10.1103/PhysRevResearch.4.013230. url: https://link.aps.org/doi/10.1103/PhysRevResearch.4.013230 (visited on 12/05/2023).

[12] Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2016): *Time series analysis: forecasting and control*. Fifth edition. Wiley series in probability and statistics. Hoboken, New Jersey: John Wiley & Sons, Inc. 669 pp.

[13] Braess, H.-H. and U. Seiffert, eds. (2013): *Vieweg Handbuch Kraftfahrzeugtechnik*. Wiesbaden: Springer Fachmedien Wiesbaden. doi: 10.1007/978-3-658-01691-3. url: http://link.springer.com/10.1007/978-3-658-01691-3 (visited on 06/13/2024).

[14] Brereton, T. (2015): *Methods of Monte Carlo Simulation*. Lecture Notes. url: https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/lehre/ws15/MonteCarloMethods/Lecture_Notes.pdf.

[15] Burger, M., K. Dreßler, and M. Speckert (2021): Load assumption process for durability design using new data sources and data analytics. *International Journal of Fatigue* **145**, p. 106116. doi: 10.1016/j.ijfatigue.2020.106116. url: https://linkinghub.elsevier.com/retrieve/pii/S0142112320306484 (visited on 06/18/2024).

[16] Burnham, K. P. and D. R. Anderson, eds. (2004): *Model Selection and Multimodel Inference*. New York, NY: Springer New York. doi: 10.1007/b97636. url: http://link.springer.com/10.1007/b97636 (visited on 06/04/2025).

[17] Caflisch, R. E. (1998): Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* **7**, pp. 1–49. doi: 10.1017/S0962492900002804. url: https://www.cambridge.org/core/product/identifier/S0962492900002804/type/journal_article (visited on 02/26/2025).

[18] Carrion-i-Silvestre, J. L., M. D. Gadea, and A. Montañés (2021): Nearly Unbiased Estimation of Autoregressive Models for Bounded Near-Integrated Stochastic

Processes*. *Oxford Bulletin of Economics and Statistics* **83**(1). Number: 1, pp. 273–297. DOI: 10.1111/obes.12399. URL: https://onlinelibrary.wiley.com/doi/10.1111/obes.12399 (visited on 12/20/2023).

[19] Cody, B. and M. Hanley (2003): *Stop Sign Violations Put Child Pedestrians at Risk: A National Survey of Motorist Behavior at Stop Signs in School Zones and Residential Areas*. National SAFE KIDS Campaign (NSKC). Washington, DC.

[20] Coleman, H. W. and W. G. Steele (2018): *Experimentation, Validation, and Uncertainty Analysis for Engineers*. 1st ed. Wiley. DOI: 10.1002/9781119417989. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119417989 (visited on 09/07/2022).

[21] Cortes, C. and V. Vapnik (1995): Support-vector networks. *Machine Learning* **20**(3). Number: 3, pp. 273–297. DOI: 10.1007/BF00994018. URL: http://link.springer.com/10.1007/BF00994018 (visited on 03/06/2025).

[22] Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **39**(1). Number: 1, pp. 1–22. DOI: 10.1111/j.2517-6161.1977.tb01600.x. URL: https://academic.oup.com/jrsssb/article/39/1/1/7027539 (visited on 04/19/2025).

[23] Dickey, D. A. and W. A. Fuller (1979): Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* **74**(366). Number: 366a, pp. 427–431. DOI: 10.1080/01621459.1979.10482531. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10482531 (visited on 02/15/2025).

[24] Dittmar, D. (2013): *Slice Sampling*. Lecture Notes. URL: https://www.ias.informatik.tu-darmstadt.de/uploads/Teaching/RobotLearningSeminar/Dittmar_RLS_2013.pdf.

[25] Durbin, J. (1960): The Fitting of Time-Series Models. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute* **28**(3). Number: 3, p. 233. DOI: 10.2307/1401322. URL: https://www.jstor.org/stable/1401322?origin=crossref (visited on 04/06/2025).

[26] Enders, W. (2015): *Applied econometric time series*. Fourth edition. Hoboken, NJ: Wiley. 485 pp.

[27] Fauriat, W., C. Mattrand, N. Gayton, and A. Beakou (2015): An Application of Stochastic Simulation to the Study of the Variability of Road Induced Fatigue Loads. *Procedia Engineering* **133**, pp. 631–645. DOI: 10.1016/j.proeng.2015.12.643. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877705815045622 (visited on 06/10/2025).

