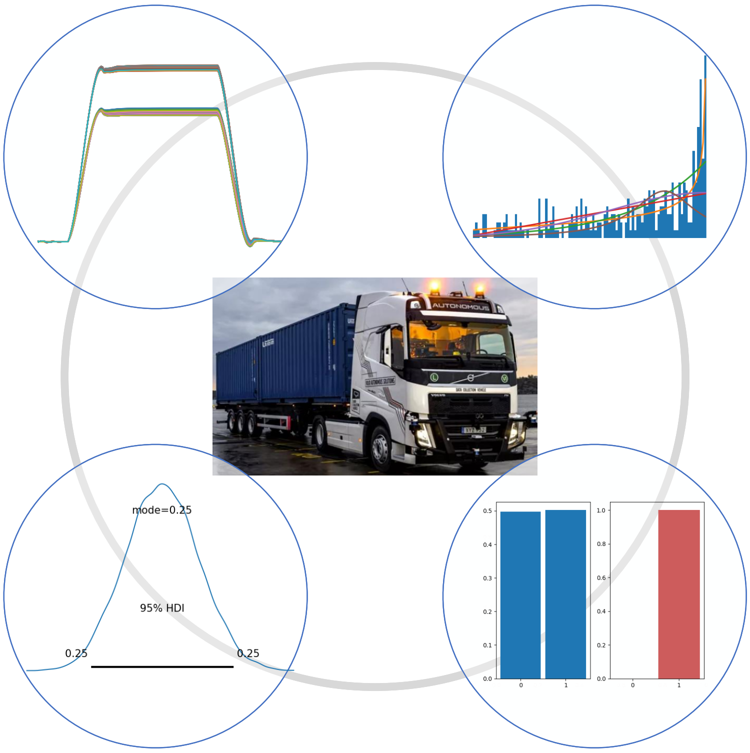
CHALMERS
UNIVERSITY OF TECHNOLOGY



Validation of Tractor-semitrailer Vehicle Model based on Bayesian Hypothesis Testing

Master's thesis in Automotive Engineering

ATHANASIA MARIA DINEFF

Department of Mechanics and Maritime Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021

MASTER'S THESIS 2021:78

Validation of Tractor-semitrailer Vehicle Model based on Bayesian Hypothesis Testing

ATHANASIA MARIA DINEFF

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021

Validation of Tractor-semitrailer Vehicle Model based on Bayesian Hypothesis Testing
ATHANASIA MARIA DINEFF

Supervisor: Peter Nilsson, Volvo Autonomous Solutions, Thorsten Helfrich, Volvo Autonomous Solutions, Mohamed Takkoush, FCC, Volvo Autonomous Solutions
Examiner: Mathias Lidberg, Department - Mechanics and Maritime Sciences

Master's Thesis 2021:78
Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Circular "flowchart" of the validation framework, starting from top left with the vehicle models' simulations, continuing to top right where distributions are fitted to the relative difference between the two models and finally ending at the bottom with the results from the two Bayesian hypothesis testing approaches. Located in the center of the flowchart is the automated truck of VAS (Source: http://www.volvoautonomoussolutions.com/).

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Validation of Tractor-semitrailer vehicle model based on Bayesian Hypothesis Testing
ATHANASIA MARIA DINEFF
Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology

Abstract

The development of automated vehicles is the future of the automotive industry, and it has seen a rapid advance over the past decade. The design of automated vehicles relies on vehicle models, which predict their responses and control their behavior. This thesis presents a vehicle model validation framework to evaluate the validity of a simple (abstract) vehicle model against a complex (implementation) model. The aim is to quantify the validity of the abstract model and determine its suitability for an application. The validation procedure uses Bayesian hypothesis testing. First, two hypotheses are stated; the null hypothesis, which supports the abstract model's validity, and the alternative, which rejects the model. Two approaches to Bayesian hypothesis testing are then studied. The first approach estimates the posterior distributions of the parameters of interest and evaluates whether the credible interval includes the null hypothesis. The second approach relies on Bayes factors, which compare the two hypotheses and indicate the most probable one.

The methods presented are applied on a tractor-semitrailer combination under the driving context of city driving. The validation framework is evaluated for three cases of the abstract model. The first case refers to the calibrated model. The second and third cases concern an incorrectly tuned parameter, with double and half the calibrated value, respectively. The outcome is that the abstract model is deemed valid when calibrated, whereas it is invalid when a parameter is incorrectly tuned. The thesis also discusses the comparison between the two proposed validation approaches.

Keywords: Bayesian hypothesis testing, Model validation, Vehicle modelling, Tractor-semitrailer, Bayes factor, Model comparison, Posterior summary

Acknowledgements

This thesis would not be feasible without the support and help of my supervisors. First and foremost, I would like to express my deep and sincere gratitude towards my thesis examiner and supervisor, Mathias Lidberg. Your invaluable advice and continuous support throughout the course of the thesis made this work possible. You always encouraged me to continue and strive for the best. Thank you for teaching me your way of thinking and sharing your methods and tricks. I want to offer my special thanks to Mohamed Takkoush for being supportive and engaged in every step of this project. Your patience, enthusiasm, feedback, and valuable insights through critical questioning helped me refine my work. I am deeply grateful to Peter Nilsson for his guidance and his inputs, especially regarding vehicle modelling and programming. I would also like to extend my deepest gratitude to Thorsten Helfrich for his insightful comments and suggestions. It has been a pleasure working with you. You made me feel like a member of VAS, even though the circumstances did not allow us to meet and work in person.

A heartfelt thanks goes to all my friends and family. To Giannis, Panos and Valeria - thank you for your moral support and for helping me overcome the obstacles I faced. I would also like to thank my brother for his assistance in the statistical analysis and for constantly believing in me.

Finally, there are no words to express my gratitude to my parents for their unwavering support and encouragement every single day. Your love, dedication, and profound belief in my abilities made me feel confident and motivated. Thank you for giving me the opportunity to study in Sweden, and I only hope to make you proud.

Athanasia Maria Dineff, Gothenburg, October 2021

Nomenclature

Acronyms

A.M.	Abstract Model
CAE	Computer-Aided Engineering
CDF	Cumulative Density Function
CI	Confidence Interval
CoG	Center of Gravity
DOF	Degree of Freedom
FEM	Finite Element Method
HDI	Highest Density Interval
HMC	Hamiltonian Monte Carlo
HPD	Highest-Posterior Density
I.M.	Implementation Model
KDE	Kernel Density Estimate
M.D.	Model Comparison (of statistical models)
MCMC	Markov Chain Monte Carlo
NUTS	No-U-Turn Sampler
P.E.	Parameter Estimation
PDF	Probability Density Function
PPL	Probabilistic Programming Language
QoI	Quantity of Interest
RMS	Root Mean Square
ROPE	Region of Practical Equivalence
TC1	Test Case 1
TC2	Test Case 2
TC3	Test Case 3
VAS	Volvo Autonomous Solutions
VTM	Virtual Truck Model

Greek Characters

$\ddot{\psi}_i$	Yaw acceleration of unit i
$\Delta\psi$	Relative yaw angle between units
δ	Steering angle
$\dot{\psi}_i$	Yaw rate of unit i

ϵ_l	Predefined lower limit of an unknown parameter's predicted value
ϵ_u	Predefined upper limit of an unknown parameter's predicted value
ϵ_x	Upper limit of relative difference of metric x
κ	Bayesian confidence
ω	Angular velocity
ψ	Yaw angle
θ	Unobserved/unknown statistical parameter
θ_x	Unknown statistical parameters of relative difference of metric x
$\theta_{l,x}$	Location parameter of likelihood describing relative difference of metric x

Latin Alhanumerics

$\boldsymbol{a}$	Vehicle acceleration vector
$\boldsymbol{F_{ec,i}}$	Coupling force of unit i in global coordinates
$\boldsymbol{F_{vc,i}}$	Coupling force of unit i in unit coordinates
$\boldsymbol{F_{vw,ij}}$	Wheel force on axle j of unit i in unit coordinates
$\boldsymbol{F_{w,ij}}$	Wheel force on axle j of unit i in wheel coordinates
$\boldsymbol{r_j}$	Moment arm from center of gravity to axle j
$\boldsymbol{v}$	Vehicle velocity vector in unit coordinates
$\tilde{y}$	Unknown observable quantity
A_δ	Maximum amplitude of steering angle
B_{01}	Bayes factor
C_y	Lateral cornering stiffness
C_{cy}	Lateral cornering coefficient
d_x	Relative difference of metric x
df	Statistical degrees of freedom
F_{zi}	Normal forces on unit i
H	Hypothesis
H_0	Null hypothesis
H_1	Alternative hypothesis
$I_{zz,i}$	Moment of inertia about z-axis
K	Gain of P controller
$l_{c,ij}$	Distance of coupling point j of unit i from its center of gravity
l_{if}	Distance of front axle of unit i from its center of gravity
l_{ij}	Distance of axle j of unit i from its center of gravity
l_{ir}	Distance of rearmost axle of unit i from its center of gravity
$LSOT$	Low speed off-tracking
M	Statistical model
m	Model index
M_f	Final value
m_i	Mass of unit i
M_p	Peak value
M_x	Statistical model of metric x
$m_{add,i}$	Added mass on unit i
M_{OV}	Maximum percent overshoot

$M_{x,m}$	Distinct statistical model of metric x, part of the hierarchical model M
$M_{z,ij}$	Moment about z-axis exerted by axle j on center of gravity of unit i
MAe	Mean absolute error
$p(\cdot \mid \cdot)$	Conditional probability density
$p(\cdot)$	Marginal probability density
$p(\cdot, \cdot)$	Joint probability density
$p(\theta \mid y)$	Posterior distribution
$p(\theta)$	Prior distribution
$p(y \mid \theta)$	Likelihood
$p(y)$	Evidence
$payload_F$	Payload fraction
Pe	Percentage error
$Pr(\cdot)$	Probability
R	Path radius
$R_{\Delta\psi}$	Rotation matrix about the unit fixed reference frame
R_{δ}	Rotation matrix about the wheel fixed reference frame
R_{ψ}	Rotation matrix about the global fixed reference frame
Re	Relative error
$RMSe$	Root mean square error
RWA_{a_y}	Rearward amplification
s_y	Lateral tire slip
SSe	Sum of squared errors
t	Time
t_p	Peak time
t_r	Rise time
t_s	Settling time
v_x	Longitudinal velocity
v_y	Lateral velocity
x_A	Response of abstract vehicle model
x_I	Response of implementation vehicle model
y	Observed data

Contents

List of Figures

List of Tables

1

Introduction

Automated vehicles are the focus of interest for the automotive industry. In the last decade, more and more manufactures and technology companies have been involved in projects for developing fully automated vehicles [1], and the first samples have already made their appearance. Both in the transportation and construction sector, automation can play a vital role in energy efficiency, sustainability, safety, and the economy in general [2]. The automated vehicles should be able to handle all possible situations, ensuring the safety of all road users. This leads to more and more complex vehicle systems. Therefore, the control and safety functions deployed as well as their effective testing are currently a major area of research and development.

1.1 Modelling in Vehicle automation

Automated vehicles pose a challenge in designing functions and systems that ensure their safe operation under various conditions. The driving task is dynamic and while in manual driving the driver can quickly adapt to changes in the driving environment and respond to different situations, especially the critical ones, the automated driving must rely on the vehicle systems to do so. These systems and control functions are designed with the help of vehicle models.

The dynamic vehicle models are the cornerstone in research and development of vehicles in the automotive industry. Virtual testing of the vehicle performance with computer models and CAE tools has reduced the need of prototypes and testing and therefore has contributed to the reduction of the time and cost of the evaluation of the vehicles [3], [4]. Vehicle models are used as virtual prototypes during the development phase, when the developers can evaluate the vehicle behavior and design, test and optimize the control functions and other systems [5].

The vehicle models are an area that has been investigated thoroughly. Different models have been developed, depending on the application, that can vary in terms of complexity and number of parameters. They can range from a 3 DOF bicycle model to a multibody system model with hundreds of degrees of freedom (DOFs) and variables [6]. The simple models have less parameters and are less computationally demanding, making them the perfect candidate for online applications. However, they are less accurate when predicting the overall vehicle behavior, under different driving situations. On the other hand, the more complex models are a more accurate representation of the actual vehicle and usually provide better predictions, but they have higher computational costs.

One modelling approach utilizes FEM software, which is known for its high computational demands. The combination of multibody dynamics and FEM for vehicle modelling has also been used. Sakuma created two-dimensional non-linear models of rigid vehicles using a finite element program, and flexible frame modes were included [7], [8].

Many of the recent vehicle models are complex and include many degrees of freedom. They are usually developed using multibody dynamics, such as Adams and TruckSim. A lot of vehicle simulations using Multibody System software are performed even from the early stages of modelling. These models, however, include a lot of details, which are not needed for handling simulations and additionally they are inappropriate when different vehicle configurations are used. According to Gäfvert and Lindgärde [6], they are usually used for structural studies.

In his doctoral thesis [1], Lima uses two different models, one dynamic and one kinematic. The first one is used to model the lateral dynamics, while the latter for the motion planning and the control design. A similar approach is followed by Hoel and Falcone [9] in their paper to design a controller for low speed maneuvering for heavy trucks. In their case, however, the dynamic model is an advanced model, called Virtual Truck Model VTM, developed in Simulink and SimMechanics at Volvo Group Trucks Technology.

Another vehicle model implemented in Simulink was derived by Gäfvert and Lindgärde [6]. Based on a simple 4 DOF model of a tractor-semitrailer combination, they developed a 9DOF model, by including also all the heave, roll and pitch motions to capture the load transfer. Chen and Tomizuka [10] propose a nonlinear complex model to simulate the articulated vehicles, which models the roll, pitch and yaw motions and follows the Lagrangian mechanics approach. They also use this model to derive a linear one, by performing simplifications and linearizations, based on which they will design their controllers.

In their paper, Li and Rakheja [11] propose a control approach to enhance the yaw stability of heavy commercial vehicles, which due to their size and mass center have lower roll and yaw stability limits than the other vehicles. The vehicle model used for the controller design was a simple single-track linear model, which was derived from the linearization of a nonlinear model developed in TruckSim.

It is evident, that due to the reduced number of variables and parameters, the analytical models are still used a lot in modelling the vehicle dynamics. Even though they are a simple representation of a vehicle, they enable a better understanding of the dynamics by the designers.

1.2 Model Validation

The vehicle models must be accurate enough in order to fulfil their intended application. This means that they should predict the actual vehicle behavior, within some acceptable margins of discrepancy, specified for each application. If the behavior of the vehicle is not predicted correctly with sufficient accuracy, then the functions created will not perform accordingly and safety will be compromised. Therefore, the need of validation of the vehicle models is imperative. The vehicle models can be validated against real systems, by using real driving data, or a reference model. The reference model is a vehicle model for which the assumption that its predictions approach reality is valid.

1.2.1 Definition of Validation

Validation is defined as the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. It determines whether the simulation outcome fits accurately enough the behavior of the real system or of the reference model [12]. It is important to note here that there is no such thing as absolute validity and that a model is always an approximation of the reality [13]. The validity of a model concerns its limits and applications. Also, according to Popper [14] a model cannot be validated, exactly as the scientific theories cannot be proven, thus a model can only be invalidated. This means that usually instead of attempting to show that a model is valid, the validation methods focus on proving that the model is not invalid, and therefore the failure to falsify it. A model is falsified by determining that it is invalid for the specific application.

1.2.2 Approaches to Validation

Various validation methods have been proposed and used for validating a simulation model. In the literature both subjective and objective approaches can be found [15]. The distinction between them is that the subjective are qualitative, whereas the objective methods are quantitative and utilize mathematical tools or statistical tests.

One common validation technique is to compare the simulation outputs to the results of other valid models [15]. Other techniques involve the comparison of the simulation outputs behavior to the corresponding distribution of the real systems to determine if they are similar, called event validation [16], or the opinion of people knowledgeable regarding the simulated system, known as face validation. Similar to the latter technique are turing tests, where knowledgeable people are asked to distinguish the model outputs from the system. Historical validation is a technique, where data for a system is collected and part of it is used to build the model and the rest to test if it performs as the real system does. Another technique is the internal validation, which determines the variability of the model under stochastic analysis. The predictive validation evaluates the similarity between the model's predictions and the actual behavior of the system. Tracing can also be used to follow the behavior of different model entities through the model [15].

Authors of [17] follow a qualitative method of graphical comparison between the simulated and experimental mean, in either time or frequency domain. This is the most straightforward and intuitive method, but it can be biased and difficult to extract a lot of conclusions [18], [19]. Therefore, the use of quantitative measures is beneficial. Authors in [17] accompany the graphical comparison with the direct comparison of values, called metrics, including steady state gains, response times, peak response times and percent overshoots from the time domain data, and peak frequency, peak amplitude ratio, bandwidth and phase angle in the frequency domain. Another quantitative method is the root mean square difference between the data (RMS error).

Scoring models are also proposed as a validation method, though they are rarely used. Sargent [15] discredited it, because of its subjectivity, which can be concealed, appearing objective, and that it can lead to overconfidence and accepting a model with flaws. During the validation process, subjective scores are established, which combined determine category scores and a total score for the simulation. If those are greater than some passing scores, the simulation is considered valid.

Another validation technique is the sensitivity analysis or parameter variability. This technique involves the variation of the input or internal parameters to determine their effect on the model's output. The relation between the parameters and the outputs should be the same as the reality and the parameters that have major impact to the output should be made accurate while building the model. It can be either qualitative or quantitative [15], [20]. An example of the latter is in [21], where the authors perform a sensitivity analysis to quantify how the uncertainty of the entire space of input parameters affect the uncertainty in the output values of a hypersonic vehicle model using a variance-based technique. Many authors have worked on quantifying and propagating the uncertainties in the simulation model, and it is considered as a fundamental step during the validation process [22]–[24].

Statistical and probabilistic methods have also been used widely for validation purposes. One category of them are methods that are based on the distance between either the probability density functions (PDFs) or the cumulative density functions (CDFs) between the simulated model and the validation system [19]. Statistical tests are used to quantify the difference, such as the Kolmogorov-Smirnov test which measures the vertical distance between the CDFs [19]. Another validation metric used by Ling and Mahadevan [25] was the area metric, which utilizes the area between the simulation and the validation distribution as the measure of the discrepancy between them.

Hypothesis testing is another approach of quantitative validation, and both conventional/classical and Bayesian methods have been researched. The classical hypothesis test consists of the null hypothesis, which is to be evaluated and a test statistic, which is computed based on the data and used to determine if the null hypothesis is rejected based on the evidence [19], [26]. In all the hypothesis tests, normality is assumed, whereas when this assumption is invalid, either the transformation of the data to normality or the use of the bootstrap method are suggested [19].

Bayesian hypothesis testing combines the classical hypothesis testing methods and the Bayes theorem. The Bayes factor is used to determine if the null hypothesis is accepted or rejected, depending on if it is higher than unity or not, respectively [19], [26]. In [27] Jiang and Mahadevan apply the Bayesian hypothesis testing to a simple structural model. In [28] the probabilistic principal component analysis in conjunction with multivariate Bayesian hypothesis testing is proposed for the validation of dynamic systems.

In the automotive industry and research many of the above methods have been used. Classical methods, such as the graphical comparison and the use of metrics for validity quantification are the most common ones, however the statistical methods and the uncertainty quantification have been of key interest in the last years [22].

1.3 Error measures

Vehicle models can be compared in a deterministic way. In this approach, the models are simulated for a set of vehicle parameters and inputs and their vector-valued responses are compared, in time and/or frequency domain. The comparison of the vehicle responses can be done either qualitatively or quantitatively. The qualitative comparison of the vehicle models is a standard and wide used methodology, which involves graphical comparison between the responses of the models. It is often accompanied though with a form of error measure. There are different ways to express the deviation of the responses. Some common measures are the mean absolute error (MAe)and the Root Mean Square Error ($RMSe$).

The absolute error of the responses of the abstract model x_A and the implementation model x_I is the arithmetic average of the absolute errors $|e_i| = |x_{A,i} - x_{I,i}|$, where $i = 1, .., n$ and n is the size of the data of the response. It is obvious that this measure has the same scale as the data, and thus it is known as a scale-dependent accuracy measure [29]. The RMS error describes the square root of the average difference of the responses. This measure is once again based on a physical quantity and thus has the same scale as the data.

Often, the relative (Re) and the percentage error (Pe) of a metric are reported, to present the deviation in a more obvious way. The relative error is the absolute error of a metric ($|x_A - x_I|$), divided by the absolute value of the metric as calculated by the implementation model (x_I).

The percentage error is the relative error in terms of per 100 [29]. Both of these errors have no units, which makes the comparisons and conclusions easier to be drawn.

Finally, another well-known measure of discrepancy between the data and an estimation model is the sum of squared errors SSe. It is defined as the sum of the squares of the deviation between the data and the predictions of the estimation model, divided by its degrees of freedom *df* [30]. Here, the (statistical) degrees of freedom refer to the number of independent observations in a data sample available to estimate a parameter [31]. A low value of SSe indicates a tight fit of the estimation model to the data. The sum of squared errors is usually used as a goodness-of-fit measure.

1.4 Aim

Designing and developing automated vehicles is complex task, for which the vehicle models play a key role. Different models, that vary in complexity, are used for the designs of control functions. However, the designers are interested in using only one vehicle model for different tasks and functions. Therefore, it is of high importance to develop a validation process in order to determine which model is complex enough to represent the actual dynamics but simple enough to be suitable for online applications.