[28] Fletcher, R. (2000): *Practical Methods of Optimization*. 1st ed. Wiley. doi: 10.1002/9781118723203. url: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118723203 (visited on 02/14/2025).

[29] Foss, S., D. Korshunov, and S. Zachary (2011): *An Introduction to Heavy-Tailed and Subexponential Distributions*. Vol. 38. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer New York. doi: 10.1007/978-1-4419-9473-8. url: https://link.springer.com/10.1007/978-1-4419-9473-8 (visited on 01/07/2025).

[30] Freimer, M., G. Kollia, G. S. Mudholkar, and C. T. Lin (1988): A study of the generalized Tukey lambda family. *Communications in Statistics - Theory and Methods* **17**(10). Number: 10, pp. 3547–3567. doi: 10.1080/03610928808829820. url: http://www.tandfonline.com/doi/abs/10.1080/03610928808829820 (visited on 01/27/2025).

[31] Garber, N. J. and L. A. Hoel (1999): *Traffic and highway engineering*. 2nd ed., rev. print. Pacific Grove, CA: PWS Pub. 1118 pp.

[32] Gazis, D. C., R. Herman, and R. W. Rothery (1961): Nonlinear Follow-the-Leader Models of Traffic Flow. *Operations Research* **9**(4). Number: 4, pp. 545–567. doi: 10.1287/opre.9.4.545. url: https://pubsonline.informs.org/doi/10.1287/opre.9.4.545 (visited on 06/03/2025).

[33] Gelman, A. (2014): *Bayesian data analysis*. Third edition. Chapman & Hall/CRC texts in statistical science. Boca Raton: CRC Press. 661 pp.

[34] Geman, S. and D. Geman (1984): Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(6). Number: 6, pp. 721–741. doi: 10.1109/TPAMI.1984.4767596. url: http://ieeexplore.ieee.org/document/4767596/ (visited on 02/27/2025).

[35] Ghanem, R. G. and P. D. Spanos (1991): Stochastic Finite Element Method: Response Statistics. In: *Stochastic Finite Elements: A Spectral Approach*. Book Authors: _:n2035. New York, NY: Springer New York, pp. 101–119. doi: 10.1007/978-1-4612-3094-6_4. url: http://link.springer.com/10.1007/978-1-4612-3094-6_4 (visited on 06/14/2024).

[36] Ghanem, R., D. Higdon, and H. Owhadi, eds. (2017): *Handbook of Uncertainty Quantification*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-12385-1. url: http://link.springer.com/10.1007/978-3-319-12385-1 (visited on 09/01/2022).

[37] Giles, M. B. (2008): Multilevel Monte Carlo Path Simulation. *Operations Research* **56**(3). Number: 3, pp. 607–617. doi: 10.1287/opre.1070.0496. url: https://pubsonline.informs.org/doi/10.1287/opre.1070.0496 (visited on 04/06/2025).

[38] Greenberg, H. (1959): An Analysis of Traffic Flow. *Operations Research* **7**(1). Number: 1, pp. 79–85. doi: 10.1287/opre.7.1.79. url: https://pubsonline.informs.org/doi/10.1287/opre.7.1.79 (visited on 09/22/2024).

[39] Gudehus, H., H. Zenner, and Verein zur Förderung der Forschung und der Anwendung von Betriebsfestigkeitskenntnissen in der Eisenhüttenindustrie, eds. (2007): *Leitfaden für eine Betriebsfestigkeitsrechnung: Empfehlungen zur Lebensdauerabschätzung von Maschinenbauteilen*. 4. korrigierte Aufl., unveränd. Nachdr. OCLC: 254389562. Düsseldorf: Verl. Stahleisen.

[40] Haibach, E. (2006): *Betriebsfestigkeit*. VDI-Buch. Berlin/Heidelberg: Springer-Verlag. doi: 10.1007/3-540-29364-7. url: http://link.springer.com/10.1007/3-540-29364-7 (visited on 11/10/2021).

[41] Hanschmann, D., E. Schelke, and J. Zamow (1994): *Rechnerisches mehraxiales Betriebsfestigkeitsvorhersage-Konzept für die Dimensionierung von KFZ-Komponenten in der frühen Konstruktionsphase*. VDI-Berichte Nr. 1153. Issue: Nr. 1153, pp. 89–112.