The aim of this thesis is to establish a framework for validation of a low complexity vehicle model, utilizing Bayesian statistics. The purpose is to quantify the model's validity and determine if it is well suited for certain applications. A process is defined for the validation procedure of an abstract (low complexity) model against an implementation (reference) model, which can be either a higher complexity model or experimental data.

The framework can be divided into three stages. The first step is the definition of a driving context. It is a very important phase of the framework since the choice of the driving context defines the applicability of the model and sets the limits for which it is considered valid. The second step concerns the vehicle modelling. Here the model assumptions and simplifications are made, and the equations of motion are derived. Finally, the model validation using statistical methods is performed, with a focus on Bayesian hypothesis testing approach. The validity of the vehicle model is assessed and quantified based on numerical values, which are the outcome of the statistical validation methods.

The proposed framework is implemented on a single track linear model of a tractor-semitrailer vehicle, for a test case of 90 ° turn under city driving, and Bayesian statistical validation methods are used to determine its validity. Additionally, the validation process is documented in a formal manner for use in other applications.

1.5 Delimitations

The purpose of the thesis is limited to the evaluation of the validity of vehicle models. To that end only minor development of a simple model is done. The high complexity models that are used, were not developed during this project, but they are provided by VAS. The metrics for describing the vehicle performance and the driving context under which the vehicle model is evaluated, are not thoroughly studied, but rather a number of them is rigidly defined and utilized to perform the model validation. Moreover, the vehicle data gathered and used for validation are generated by only varying the payload (additional mass), that can be loaded in the trailer. The study did not research whether this parameter had a significant effect on the vehicle dynamics, or if other parameters have greater influence on the vehicle behavior. Additionally, the abstract model was calibrated manually, without using any automated calibration methods.

As far as the validation is concerned, this study does not cover the comparison of the abstract model to real driving data. Furthermore, the statistical models used in the validation methodology were not optimized. More specifically, only a number of distribution families were available to describe the data. This work did not evaluate the use of mixtures of distribution to obtain the best fit to the data. Finally, a univariate hypothesis testing was used, consequently only one unknown parameter was inferred.

1.6 Thesis Structure

The work in this thesis is structured as follows. Chapter 2 introduces the Bayesian statistics and hypothesis testing theory on which the validation methodology is based. In Chapter 3, the modelling of a single track tractor semi-trailer model is discussed. In addition, the performance metrics used to describe the vehicle behavior and response are presented. Chapter 4 analyzes the structure of the validation methodology, where the Bayesian hypothesis testing approaches from Ch. 2 are applied on the tractor semi-trailer model derived in Ch. 3. In Chapter 5 the results of the research conducted in this thesis are outlined. Finally, in Chapter 6 the conclusions of this work are discussed and summarized, and suggestions for future research regarding the validation methodology are proposed.

2

Bayesian statistics

This chapter introduces the basics of the probability theory used in the thesis. The Bayesian theory is presented and the construction of the Bayesian hypothesis testing is explained.

2.1 Probability & Bayesian Theory

To better understand the Bayesian methods used in this thesis, it is valuable to go through some basic concepts of probability theory. Additionally, the fundamental theory of Bayesian statistics is introduced.

2.1.1 Probability Theory & Distributions

The mathematical definition of probability, as given in [32] is that probabilities are numerical non-negative quantities, which measure the likelihood of an event occurring in a random experiment. If the events are mutually exclusive, probabilities sum to 1 over all possible mutually exclusive events. Therefore, in other words, probabilities are numbers between 0 and 1, including both extremes, i.e., the [0,1] interval. From the Bayesian perspective, a probability measures the degree of belief or the level of uncertainty of a statement or event [32], [33].

A probability distribution is a function that describes how likely different events are [33]. A variable that can be described by a probability distribution is called random variable. Variable might be categorical or numerical. A categorical variable places something into a category. The numerical variables can be distinguished into two types, namely discrete and continuous. Discrete variables can take only certain values and continuous variables can take any value from some interval.

The probability distribution function of a discrete variable provides the probability that the variable is equal to some values. On the other hand, the probability function of a continuous variable provides probability densities, which can be interpreted as providing the relative likelihood that the random variable would equal a value. For that matter, to get a proper probability, integration between a given interval is required, or in other words the area below the curve, for that interval should be computed. While probabilities cannot be greater than one, probability densities can be, whereas the total area under the probability density curve is restricted to be 1.

Following the notation in [32], $p(\cdot,\cdot)$ denotes a joint probability density with arguments determined by the context and $p(\cdot|\cdot)$ a conditional density. The joint distribution $p(X, Y)$, where X, Y random variables, is the probability distribution that describes the probability that both variables X and Y simultaneously belong to a specific range or discrete set of values. It is important to distinguish the joint distribution from the probability distribution of each variable. The individual probability distribution for each random variable is called marginal distribution and is denoted as $p(\cdot)$. Given the joint distribution $p(X, Y)$, the marginal distribution of X is found when the values of Y are not taken into account. Finally, the conditional distribution $p(X|Y)$ describes the probability of X given that Y has a particular value [34]. To avoid confusion, when $Pr(\cdot)$ is used it refers to the probability of an event. To describe the distribution that a parameter follows, the symbol $\sim$ is used, e.g., if parameter θ follows a normal distribution with mean μ and variance σ^2, then this is written as $\theta \sim N(\mu, \sigma^2)$.

A probability distribution is characterized by its *moments*, which describe the location, scale and shape of the distribution. The two most important moments are the first and second. The first moment is a measure of the central location, which is called expectation (or mean in statistics) and the second is a measure of dispersion (variance in statistics). The third moment is the skewness, which measures the asymmetry about the mean. It can be positive, negative, zero or undefined. The sign of the skewness indicates which tail is bigger; for negative skewness, the tail is usually on the left side of the distribution, while for positive skewness it is on the right. When the skewness is zero, it means that the tails on each side balance each other, and therefore the distribution is symmetric. The magnitude of the skewness indicates how much difference between the tails exists. The fourth moment, also a shape measure, is called kurtosis. It describes the heaviness of the distribution tails, meaning it identifies if the tails contain extreme values.

There are a lot of common families of distribution functions that are used in statistics. One of the most common distributions is the normal or Gaussian. Depending on the variables and the application, different distributions can be used to describe the events.

2.1.2 Statistical Inference

Statistical inference draws conclusions on unobserved quantities based on numerical data observed from a population sample. Bayesian inference, specifically, is the process of reallocating credibility across possibilities. Possibilities are the parameter values in mathematical models, while the credibility assigned to them refers to weighing how probable a possibility is based on the available knowledge and data. Initially, the distribution of credibility expresses the prior knowledge. The reallocation of the credibility is then done based on newly observed data. The possibilities that are consistent with the data gain more credibility, while the rest of them lose credibility. Bayesian inference computes the posterior credibility by updating the prior allocation (distribution of credibility), based on the likelihood function (Fig. 2.1) that describes the observed data, using the Bayes' theorem, described in the next section.

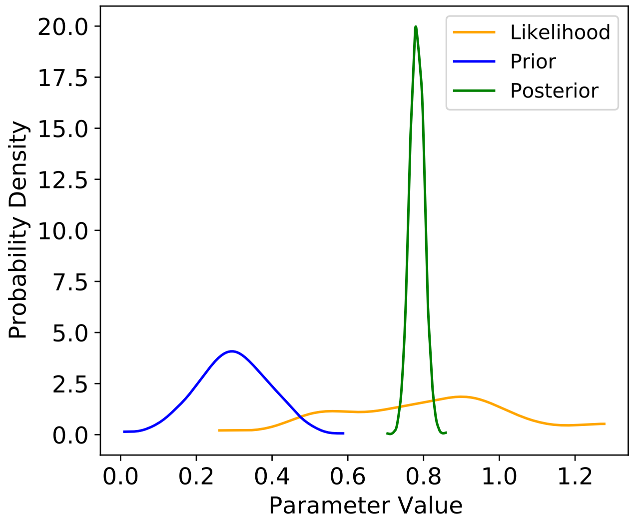


Figure 2.1: Bayesian inference: Prior, likelihood and posterior distribution example.

2.1.3 Bayes' theorem

Let θ denote the set of unobserved quantities or population parameters, which are random variables, and $y = (y_1, y_2, ..., y_n)$ the observed data from a population sample. The Bayes' theorem then is given in the Equation 2.1. The left hand term $p(\theta|y)$ is the conditional probability density of θ given the observed data y, which is called the posterior distribution of θ. On the right hand side of the equation, the term $p(y|\theta)$ is the conditional probability density, or likelihood, of the data y given that θ is known. The term $p(\theta)$ is the prior probability density of θ and the term $p(y)$ on the denominator is the marginal probability density of data y [33].

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{p(y)} \tag{2.1}$$

The prior probability distribution $p(\theta)$ reflects the knowledge (credibility) about the unknown parameter θ prior of observing the data y, such as experts' opinions or information from previous experiments. If no information regarding the parameter is available, then the use of 'flat', or non-informative, priors that do not provide too much information is selected, since they have the least amount of impact on the analysis. A number of variations of the beta distribution are commonly used as non-informative priors. These variations share the properties that the parameters α, β are equal to each other and less or equal than 1. Such priors are the uniform distribution ($Beta(\alpha = 1, \beta = 1)$), the Jeffreys prior ($Beta(\alpha = 0.5, \beta = 0.5)$) and the Haldane prior ($Beta(\alpha = 0, \beta = 0)$), which are illustrated in Figure 2.2 [32], [33].

Usually, in many problems, there is some knowledge about the values the unknown parameters can take, for example that it is only positive or restricted to a known range. In such cases, instead of non-informative priors, weakly-informative priors are selected. The last category of prior distributions are the informative priors, which are considered strong priors, and convey a lot of information. They are used when good-quality and reliable information is available about the unknown parameter.

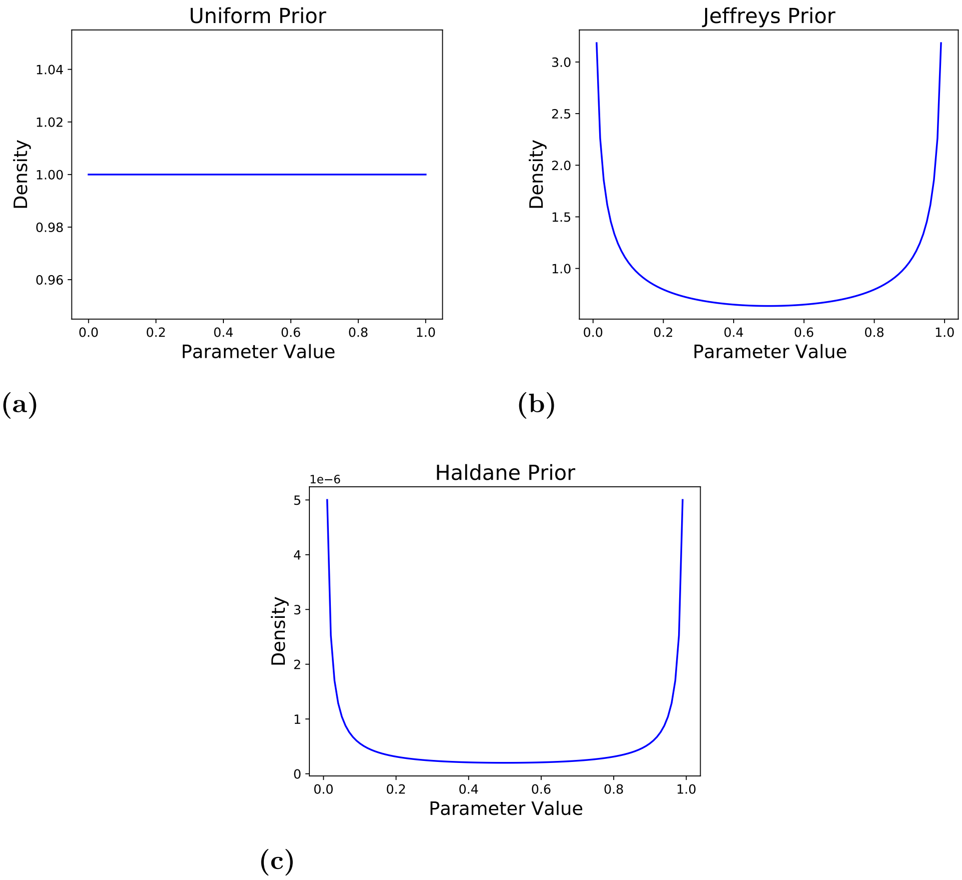


Figure 2.2: Non-informative Priors. 2.2a Uniform prior, 2.2b Jeffreys prior, 2.2c Haldane prior.

The likelihood $p(y|\theta)$ is an expression of the plausibility of the data given the parameters. It describes the probability distribution of the data y being observed given the parameters θ. Among other names for the term $p(y|\theta)$, are the sampling model or statistical model, however to avoid confusion, in the thesis, the combination of the prior ($p(\theta)$) and likelihood functions ($p(y|\theta)$) will be referred to as statistical model.

The posterior distribution $p(\theta|y)$ is the result of the Bayesian analysis, meaning it is the result of the statistical model and data. The posterior is a distribution of plausible values of the unknown parameters, conditioned on the observed values, and not a single value. The posterior distribution represents the updated knowledge about the parameters, after observing the data, and therefore it can be used as a new prior when observing new data. Generally, the posterior is the balance between the prior and the likelihood function, however, when the observed data are stronger than the prior knowledge, the posterior reflects the likelihood more than the prior. Therefore, when the size of the data is large enough, the use of different priors will result in the same posterior.

The last term of the Bayes' Theorem is the marginal likelihood $p(y)$, which is also called evidence. The marginal likelihood refers to the average of the likelihood across all the values of θ, weighted by the prior [33], [35]. The marginal likelihood does not depend on θ and, with fixed y, it is considered constant, and functions as a normalizing constant. For discrete variables, the marginal likelihood is a sum (Eq. 2.2), while for continuous variables it is a integral (Eq. 2.3). The Bayes' Theorem (Eq. 2.1) can be used for both discrete or continuous variables, where probability masses or densities are used, respectively.

$$p(y) = \sum_{\theta} p(\theta)p(y|\theta) \tag{2.2}$$

$$p(y) = \int p(\theta)p(y|\theta)d\theta \tag{2.3}$$

Figure 2.3 illustrates the statistical model used in Bayesian statistics. This model, consisting of the likelihood and the priors on the unknown parameters, is used in the Bayesian inference to procure the posterior distribution of the unknown quantities.

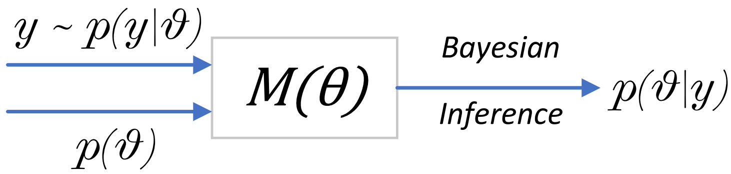
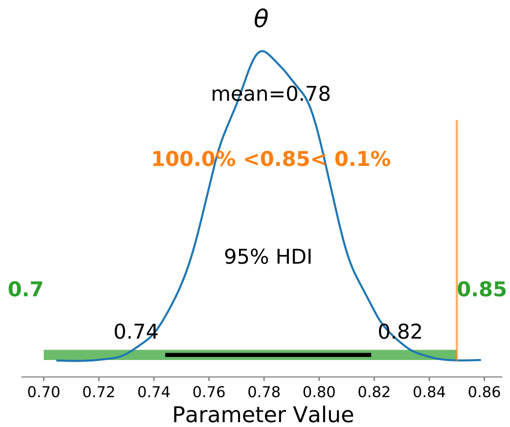


Figure 2.3: Statistical model $M(\theta)$. The data y, which follow the likelihood distribution $p(y \mid \theta)$ and the priors $p(\theta)$ on the unknown parameters θ comprise the statistical model (here illustrated as inputs). Bayesian inference produces the posterior distribution $p(\theta \mid y)$ of the unknown parameters θ, based on the observed data y.

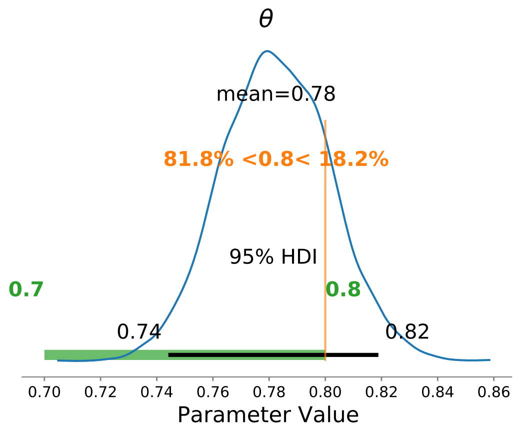
2.1.4 Posterior Distribution Summary

The result of the Bayesian analysis is a posterior distribution. Since it is not a single value, it is useful to summarize it to be able to collect the information provided by the model and data. One of the usual components of this summary is to report the most probable value, which is given by the mode of the posterior, meaning the peak of the distribution. This is accompanied by a measure of dispersion or spread of the posterior, which is proportional to the uncertainty about the value of the parameters. A common practice is to report the standard deviation. However, for non-normal distributions, such as skewed distributions, this can be deceitful, and thus higher moments should also be reported.

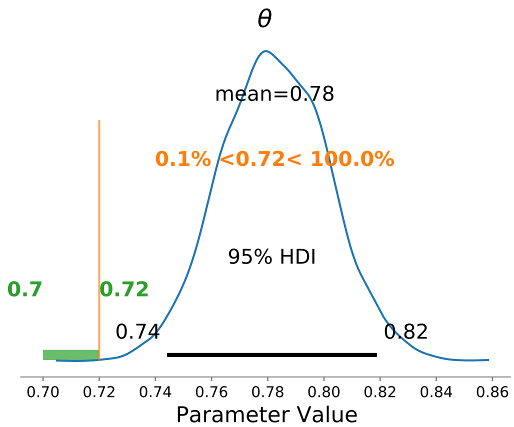
An alternative measure of dispersion is the Highest-Posterior Density (HPD) or Highest Density Interval (HDI). An HDI is the shortest continuous interval that covers a given portion of the posterior distribution. The HDI that are commonly used are the 95% and 50% intervals. The values inside the 95% HDI have higher credibility than the values outside the HDI, and they have total probability of 95%. It should be noted, that the Bayesian credible intervals are not the same as the frequentist confidence intervals (CI). In the frequentist framework, the parameters are fixed and the confidence intervals contains or not the true value of the parameter. The CI does not refer to a distribution over the parameters. On the other hand, in the Bayesian framework, the 95% credible interval indicates the 95% most probable values of the parameter, given the data [33], [35], [36].



(a) HDI entirely in ROPE



(b) HDI partially in ROPE



(c) HDI outside ROPE

Figure 2.4: ROPE and 95% HDI location in the Posterior distribution plot of an unknown parameter θ. The horizontal black and green lines represent the HDI and the ROPE, respectively. The orange vertical line is the upper limit of the ROPE interval. 2.4a Statement is accepted, 2.4b Inconclusive - No decision can be made regarding the statement, 2.4c Statement is rejected.

To make decisions based on the posterior distribution, the above summary is not enough. To make matters clearer, the use of Region Of Practical Equivalence (ROPE) is needed to transform the continuous estimation into a dichotomous one, e.g., yes-no. The ROPE contains the range of values for which the statement or hypothesis is accepted. Once the ROPE is defined, the decision is based on its comparison with the HDI. If the ROPE contains the entire HDI then the statement is accepted, while if the ROPE does not overlap, it is rejected. In case the ROPE overlaps partially with the HDI, then no conclusion can be drawn. Figure 2.4 illustrates the HDI and ROPE relative position of the posterior distribution of parameter θ. The black horizontal line represents the ROPE interval, the upper limit of which is depicted with the vertical orange line, and the green horizontal line depicts the 95% HDI.

2.1.5 Posterior Predictive Check

The posterior distribution can also be used to evaluate how good the statistical model is. By default, the model is wrong, but it is useful to understand its limitations, and either acknowledge them or proceed to model modifications and improvements. For that matter, posterior predictive checks can be performed.

The posterior predictive checks make inferences about an unknown but observable quantity $\tilde{y}$. Before the data y are observed, the distribution of the unknown observable $\tilde{y}$ is the marginal likelihood given in Equation 2.3, which is here called the prior predictive distribution. This name is more informative, and it is given because the distribution is not conditional on any observation (*prior*) and it refers to a quantity that is observable (*predictive*). After the data y are observed, the unknown observable $\tilde{y}$ can be predicted, and its distribution is called the posterior predictive distribution (Eq. 2.4). This name is given to indicate that it is conditional on an observed y (*posterior*) and it is a prediction of the observable $\tilde{y}$ (*predictive*) [32].

$$p(\tilde{y}|y) = \int p(\tilde{y}, \theta|y)d\theta = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \tag{2.4}$$

The posterior predictive checks compare the posterior predictive distribution to the sample data. The comparison can be qualitative, by comparing the distributions visually using graphs. A quantitative comparison involves a summary statistic (usually regarding the location and/or spread) and a quantity to measure the fit of the posterior predictive check. Such measure is a Bayesian p-value, which is calculated as the proportion of times that the summary statistic of the posterior predictive is equal or greater than the one computed from the data. Equation 2.5 calculates the Bayesian p-value, where T_{sim} is the simulated statistic and T_{obs} is the statistic of the data y. These p-values are just a way of assessing the fit of the posterior predictive distribution to the dataset, and thus no predefined threshold is used to indicate statistical significance.When the data and the posterior predictive agree, the predictive distribution of the statistic should be centered around the statistic computed from the data. This is translated to the p-value being around 0.5. If the Bayesian p-value is close to zero or one, the predictions are far away from the statistic of the data. Therefore, the posterior predictive distribution is biased, and improvements to the model are necessary. Depending on the context, the acceptable discrepancies between the predictive checks and the observed data can differ [33].