[42] Hastings, W. K. (1970): Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**(1). Number: 1, pp. 97–109. doi: 10.1093/biomet/57.1.97. url: https://academic.oup.com/biomet/article/57/1/97/284580 (visited on 02/27/2025).

[43] HERE Global B.V. (2025): *Routing API*. Version v8.

[44] Herman, R., E. W. Montroll, R. B. Potts, and R. W. Rothery (1959): Traffic Dynamics: Analysis of Stability in Car Following. *Operations Research* **7**(1). Number: 1, pp. 86–106. doi: 10.1287/opre.7.1.86. url: https://pubsonline.informs.org/doi/10.1287/opre.7.1.86 (visited on 06/03/2025).

[45] Homma, T. and A. Saltelli (1996): Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety* **52**(1). Number: 1, pp. 1–17. doi: 10.1016/0951-8320(96)00002-6. url: https://linkinghub.elsevier.com/retrieve/pii/0951832096000026 (visited on 03/05/2025).

[46] Hück, M., L. Thrainer, and W. Schütz (1980): *Berechnung von Wöhlerlinien für Bauteile aus Stahl, Stahlguß und Grauguß*. VDEh-Vorhaben ABF 11. Issue: ABF 11. Düsseldorf.

[47] Hyndman, R. J. and Y. Fan (1996): Sample Quantiles in Statistical Packages. *The American Statistician* **50**(4). Number: 4, pp. 361–365. doi: 10.1080/00031305.1996.10473566. url: http://www.tandfonline.com/doi/abs/10.1080/00031305.1996.10473566 (visited on 02/26/2025).

[48] INRIX (2019): *Floating Car Data*. url: https://inrix.com/.

[49] INRIX (2022): *U.S. Signals Scorecard*. url: https://inrix.com/signals-scorecard/.

[50] International Organization for Standardization (2019): *Calculation of load capacity of spur and helical gears — Part 6: Calculation of service life under variable load*. Version Second edition 2019-11. Issue: ISO 6336-6 ISO 6336-6. Geneva.

[51] Kantorovich, L. V. (1960): Mathematical Methods of Organizing and Planning Production. *Management Science* **6**(4). Number: 4, pp. 366–422. doi: 10.1287/mnsc.6.4.366. url: https://pubsonline.informs.org/doi/10.1287/mnsc.6.4.366 (visited on 10/03/2024).

[52] Karian, Z. A. and E. J. Dudewicz (2000): *Fitting Statistical Distributions*. 0th ed. Chapman and Hall/CRC. doi: 10.1201/9781420038040. url: https://www.taylorfrancis.com/books/9781420038040 (visited on 04/06/2025).

[53] Kass, R., L. Tierney, and J. Kadane (1991): Laplace's method in Bayesian analysis. In: *Contemporary Mathematics*. N. Flournoy and R. K. Tsutakawa eds. Vol. 115. Providence, Rhode Island: American Mathematical Society, pp. 89–135. doi: 10.1090/conm/115/07. url: http://www.ams.org/conm/115/ (visited on 03/10/2025).

[54] Kloek, T. and H. K. van Dijk (1978): Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo. *Econometrica* **46**(1). Number: 1, p. 1. doi: 10.2307/1913641. url: https://www.jstor.org/stable/1913641?origin=crossref (visited on 02/26/2025).

[55] Köder, M. (2014): *Schwingfestigkeitsnachweis für innendruckbelastete Common-Rail-Bauteile aus 100Cr6 unter Berücksichtigung hochzyklischer Betriebsbeanspruchungen*. PhD thesis. Siegen: Universität Siegen.

[56] Kucherenko, S., D. Albrecht, and A. Saltelli (2015): Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. doi: 10.48550/ARXIV.1505.02350. url: https://arxiv.org/abs/1505.02350 (visited on 02/26/2025).

[57] Le Maître, O. P., O. M. Knio, H. N. Najm, and R. G. Ghanem (2001): A Stochastic Projection Method for Fluid Flow. *Journal of Computational Physics* **173**(2). Number: 2, pp. 481–511. doi: 10.1006/jcph.2001.6889. url: https://linkinghub.elsevier.com/retrieve/pii/S0021999101968895 (visited on 03/06/2025).