Figure 2.5 illustrates the meaning of Bayesian p-values. The black dashed line is the mean value (statistic T) computed from the observed data, and the blue line is the posterior predictive distribution of the mean. The computed Bayesian p-value is equal to 0.25, which is lower than the desired value (0.5). It is apparent from Figure 2.5 that the statistical model predicts lower values of the mean. This is represented numerically by the p-value.

$$\text{Bayesian p-value} \triangleq Pr\left(T_{sim} \geq T_{obs} \mid y\right) \tag{2.5}$$

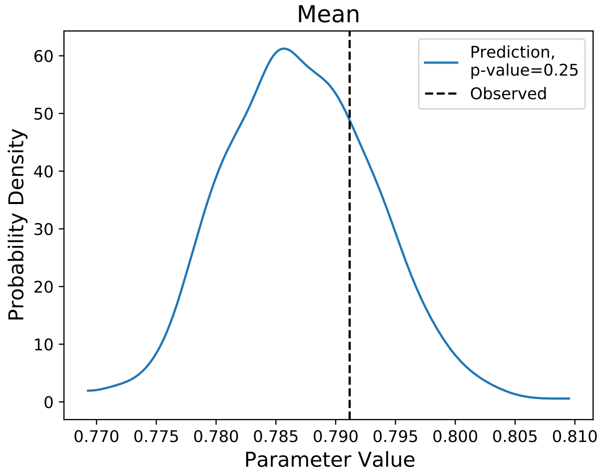


Figure 2.5: Posterior predictive distribution of mean (summary statistic T) for Bayesian p-value of 0.25.

2.1.6 Inference engines

Conceptually, the Bayesian analysis is quite simple. It includes the knowns, which are the data (y), and are constant, and the unknowns, which are the statistical parameters (θ), and are random variables, described by probability distributions. Using Bayes' theorem, the prior distributions of the parameters ($p(\theta)$) are transformed into the posteriors ($p(\theta|y)$), after observing the data. If this analysis is carried out analytically, it can yield very complex statistical models, and depending on the prior and likelihood distributions ($p(y|\theta)$), it might be impossible to acquire an analytical solution. The reason why the Bayesian problem can be difficult to solve analytically is the calculation of the marginal likelihood $p(y)$ (denominator in the Bayes' theorem), which is usually a very computational demanding integral to compute (Eq. 2.3).

However, using computational and numerical methods any inference problem can be solved. These methods are called inference engines. The automation of the inference process can be done using probabilistic programming languages (PPL). In this thesis, the PyMC3 package [37] for python is used to perform the Bayesian inference. Generally, in the PPL, the probabilistic model is specified and then the inference is done automatically. The posterior distribution is usually calculated numerically using one of the algorithms from the Markov Chain Monte Carlo (MCMC) family, or variational algorithms. Some of the most wide-used algorithms are the Metropolis-Hastings and Gibbs. In this work, the samplers used are the Metropolis, for the discrete variables, and the No-U-Turn Sampler (NUTS), for the continuous variables.

MCMC methods draw values of θ from approximate distributions and then correct the draws to better approximate and converge to the desired, stationary, posterior distribution $p(\theta|y)$. The sampling of the draws is done sequentially, where each sampled draw depends only on the previous value drawn (Markov Chain). MCMC methods are usually used when it is not possible, or efficient, to sample directly from the posterior distribution. They are iterative methods, where it is expected that the draw at each step of the process becomes closer to the target distribution [32], [33].

The Metropolis-Hastings algorithm is an adapted random walk, with an acceptance (or rejection) criterion to converge to the target distribution. The algorithm involves the following steps.

1. Choose a starting (random) value for the parameter θ^i ($i = 0$), for which $p(\theta^0|y) > 0$.
2. For each sequence $i(= 1, 2, \dots)$, draw a new parameter value θ^i, that depends on the previous draw (θ^{i-1}), from a transition distribution $p_i(\theta^i|\theta^{i-1})$. The transition distribution can be an easy-to-sample distribution, such as a uniform or a gaussian distribution. Specifically for the Metropolis algorithm (unlike the generic Metropolis-Hastings), the transition distribution must be symmetric, which means it satisfies the condition $p_i(\theta_\alpha|\theta_b) = p_i(\theta_b|\theta_\alpha)$, for all θ_α, θ_b and i.
3. Calculate the probability of accepting the new parameter value (θ^i) using the Metropolis-Hastings criterion:

$$Pr_a(\theta^i|\theta^{i-1}) = min\left(1, \frac{p(\theta^i)p_i(\theta^i|\theta^{i-1})}{p(\theta^{i-1})p_i(\theta^i|\theta^{i-1})}\right) \tag{2.6}$$

4. The new parameter value is accepted if the probability computed in step 3 (Pr_a) is greater than the value taken from a uniform distribution on the [0,1] interval, otherwise the previous drawn value is kept.
5. Repeat from step 2 until enough samples are drawn
6. The drawn samples are the approximated target distribution (posterior)

The NUTS algorithm is a variation of the Hamiltonian Monte Carlo (HMC) algorithm. The HMC is similar to the Metropolis-Hastings algorithm, with the difference that HMC avoids the use of random walk to propose new parameter values. HMC algorithm utilizes first-order gradient information to proceed to a new step, enabling quicker convergence to the target distributions than other more simple methods (e.g., Metropolis-Hastings or Gibbs). However, HMC is sensitive to two user-defined parameters; the step size and the number of steps . On the other hand, NUTS automatically tunes these two parameters. It uses a recursive algorithm to generate a set of potential points over a wide range of the target distribution and stops (automatically) when it starts to go back and repeat its steps. NUTS is usually as efficient (or more) as a fine-tuned HMC algorithm, without the need for user interference [38].

2.2 Bayesian Hypothesis testing

In [39] hypothesis testing is defined as "a method for testing a claim or hypothesis about a parameter in a population, using data measured in a sample". Hypothesis testing involves two hypotheses, namely the null and alternative hypothesis. The null hypothesis is the one that is tested to determine if it is true, while the alternative hypothesis states what is wrong regarding the null hypothesis. For example, if the null hypothesis is that $\theta = 0.5$, then the alternative hypothesis can be that $\theta \neq 0.5$. In classical hypothesis testing, the decision about rejecting the null hypothesis or not is based on a test statistic, which is a function of the sample data [40].

In the Bayesian context, hypothesis testing again involves two (or more) competing hypotheses and some sample data. For simplicity, the hypothesis testing framework is described only for the binary univariate case, which includes the null (H_0) and one alternative (H_1) hypothesis. The two hypotheses are mutually exclusive. The formulation of hypothesis testing can be divided in two main types, in terms of the hypothesis statement formulation, namely the point-based and the interval-based hypothesis testing, which are explained in Chapters 2.2.1 and 2.2.2.

2.2.1 Bayesian Point-based Hypothesis Testing

Let θ_0 and θ be the predicted and the true value of an unknown parameter. The θ_0 value is predicted by the statistical model M. The point-based hypothesis testing can be formulated as follows:

$$\begin{aligned} H_0 &: \ \theta = \theta_0 \\ H_1 &: \ \theta \neq \theta_0 \end{aligned}$$

In a Bayesian point-based hypothesis testing, the null hypothesis H_0 is expressed as a prior distribution that assigns all credibility only in an infinitely narrow area around the null value, while all the other parameter values have zero credibility. The prior distribution of the alternative hypothesis spreads the credibility to other than the null values of the parameter [36].

The point-based hypothesis testing is quite strict and it is preferred when the predicted and true value deviate slightly. Otherwise, due to the formulation, if the values are expected to differ, the null hypothesis is always rejected and no significant conclusion can be drawn. Additionally, as the sample size increases, the possibility of rejecting the null hypothesis also rises. A more relaxed formulation of the hypothesis testing is given by the interval-based hypothesis testing, described in the next chapter.

2.2.2 Bayesian Interval-based Hypothesis Testing

The generic formulation of the interval-based hypothesis testing, considering only one alternative hypothesis, is:

$$\begin{aligned} H_0 &: \ \epsilon_l \leq \theta \leq \epsilon_u \\ H_1 &: \ \theta \leq \epsilon_l \ or \ \theta \geq \epsilon_u \,, \end{aligned}$$

where ϵ_l and ϵ_u are predefined lower and upper limits of the predicted value of the unknown parameter θ, respectively, and depend on the application.

The prior distribution of the parameter for each hypothesis should reflect the above formulation. The credibility, in the null hypothesis, is given to the values of the interval $[\epsilon_l, \epsilon_u]$, with zero credibility assigned on all the values outside of it $(-\infty, \epsilon_l) \cup (\epsilon_u, +\infty)$. Equation 2.2.2 illustrates a general formulation of an interval-based hypothesis testing. In many cases, however, the alternative hypothesis can be more than one. Additionally, a hypothesis (null or alternative) can have either a lower or an upper limit or both, or even include more than one intervals. The form of each hypothesis depends on the evaluated case.

Unlike point-based hypothesis testing, which assesses the equality of two values, interval-based hypothesis testing tests if the predicted value is within a specified interval. Moreover, the interval-based formulation provides more consistent results with increasing sample size [19], [41].

2.2.3 Bayes Factor

The Bayes factor quantifies to what extent one hypothesis is supported compared to another and it can range from 0 to infinity. Let $p(H_0)$ and $p(H_1)$ be the priors for the null and alternative hypothesis respectively, such that $p(H_0) + p(H_1) = 1$. After obtaining the sample data y, these distributions are updated to the posteriors $(p(H_0|y), p(H_1|y))$ using Bayes' theorem (Eq. 2.1). The ratio of the posterior odds is shown in Equation 2.7. The likelihood ratio, which is the term inside the brackets, is defined as the Bayes factor B_{01} (Eq. 2.8).

$$\frac{p(H_0|y)}{p(H_1|y)} = \left[\frac{p(y|H_0)}{p(y|H_1)}\right] \frac{p(H_0)}{p(H_1)} \tag{2.7}$$

$$B_{01} = \frac{p(y|H_0)}{p(y|H_1)} = \frac{\int p(\theta|H_0)p(y|\theta, H_0)d\theta}{\int p(\theta|H_1)p(y|\theta, H_1)d\theta} \tag{2.8}$$

One way of interpreting the Bayes factor is continuously, where for example a Bayes factor equal to 10, means that the sample data support the null hypothesis 10 times more than the alternative hypothesis. The Bayes factor can also be interpreted discretely. The general idea behind this, is that if the Bayes factor is higher than 1.0, then the null hypothesis is accepted relative to the alternative hypothesis, while if it is lower than 1.0, the evidence supports the alternative hypothesis more than the null [41]. This is a very crude methodology, because if the Bayes factor is slightly larger than 1, the supporting evidence is not enough to conclude that the null hypothesis is more probable. For that matter, a categorization of the values of the Bayes factor is needed to illustrate the strength of evidence that support the null hypothesis (Table 2.1).

It is apparent, that as the value of the Bayes factor increases, so does the confidence in the model/null hypothesis based on the evidence. The confidence κ is quantified by the posterior probability of the null hypothesis $p(H_0|y)$, which can be found by rearranging Equation 2.7, and it is shown in Equation 2.9 [42].

$$\kappa = 100\% \cdot p(H_0|y) = 100\% \cdot \frac{B_{01}p(H_0)}{p(H_1) + B_{01}p(H_0)} \tag{2.9}$$

If no prior knowledge is available about the hypotheses, equal priors are assumed, and thus the confidence is simplified to:

$$\kappa = 100\% \cdot \frac{B_{01}}{1 + B_{01}} \tag{2.10}$$

From Equation 2.10, it is obvious that for a Bayes factor near 0 ($B_{01} \to 0$) the confidence in the null hypothesis being accepted is 0%, while when the Bayes factor approaches infinity ($B_{01} \to \infty$) the model confidence is 100% . The confidence levels for each category of Bayes factor's values regarding the level of support in the null hypothesis are also shown in Table 2.1.

Table 2.1: Categorization of Bayes factor's values depending on the strength of evidence [33].

Bayes Factor (B_{01})	**Strength of Evidence**	**Confidence κ**
1-3	Anecdotal	50-75%
3-10	Moderate	75-90.91%
10-30	Strong	90.91-96.78%
30-100	Very strong	96.78-99%
>100	Extreme	>99%

2.2.4 Model Comparison

One framework used for Bayesian hypothesis testing is based on (statistical) model comparison. According to this, each hypothesis is modelled as a distinct statistical model, namely the null and the alternative model. Therefore, the decision between the two hypotheses is translated into selecting one of the models [43].

The model comparison is based on the construction of hierarchical models. Hierarchical or multi-level models include "multiple parameters such that the credible values of some parameters meaningfully depend on the values of other parameters" [35]. These models are used when the parameters involved are somehow related and thus the joint distribution should reflect their connection [32]. The parameters are organized at different levels, defined by the chain of dependencies, but they all belong to a joint parameter space.

Let M_m for $m \in \{0, 1\}$ represent the two models, and m be a discrete random variable, which takes only the values 0 or 1, depending on the model. If $m = 0$ then the null model M_0 is true, whereas if $m = 1$ then the alternative model M_1 is true. In Chapter 2.1.3, the unknown parameters of a model were denoted as θ, the prior distribution as $p(\theta)$ and the likelihood as $p(y|\theta)$. Expanding these notations to refer to multiple models, the prior and likelihood distributions become $p_m(\theta_m|m)$ and $p_m(y|\theta_m, m)$, respectively, where θ_m are the parameters of model M_m. The subscript given to the parameters indicate that each model can involve different parameters.

Each model also has a prior probability $p(m)$ or $p(M_m)$. Considering the joint parameter space which includes the unknown parameters θ_m of each model and the model index m, Bayes' theorem (Eq. 2.1) is then transformed for multiple models (Eq. 2.11). The hierarchical modelling enables the transformation of the joint parameter space (θ, m), in Eq. 2.11, to dependencies among parameters through factoring (convert $p(y|\theta_m, m)p(\theta_m, m)$ to $\prod_m p_m(y|\theta_m, m)p_m(\theta_m|m)p(m)$) [35].

$$\begin{aligned} p(\theta_m, m|y) &= \frac{p(y|\theta_m, m)p(\theta_m, m)}{\sum_m \int d\theta_m p(y|\theta_m, m)p(\theta_m, m)} \\ &= \frac{\prod_m p_m(y|\theta_m, m)p_m(\theta_m|m)p(m)}{\sum_m \int d\theta_m \prod_m p_m(y|\theta_m, m)p_m(\theta_m|m)p(m)} \end{aligned} \tag{2.11}$$

A simplified representation of the Bayes' theorem for model comparison, when marginalized across the parameters of the models, is given in Equation 2.12, [35], [44]. This form of the Bayes' theorem is useful to compute the relative credibility of the models [35].

$$p(M_m|y) = \frac{p(y|M_m)p(M_m)}{\sum_m p(y|M_m)p(M_m)} \tag{2.12}$$

In this hierarchical formulation, the inference reallocates credibility not only across the values of the parameters of each model, but also across the values of the 'top-level' parameter, the model index m. Figure 2.6 illustrates the hierarchical models used in model comparison, in the form of diagrams. Figure 2.6a depicts the joint parameter space across all statistical models, reflecting Equation 2.11, before the factorization. The latter is shown in Figure 2.6b, where the hierarchical model is factored into distinct statistical models, which are depicted in the blue boxes. Each statistical model has its own parameters and distributions, however they are all under the top-level parameter m. Figure 2.6c is a special case, where the distinct statistical models have the same likelihood, but different prior distributions. In these figures, the prior distribution of the model index m is presented as a bar graph over its possible values ($m = 0, 1, \ldots, n$).

Although the inference is done for all the statistical parameters involved, the hypothesis testing relies only on the inference of the model index and not on the estimation of the statistical model parameters. The selection of the model, and thus the hypothesis that is more probable to be true based on the observed data, is done using the decision rule described in Chapter 2.2.3. The Bayes factor measures the degree to which the model index changes from prior to posterior. The Bayes factor in model comparison is defined as the ratio of posterior odds to the prior odds in Equation 2.13. To avoid confusion, it can also be defined as in Eq. 2.8, as the ratio of the marginal likelihoods ($\frac{p(y|M_0)}{p(y|M_1)}$). However, due to the difficulty in calculating them and their sensitivity to the priors, the Bayes factor is usually preferred to be computed using Equation 2.13.

$$B_{01} = \frac{\frac{p(M_0|y)}{p(M_1|y)}}{\frac{p(M_0)}{p(M_1)}} \tag{2.13}$$

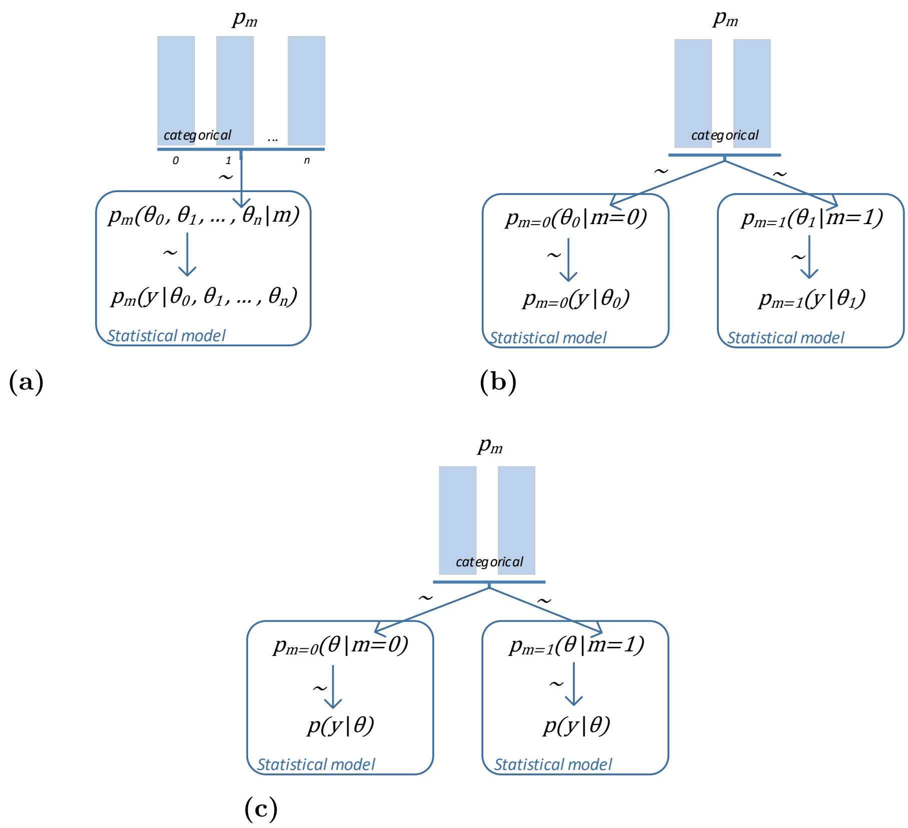


Figure 2.6: Hierarchical models with hyperparameter a categorical parameter m with prior $p(m)$. The blue boxes include the distinct statistical models (submodels) that are being compared. 2.6a Generic concept of a hierarchical model, with parameters θ_m for all submodels, in the joint parameter space. 2.6b Hierarchical model with two submodels involving different likelihoods. The likelihood and the priors are reduced into functions of only the parameters θ_m for each m. 2.6c Special case of 2.6b, where the two submodels have the same likelihood for all m, but different priors on the parameters for each m.

3

Vehicle Modelling in Automation

This chapter introduces the vehicle model that is used in the thesis. Depending on the application, different aspects of the vehicle behavior are required to be modelled. Therefore, the vehicle models can be very different for each use case they are designed for. Their complexity can vary, ranging from a simple linear model, usually used for designing control algorithms or online applications, to high complexity models, consisting of multiple subsystems, that can capture the dynamic vehicle behavior to a great extent. The latter are often used for offline calculations, due to their high computational demands.

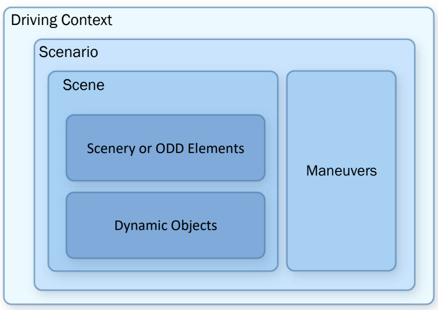
The model to be validated in this work is a simple model, called single track model of a tractor-semitrailer combination. In the following sections, the equations of motion and simplifications of the model are presented. The validation of the abstract model was implemented by comparing it to an implementation model, which is a higher complexity model, provided by VAS.

3.1 Driving Framework

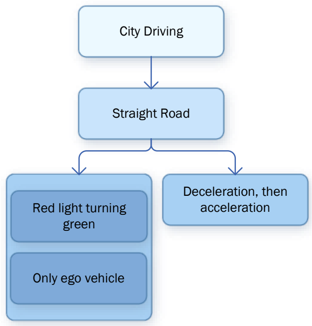
One of the most important parts during the design phase is the definition of the driving framework. The selection of the framework affects both the modelling and the validation stage. Therefore, it must be clearly defined at the early stages of the design process. Different definitions can be found in literature. Many formulations involve theater references, including terms such as scenes, and actors [45], [46]. The term scenario is also usually used to describe and substantiate the environment and conditions of the driving task. Menzel et al. [47] defined three levels of scenarios, depending on the extent of details involved. However, these definitions can often be vague or inconsistent.