[58] Li, Z.-H., S.-T. Zheng, R. Jiang, J.-F. Tian, K.-X. Zhu, and R. Di Pace (2022): Empirical and simulation study on traffic oscillation characteristic using floating car data. *Physica A: Statistical Mechanics and its Applications* **605**, p. 127973. doi: 10.1016/j.physa.2022.127973. url: https://linkinghub.elsevier.com/retrieve/pii/S0378437122006136 (visited on 11/30/2023).

[59] Lighthill, M. J. and G. B. Whitham (1955): On kinematic waves II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **229**(1178). Number: 1178, pp. 317–345.

DOI: 10.1098/rspa.1955.0089. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1955.0089 (visited on 06/10/2025).

[60] Liu, J. and H. Zenner (1995): Berechnung von Beuteilwöhlerlinien unter Berücksichtigung der statistischen und spannungsmechanischen Stützziffer. *Materialwissenschaft und Werkstofftechnik* **26**, pp. 14–21.

[61] Ljung, G. M. and G. E. P. Box (1978): On a measure of lack of fit in time series models. *Biometrika* **65**(2). Number: 2, pp. 297–303. DOI: 10.1093/biomet/65.2.297. URL: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/65.2.297 (visited on 02/15/2025).

[62] Lüthen, N., S. Marelli, and B. Sudret (2023): A spectral surrogate model for stochastic simulators computed from trajectory samples. *Computer Methods in Applied Mechanics and Engineering* **406**, p. 115875. DOI: 10.1016/j.cma.2022.115875. URL: https://linkinghub.elsevier.com/retrieve/pii/S0045782522008313 (visited on 01/29/2025).

[63] Mackay, D. (1998): *Introduction to Monte Carlo Methods*. Lecture Notes. URL: https://www.cs.cmu.edu/~motionplanning/papers/sbp_papers/kalman/intro_montecarlo.pdf.

[64] McClarren, R. G. (2018): *Uncertainty Quantification and Predictive Computational Science: A Foundation for Physical Scientists and Engineers*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-99525-0. URL: http://link.springer.com/10.1007/978-3-319-99525-0 (visited on 09/17/2022).

[65] McCleary, R. and R. Hay (1980): *Applied time series analysis for the social sciences*. Beverly Hills, Calif: Sage Publications. 331 pp.

[66] McElreath, R. (2020): *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd ed. Chapman and Hall/CRC. DOI: 10.1201/9780429029608. URL: https://www.taylorfrancis.com/books/9780429642319 (visited on 03/10/2025).

[67] McKay, M. D., R. J. Beckman, and W. J. Conover (1979): A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **21**(2). Number: 2, p. 239. DOI: 10.2307/1268522. URL: https://www.jstor.org/stable/1268522?origin=crossref (visited on 01/17/2025).

[68] Meeker, W. Q., L. A. Escobar, and F. G. Pascual (2022): *Statistical methods for reliability data*. Second edition. Wiley series in probability and statistics. Hoboken, NJ: Wiley. 704 pp.

[69] Metropolis, N. and S. Ulam (1949): The Monte Carlo Method. *Journal of the American Statistical Association* **44**(247). Number: 247, pp. 335–341. DOI: 10.1080/01621459.1949.10483310. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1949.10483310 (visited on 02/25/2025).

[70] Miner, M. A. (1945): Cumulative Damage in Fatigue. *Journal of Applied Mechanics* **12**(3). Number: 3, A159–A164. doi: 10.1115/1.4009458. url: https://asmedigitalcollection.asme.org/appliedmechanics/article/12/3/A159/1103202/Cumulative-Damage-in-Fatigue (visited on 05/17/2025).

[71] Moran, P. A. and P. Whittle (1951): Hypothesis Testing in Time Series Analysis. *Journal of the Royal Statistical Society. Series A (General)* **114**(4). Number: 4, p. 579. doi: 10.2307/2981095. url: https://www.jstor.org/stable/10.2307/2981095?origin=crossref (visited on 01/04/2025).

[72] Morr, A. and N. Boers (2024): Detection of Approaching Critical Transitions in Natural Systems Driven by Red Noise. *Physical Review X* **14**(2). Number: 2, p. 021037. doi: 10.1103/PhysRevX.14.021037. url: https://link.aps.org/doi/10.1103/PhysRevX.14.021037 (visited on 08/30/2024).

[73] Nagel, J. B. (2017): *Bayesian techniques for inverse uncertainty quantification*. PhD thesis. Zürich: ETH Zürich.