In this work, the driving framework is composed of different levels of detail regarding the driving situations. The first component of the framework is the driving context. This refers to the type of the road environment, such as urban or high-speed roads, in conjunction with the vehicular characteristics, which can include speed, acceleration, or steering angles. This makes the vehicle under study an integral part of the driving framework. Based on this definition, it is evident that different driving contexts can include some common vehicular characteristics and therefore overlap [16].

The driving context contains a set of different driving scenarios, which correspond to different road types and layouts. Examples of scenarios are that the road is straight, curved, or an intersection. Each driving scenario is specified by the scene and the driving maneuver(s). The scene contains the scenery and all the dynamic objects. The scenery includes, but is not limited to, environmental and weather conditions, such as the ambient temperature, additional information and characteristics of the road, for instance, road friction, and geographical restrictions. Stationary elements are also a part of the scenery. These can be traffic lights, signs, or even buildings. The dynamic objects are elements that are moving or can move [45]. They can be vehicles, other than the ego one, or pedestrians. Here, the ego vehicle refers to the vehicle being studied, and optionally, the driver and/or the control system or automation. The driving maneuvers concern the motion of the ego vehicle and are defined as a set of one or more actions of the ego vehicle. An action can be, for example, deceleration, navigation, or steady-state cornering. The structure of the driving framework used in this thesis is illustrated in Figure 3.1a.



(a) Driving Framework Structure



(b) Test Case Example

Figure 3.1: Driving Framework Structure Specification.

A specific scenario under a specified driving context is referred to as a test case. Each test case can include a different set of maneuvers and scenes, which can be performed under the selected driving context. An example of a test case is illustrated in Figure 3.1b. The driving context is city driving and the scenario is that of a straight road. The scenery contains a traffic light which turns from red to green, and since no specification is given regarding the environmental and road conditions, they are defined using typical values and characteristics. No moving objects are included in this test case and only the ego vehicle is studied. The maneuvers of the ego vehicle include the actions of deceleration, followed by acceleration. A use case, on the other hand, refers to the desired behavior and typical usage of the test cases. An example of a use case can be the automated parking.

3.2 Single Track Model of a Tractor-semitrailer Combination

Simple vehicle models are very useful, especially in basic analysis, because of the low computational cost and the low number of parameters. One of the wide-used simple models is the single track vehicle model.

3.2.1 Modelling Assumptions

The simplicity of a vehicle model lies in the assumptions that are made during the design phase. The driving situation under which the model is studied, sets the limits of the vehicle dynamic behavior, such as the maximum acceleration, and therefore dictates the assumptions that can be made. Since the assumptions can affect the behavior of the model, serious consideration is required such that the model captures and reflects the real dynamics to a certain degree.

In the single track model of a tractor-semitrailer combination, the two tires of an axle are combined into one single tire, located at the center of the axle in the vertical center plane (disregarding the height of the CoG from the ground). Moreover, the tire models are considered linear, where the tire forces are linear functions of the slip angles. The longitudinal forces are neglected, meaning that no braking or accelerating forces are considered, and only the lateral forces on the wheels are taken into account. No aerodynamic forces are included, and the road is assumed flat. The coupling between the units is considered rigid and only kinematic relations are used to model it. Finally, only planar motion is concerned, neglecting the vertical dynamics and load shifting. Figure 3.2 illustrates the single track model of a tractor-semitrailer combination. The tractor has two axles and the semitrailer three. The two units are coupled to each other with angle $\Delta\psi$.

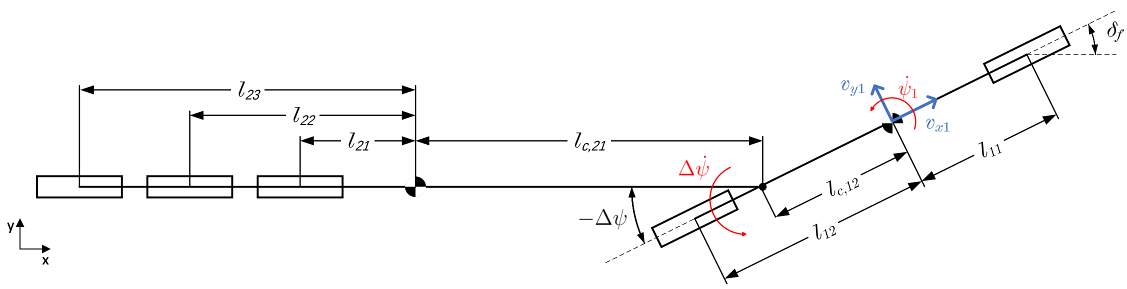


Figure 3.2: Single track model of a tractor-semitrailer combination.

3.2.2 Equations of motion

The equations of motion for the single track model of a tractor-semitrailer combination in planar motion are derived from Newton's second law. To describe the quantities involved in the vehicle model, three different reference frames are used, namely the global, which is fixed, the unit, and the wheel reference frame. The relation between them is described using rotation matrices. The transformation from the wheel to the unit frame is done through the matrix R_δ (Eq. 3.1), from the unit to the global frame through the matrix R_ψ (Eq. 3.2) and from unit i to unit $i + 1$ with the matrix $R_{\Delta\psi}$ (Eq. 3.3) [16].

$$R_\delta = \begin{bmatrix} \cos(\delta) & -\sin(\delta) \\ \sin(\delta) & \cos(\delta) \end{bmatrix} \tag{3.1}$$

$$R_\psi = \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix} \tag{3.2}$$

$$R_{\Delta\psi} = \begin{bmatrix} \cos(\Delta\psi) & -\sin(\Delta\psi) \\ \sin(\Delta\psi) & \cos(\Delta\psi) \end{bmatrix} \tag{3.3}$$

The Newton's second law is applied on each unit i, as shown in Equations 3.4, 3.5. The wheel force on axle j is a vector, containing the forces in the x and y direction, in the wheel coordinate system ($\boldsymbol{F_{w,ij}} = [F_{wx,ij} \quad F_{wy,ij}]^T$). These are then transformed to the unit's frame ($\boldsymbol{F_{vw,ij}}$), with the use of the rotation matrix $R_{\delta,i}$ (Eq. 3.6).

$$\sum_j \boldsymbol{F_{v,ij}} = m_i a_i \tag{3.4}$$

$$\sum_j M_{z,ij} = I_{zz,i}\ddot{\psi}_i \tag{3.5}$$

$$\boldsymbol{F_{vw,ij}} = R_{\delta,i}\boldsymbol{F_{w,ij}} \Rightarrow \begin{bmatrix} F_{vwx,ij} \\ F_{vwy,ij} \end{bmatrix} = R_{\delta,i} \begin{bmatrix} F_{wx,ij} \\ F_{wy,ij} \end{bmatrix} \tag{3.6}$$

The units in the tractor-semitrailer model are connected through a fifth-wheel coupling. The coupling forces are defined in the global reference frame as $\boldsymbol{F_{ec,i}} = [F_{ecx,i} \quad F_{ecy,i}]^T$ and they are transformed to the unit reference frame through the $R_{\psi,i}$ rotation matrix (Eq. 3.7). It should be noted that, according to Newton's third law, the rear coupling force of unit i is equal to the front coupling force of the unit $i+1$, but in opposite direction.

$$\boldsymbol{F_{vc,i}} = R_{\psi,i}^{-1}\boldsymbol{F_{ec,i}} \Rightarrow \begin{bmatrix} F_{vcx,i} \\ F_{vcy,i} \end{bmatrix} = R_{\psi,i} \begin{bmatrix} F_{ecx,i} \\ F_{ecy,i} \end{bmatrix} \tag{3.7}$$

The moments about the z-axis on the center of gravity for each unit i are defined as in Equation 3.8, where $\boldsymbol{r_j}$ is the moment arm from the center of gravity to the corresponding axle.

$$M_{z,ij} = \boldsymbol{r_j} \times \boldsymbol{F_{v,ij}} \tag{3.8}$$

In planar motion, the vehicle velocity is defined as $\boldsymbol{v} = [v_x \quad v_y \quad 0]^T$ and the angular velocity is assumed to be equal to the yaw rate, such that $\boldsymbol{\omega} = [0 \quad 0 \quad \dot{\psi}]^T$. The vehicle acceleration is then found as shown in Equation 3.9.

$$\boldsymbol{a} = \frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{\omega} \times \boldsymbol{v} \Rightarrow \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} \dot{v_x} - \dot{\psi} v_y \\ \dot{v_y} + \dot{\psi} v_x \\ 0 \end{bmatrix} \tag{3.9}$$

Therefore, the equations of motion can be written for the tractor as in Equation 3.10 and for the semitrailer as in Equation 3.11. These expressions are for a tractor with two axles and a semitrailer with three axles. The first index refers to the unit, where 1 is for the tractor and 2 for the semitrailer, and the second index to the axle.

$$\begin{aligned} F_{vwx,11} + F_{vwx,12} + F_{vcx,12} &= m_1(\dot{v}_{x1} - \dot{\psi}_1 v_{y1}) \\ F_{vwy,11} + F_{vwy,12} + F_{vcy,12} &= m_1(\dot{v}_{y1} + \dot{\psi}_1 v_{x1}) \\ F_{vwy,11} l_{11} - F_{vwy,12} l_{12} - F_{vcy,12} l_{c,12} &= I_{zz,1} \ddot{\psi}_1 \end{aligned} \tag{3.10}$$

$$\begin{aligned} F_{vwx,21} + F_{vwx,22} + F_{vwx,23} + F_{vcx,21} &= m_2(\dot{v}_{x2} - \dot{\psi}_2 v_{y2}) \\ F_{vwy,21} + F_{vwy,22} + F_{vwy,23} + F_{vcy,21} &= m_2(\dot{v}_{y2} + \dot{\psi}_2 v_{x2}) \\ -F_{vwy,21} l_{21} - F_{vwy,22} l_{22} - F_{vwy,23} l_{23} + F_{vcy,21} l_{c,21} &= I_{zz,2} \ddot{\psi}_2 \end{aligned} \tag{3.11}$$

3.2.3 Tire Model, Constitution & Compatibility Relations

Tires play a crucial role in the dynamic behavior of the vehicles [48]. They are the primary contact between vehicle and road, while the other chassis components affect the vehicle performance through the tire forces and torques applied over the contact patch [49]. This, reveals the direct relation of the tire to the vehicle dynamics.