[74] Neal, R. M. (2003): Slice sampling. *The Annals of Statistics* **31**(3). Number: 3. doi: 10.1214/aos/1056562461. url: https://projecteuclid.org/journals/annals-of-statistics/volume-31/issue-3/Slice-sampling/10.1214/aos/1056562461.full (visited on 01/28/2025).

[75] Oberkampf, W. L. and C. J. Roy (2010): *Verification and Validation in Scientific Computing*. 1st ed. Cambridge University Press. doi: 10.1017/CBO9780511760396. url: https://www.cambridge.org/core/product/identifier/9780511760396/type/book (visited on 09/12/2022).

[76] Oertel, R. (2010): *Delay Based Signal Control for an Isolated Intersection*. Young European Arena of Research 2010. url: https://elib.dlr.de/64360/1/YEAR2010_FCD2TrafficLights.pdf.

[77] Ramberg, J. S. and B. W. Schmeiser (1974): An approximate method for generating asymmetric random variables. *Communications of the ACM* **17**(2). Number: 2, pp. 78–82. doi: 10.1145/360827.360840. url: https://dl.acm.org/doi/10.1145/360827.360840 (visited on 04/06/2025).

[78] Rasmussen, C. E. and C. K. I. Williams (2005): *Gaussian Processes for Machine Learning*. The MIT Press. doi: 10.7551/mitpress/3206.001.0001. url: https://direct.mit.edu/books/book/2320/Gaussian-Processes-for-Machine-Learning (visited on 03/06/2025).

[79] Rennert, R. and Forschungskuratorium Maschinenbau, eds. (2013): *Analytical strength assessment of components: made of steel, cast iron and aluminum materials in mechanical engineering ; FKM Guideline*. 6th, rev. ed. OCLC: 931450321. Frankfurt am Main: VDMA-Verl. 232 pp.

[80] Richards, P. I. (1956): Shock Waves on the Highway. *Operations Research* **4**(1). Number: 1, pp. 42–51. doi: 10.1287/opre.4.1.42. url: https://pubsonline.informs.org/doi/10.1287/opre.4.1.42 (visited on 06/10/2025).

[81] Ripley, B. D. (1996): *Pattern Recognition and Neural Networks*. 1st ed. Cambridge University Press. doi: 10.1017/CBO9780511812651. url: https://www.cambridge.org/core/product/identifier/9780511812651/type/book (visited on 03/06/2025).

[82] Saltelli, A. (2002): Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* **145**(2). Number: 2, pp. 280–297. doi: 10.1016/S0010-4655(02)00280-1. url: https://linkinghub.elsevier.com/retrieve/pii/S0010465502002801 (visited on 01/24/2025).

[83] Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola (2010): Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* **181**(2). Number: 2, pp. 259–270. doi: 10.1016/j.cpc.2009.09.018. url: https://linkinghub.elsevier.com/retrieve/pii/S0010465509003087 (visited on 01/24/2025).

[84] Samaras, E., M. Shinzuka, and A. Tsurui (1985): ARMA Representation of Random Processes. *Journal of Engineering Mechanics* **111**(3). Number: 3, pp. 449–461. doi: 10.1061/(ASCE)0733-9399(1985)111:3(449). url: https://ascelibrary.org/doi/10.1061/%28ASCE%290733-9399%281985%29111%3A3%28449%29 (visited on 02/10/2025).

[85] Sandmann, K., T. Leyendecker, M. Burger, and M. Speckert (2018): Ableitung von feldrelevanten Lastkollektiven mittels Stochastischer Verkehrssimulation. In: *DVM-Bericht 145*. 45. Tagung des DVM-Arbeitskreises Betriebsfestigkeit. Berlin. url: https://publica.fraunhofer.de/handle/publica/403187.

[86] Schramm, D., M. Hiller, and R. Bardini (2018): *Vehicle Dynamics*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-54483-9. url: http://link.springer.com/10.1007/978-3-662-54483-9 (visited on 04/05/2025).

[87] Schwarz, G. (1978): Estimating the Dimension of a Model. *The Annals of Statistics* **6**(2). Number: 2. doi: 10.1214/aos/1176344136. url: https://projecteuclid.org/journals/annals-of-statistics/volume-6/issue-2/Estimating-the-Dimension-of-a-Model/10.1214/aos/1176344136.full (visited on 01/06/2025).

[88] Šmídl, V. and A. Quinn (2006): *The Variational Bayes Method in Signal Processing*. Signals and Communication Technology. Berlin/Heidelberg: Springer-Verlag. doi: 10.1007/3-540-28820-1. url: http://link.springer.com/10.1007/3-540-28820-1 (visited on 03/10/2025).

[89] Sobol, I. (1993): Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modelling and Computational Experiments*, **1**(4). Number: 4, pp. 407–414.

[90] Sobol, I. (2001): Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* **55**(1). Number: 1-3, pp. 271–280. doi: 10.1016/S0378-4754(00)00270-6. url: https://linkinghub.elsevier.com/retrieve/pii/S0378475400002706 (visited on 05/22/2024).

[91] Soize, C. and R. Ghanem (2004): Physical Systems with Random Uncertainties: Chaos Representations with Arbitrary Probability Measure. *SIAM Journal on Scientific Computing* **26**(2). Number: 2, pp. 395–410. doi: 10.1137/S1064827503424505. url: http://epubs.siam.org/doi/10.1137/S1064827503424505 (visited on 03/07/2025).

[92] Soize, C. (2017): *Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering*. Vol. 47. Interdisciplinary Applied Mathematics. Cham: Springer International Publishing. doi: 10.1007/978-3-319-54339-0. url: http://link.springer.com/10.1007/978-3-319-54339-0 (visited on 08/25/2022).

[93] Speckert, M., K. Dreßler, N. Ruf, T. Halfmann, and S. Polanski (2014): The Virtual Measurement Campaign (VMC) concept - A methodology for georeferenced description and evaluation of environmental conditions for vehicle loads and energy efficiency. In: *Commercial vehicle technology 2014*. Commercial Vehicle Technology Symposium (CVT). url: https://publica.fraunhofer.de/handle/publica/385532.

[94] Speckert, M. and M. Burger (2020): *VMC speed profile calculation - Documentation of approach and parameters*. Operation manual.

[95] Steihaug, T. (1983): The Conjugate Gradient Method and Trust Regions in Large Scale Optimization. *SIAM Journal on Numerical Analysis* **20**(3). Number: 3, pp. 626–637. doi: 10.1137/0720042. url: http://epubs.siam.org/doi/10.1137/0720042 (visited on 03/10/2025).

[96] Stein, M. L. (1999): *Interpolation of Spatial Data*. Springer Series in Statistics. New York, NY: Springer New York. doi: 10.1007/978-1-4612-1494-6. url: http://link.springer.com/10.1007/978-1-4612-1494-6 (visited on 07/20/2025).

[97] Sullivan, T. (2015): *Introduction to Uncertainty Quantification*. Vol. 63. Texts in Applied Mathematics. Cham: Springer International Publishing. doi: 10.1007/978-3-319-23395-6. url: http://link.springer.com/10.1007/978-3-319-23395-6 (visited on 08/25/2022).

[98] Torre, E., S. Marelli, P. Embrechts, and B. Sudret (2019): Data-driven polynomial chaos expansion for machine learning regression. *Journal of Computational Physics*

388, pp. 601–623. DOI: 10.1016/j.jcp.2019.03.039. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021999119302220 (visited on 03/07/2025).

[99] Treiber, M., A. Hennecke, and D. Helbing (2000): Congested traffic states in empirical observations and microscopic simulations. *Physical Review E* **62**(2). Number: 2, pp. 1805–1824. DOI: 10.1103/PhysRevE.62.1805. URL: https://link.aps.org/doi/10.1103/PhysRevE.62.1805 (visited on 06/19/2025).

[100] Treiber, M. and A. Kesting (2013): *Traffic Flow Dynamics: Data, Models and Simulation*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-32460-4. URL: https://link.springer.com/10.1007/978-3-642-32460-4 (visited on 09/21/2024).

[101] Vapnik, V. N. (1997): The Support Vector method. In: *Artificial Neural Networks — ICANN'97*. W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud eds. Red. by G. Goos, J. Hartmanis, and J. van Leeuwen. Vol. 1327. Series Editors: _:n2534 Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 261–271. DOI: 10.1007/BFb0020166. URL: http://link.springer.com/10.1007/BFb0020166 (visited on 03/06/2025).

[102] Wen, X., L. Fu, T. Fu, J. Keung, and M. Zhong (2021): Driver Behavior Classification at Stop-Controlled Intersections Using Video-Based Trajectory Data. *Sustainability* **13**(3). Number: 3, p. 1404. DOI: 10.3390/su13031404. URL: https://www.mdpi.com/2071-1050/13/3/1404 (visited on 04/10/2025).

[103] Westermann-Friedrich, A. and H. Zenner (1988): *Zählverfahren zur Bildung von Kollektiven aus Zeitfunktionen - Vergleich der verschiedenen Verfahren und Beispiele*. Issue: 0/14 Merkblatt des AK Lastkollektive der FVA. Frankfurt.

[104] Woldeamanuel, M. (2012): Stopping behavior of drivers at stop-controlled intersections : compositional and contextual analysis. *Journal of the Transportation Research Forum* **51**(3). Number: 3.

[105] Wooldridge, J. M. (2013): *Introductory econometrics: a modern approach*. 5th ed. Mason, Ohio: South-Western, Cengage Learning. 881 pp.

[106] Xiu, D. and G. E. Karniadakis (2002): The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing* **24**(2). Number: 2, pp. 619–644. DOI: 10.1137/S1064827501387826. URL: http://epubs.siam.org/doi/10.1137/S1064827501387826 (visited on 08/18/2024).

[107] Yule, G. U. (1927): VII. On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **226**(636). Number: 636-646, pp. 267–298. DOI: 10.1098/rsta.1927.0007. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.1927.0007 (visited on 02/11/2025).

[108] Zenner, H. (1994): *Lebensdauervorhersage im Automobilbau*. VDI-Berichte Nr. 1153. Issue: Nr. 1153, pp. 29–42.

[109] Zhu, X. and B. Sudret (2020): Replication-based emulation of the response distribution of stochastic simulators using generalized lambda distributions. *International Journal for Uncertainty Quantification* **10**(3). Number: 3, pp. 249–275. doi: 10.1615/Int.J.UncertaintyQuantification.2020033029. url: http://www.dl.begellhouse.com/journals/52034eb04b657aea,5047a058688097ba,17fb6a69404aa954.html (visited on 03/09/2025).

[110] Zhu, X. and B. Sudret (2021): Emulation of Stochastic Simulators Using Generalized Lambda Models. *SIAM/ASA Journal on Uncertainty Quantification* **9**(4). Number: 4, pp. 1345–1380. doi: 10.1137/20M1337302. url: https://epubs.siam.org/doi/10.1137/20M1337302 (visited on 11/06/2024).

[111] Zhu, X. and B. Sudret (2021): Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. *Reliability Engineering & System Safety* **214**, p. 107815. doi: 10.1016/j.ress.2021.107815. url: http://arxiv.org/abs/2005.01309 (visited on 01/20/2025).

[112] Zhu, X. and B. Sudret (2023): Stochastic polynomial chaos expansions to emulate stochastic simulators. *International Journal for Uncertainty Quantification* **13**(2). Number: 2, pp. 31–52. doi: 10.1615/Int.J.UncertaintyQuantification.2022042912. url: https://www.dl.begellhouse.com/journals/52034eb04b657aea,629ce0e12931a3e6,49d88de9790d7920.html (visited on 01/29/2025).

Publications

[113] Strähle, P., S. Wolff-Vorbeck, T. Leyendecker, and C. Proppe (2024): Modeling Velocity Fluctuations for the Derivation of Automotive Component Loads. In: *ECCOMAS 2024*. 9th European Congress on Computational Methods in Applied Sciences and Engineering. URL: https://www.scipedia.com/public/Strahle_et_al_2024a.

[114] Strähle, P., S. Wolff-Vorbeck, T. Leyendecker, and C. Proppe (2025): Uncertainty Quantification for the Simulation of Automotive Component Loads using Stochastic Surrogate Models. In: *UNCECOMP 2025*. 6th ECCOMAS International Conference on Uncertainty Quantification in Computational Sciences and Engineering. URL: https://2025.uncecomp.org/proceedings/pdf/21214.pdf.

Supervised Theses

[115] Blatt, L. (2024): *Modellierung von Verkehrseinflüssen in der virtuellen Lastgenerierung*. (unpublished). Master thesis. Hochschule Konstanz Technik, Wirtschaft und Gestaltung. 83 pp.

[116] Le, T.-H. B. (2024): *Datenbasierte Modellierung von Fahrzeug-Geschwindigkeitsprofilen*. (unpublished). Master thesis. Karlsruher Institut für Technologie. 101 pp.