The modelling of the interactions between the tire and the road is a significant task of the modelling process, which justifies the plethora of tire models developed. The models can be distinguished between physical and empirical [50]. The physical models utilize the knowledge of the physical structure of the tire. They can be simplified, using simple mechanical representation, such as the brush model, or complex, describing the tire in great detail [48], [50]. The empirical models represent the tire characteristics through mathematical formulas and interpolation schemes or tables, through fitting experimental data with regression procedures. The Magic Formula by Pacejka is an example of an empirical model [48].

The tire forces and torque are function of the slip and the wheel load. In this thesis, a linear tire model is considered, and thus the lateral force is a linear function of the lateral slip (Eq. 3.12). The cornering stiffness C_y in Equation 3.12 is proportional to the normal force F_z with proportionality coefficient C_{cy}, called cornering coefficient (Eq. 3.13) [51].

$$F_{wy} = C_y s_y \tag{3.12}$$

$$C_y = C_{cy} F_z \tag{3.13}$$

The lateral tire slip for the front (steered) and rear axle of unit i is given in Equation 3.14 and 3.15, respectively.

$$s_{y,if} = \frac{v_{yi} + l_{if}\dot{\psi}}{v_{xi}} - \delta_f \tag{3.14}$$

$$s_{y,ir} = \frac{v_{yi} - l_{ir}\dot{\psi}}{v_{xi}} \tag{3.15}$$

3.2.4 Kinematic Relations

The relationship between the motion of the units is established through kinematic equations, which relate the velocity and the acceleration of each unit based on the yaw angles between them. The relative velocity of a point P in a moving coordinate system with origin O is given by the Equation 3.16, where $\boldsymbol{r_{PO}}$ is the distance between the points P and O [52].

$$\boldsymbol{v_P} = \boldsymbol{v_O} + \boldsymbol{\omega} \times \boldsymbol{r_{PO}} \tag{3.16}$$

$$\boldsymbol{a_P} = \boldsymbol{a_O} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r_{PO}}) \tag{3.17}$$

To construct the kinematic relation of the velocities, first the velocity $\boldsymbol{v_{c,i,r}}$ at the rear coupling point of the (first) unit i, is found, using Equation 3.16. It is then transformed to the reference frame of the unit $i+1$ ($\boldsymbol{v_{c,i+1,f}}$) and finally, using once more Equation 3.16, the velocity of the center of gravity of the unit ($\boldsymbol{v_{i+1}}$) is calculated. Using Equation 3.17, the relation of the acceleration of units i and $i+1$ is obtained. For instance, the acceleration at the rear coupling point of unit i is shown in Equation 3.18.

$$\boldsymbol{a_{c,i,r}} = \begin{bmatrix} \dot{v_x} \\ \dot{v_y} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta\dot{\psi}_i \end{bmatrix} \times \begin{bmatrix} v_x \\ v_y \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \ddot{\psi}_i \end{bmatrix} \times \begin{bmatrix} -l_{c,ir} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta\dot{\psi}_i \end{bmatrix} \times \left(\begin{bmatrix} 0 \\ 0 \\ \dot{\psi}_i \end{bmatrix} \times \begin{bmatrix} -l_{c,ir} \\ 0 \\ 0 \end{bmatrix} \right) \tag{3.18}$$

3.2.5 Final Equations of Single Track Model

The equations for the single track tractor-semitrailer vehicle model are derived using the relations presented in the previous sections. The resulting set of equations, however, are not presented here due to space limitations. The system of equations describing the single track model are nonlinear and can be expressed as ordinary differential equations of the form:

$$\dot{x} = f(x, u) \tag{3.19}$$

3.3 Performance Metrics

The performance metrics are used to describe the overall behavior of the vehicle. They are single values that quantify different characteristics of the vehicle's response. The performance metrics are chosen based on the driving situation, the vehicle configuration and the intended application, so that they reveal and summarize the necessary and important aspects of the vehicle performance.

The metrics can be separated into categories, such as time-series based and quantities of interest (QoI). In the first category, indicators for describing the shape of responses in time and frequency domain are included. In time domain, the metrics for transient response that are commonly used are peak time, peak value, rise time, percent overshoot, and settling time.

The peak time t_p in a step response is the time required for the response $y(t)$ to reach the first peak, called the peak value ($M_p = y(t_p)$), and it indicates the responsiveness of the system. The rise time t_r is the required time for the response to rise from the 10% to 90% of its final value (M_f). The maximum percent overshoot M_{OV} measures the distance between the peak value and the final value and it is calculated as in Equation 3.20. The percent overshoot shows the relative stability of the system.The settling time t_s is the time required for the response to reach the final value and remain within a range around the final value, usually specified as 2% or 5% [53]. Figure 3.3 illustrates a step response and the aforementioned quantities.

$$M_{OV} = \frac{y_p - y_f}{y_f} \times 100\% \tag{3.20}$$

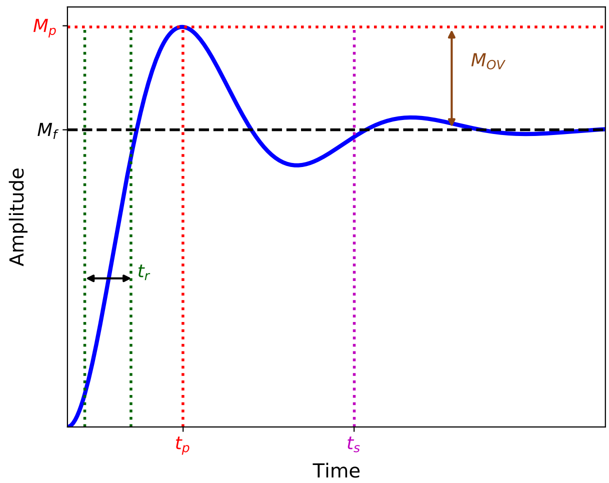


Figure 3.3: Time domain metrics marked in a step response example.

Time domain analysis, however, cannot study efficiently high frequency transient responses [17]. Therefore, frequency analysis should be implemented. In frequency domain, the peak amplitude and the corresponding peak frequency, as well as the eigenfrequencies are usually used. The peak amplitude is the maximum gain, which is usually defined as the ratio of the output relative to the input. The eigenfrequencies, also known as natural frequencies, is the frequencies in which the system oscillates in the absence of damping.

The metrics defined as quantities of interest depend on the application and the behavior that is desired to be observed. The quantities of interest can be physical quantities, common to all vehicles, such as the lateral acceleration or the yaw rate. They can also be, quantities specific for articulated heavy vehicles. For the tractor-semitrailer, the common quantities of interest of the lateral performance are the rearward amplification, the swept path width, the high-speed transient and steady state off-tracking and the yaw damping coefficient.

The metrics used for describing and assessing the vehicle performance can belong to different categories simultaneously. It is very common to use a combination of the above, such as the maximum lateral acceleration, or the rise time of the yaw rate.

In this work, the metrics that are used for the analysis of the vehicle behavior, are the peak value and rise time of the lateral acceleration of the tractor and the semi-trailer at the center of gravity, the peak yaw rate of the tractor and the semitrailer, the rearward amplification and the peak off-tracking.

The rearward amplification (RWA) is the ratio of the maximum value of a motion variable of the last vehicle unit to that of the first unit, under a specified steering maneuver and constant vehicle speed. The motion variable is usually the lateral acceleration of the center of gravity (Eq. 3.21) or the yaw rate. This metric quantifies the change of the motion variable from the first unit towards the last unit and indicates the risk of roll over of the rear units. Higher values show higher risk, and the best value of RWA is equal to one [54].

$$RWA_{a_y} = \frac{a_{y2}^{max}}{a_{y1}^{max}} \tag{3.21}$$

When a truck is turning at low speed, the rear wheels do not follow the same path as the front, and thus the vehicle requires more space, than it would for driving straight. The low speed off-tracking ($LSOT$) is defined as "the distance between the outermost and innermost paths of mid-points of axles" [51]. This metric expresses the required road space that the truck needs, when turning, and the lower the speed the higher the value of the metric [55]. Let R_f and R_r be the path radius for the front axle of the first unit and the path radius for the rearmost axle of the last unit, then the $LSOT$ is found using Equation 3.22.

$$LSOT = R_f - R_r \tag{3.22}$$

4

Model Validation through Bayesian Hypothesis Testing

In this chapter, the methodology used for the vehicle model validation is introduced, with specific interest on the Bayesian hypothesis testing approach. The methodology presented is then applied to the tractor-semitrailer model (Chapter 3) following the theory discussed in Chapter 2.

4.1 Vehicle models for validation

The first step of the model validation is to determine the models used. The implementation model can be either a high order and complex vehicle model or driving data acquired through real driving experiments. This thesis implements the validation process using a high order vehicle model, provided by VAS. The abstract model, i.e. the model to be validated, is the tractor-semitrailer model presented in Chapter 3. The validation is performed by comparing the dynamic behavior of those models, in terms of metrics, under a specified driving test case, by utilizing the methods described in Chapter 2.

4.2 Data generation

In this work, since the implementation model is provided for simulations, a direct comparison between the chosen metrics of the models is possible. Both models are simulated with the same vehicle parameter values and the relative difference between the metrics d_x (Eq. 4.1) is used to decide the validity of the abstract model. Here, x is the performance metric and the indices I, A refer to the implementation and the abstract model, respectively. The absolute value of the difference ($|x_A - x_I|$) is used, since only the deviation between the metric values is of interest. Moreover, the parallel simulation of the two models is performed not only for simplicity but also because some parameters of the abstract model, that change under different operating conditions, require information provided by the implementation model.

$$d_x = Re = \frac{|x_A - x_I|}{x_I} \tag{4.1}$$

To create the data sets required for the statistical analysis in the validation process, one parameter of the vehicle models is varied. The additional semitrailer mass, called payload, is selected, and its distribution is assumed uniform. Through Monte

Carlo sampling, random payload values are drawn and used to simulate both the implementation and the abstract model (same values for both models).

There are various reasons for choosing the payload as the varied vehicle parameter. First, the additional mass changes the normal forces. However, the abstract model is statically indeterminate since the semitrailer has more than two axles, and thus the normal forces cannot be found by static equilibrium. For that matter, the implementation model is simulated first, and the calculated normal forces are then used as input to the abstract model. Another reason is that the implementation model includes load sharing between the tractor and the semitrailer, unlike the abstract model. To incorporate that in the latter, the additional load, in terms of mass, on each unit i is calculated. Let F_{zi}^{old} and $\sum F_{zi}^{new}$ be the normal forces on unit i before and after the increase of payload, respectively, then the added mass on the unit is calculated as $m_{add,i} = (\sum F_{zi}^{new} - \sum F_{zi}^{old})/g$. Finally, the complexity of the implementation model is restrictive as to which vehicle parameters can be varied. For the abstract model, the variation of the parameters is straightforward. On the contrary, the implementation model's structure includes multiple bodies, which are interconnected. Changing a parameter that is calculated based on other parameters (and which are not used in the abstract model), cannot be done in an obvious way.

After the simulation of the models, the relative difference d_x for each evaluated metric x is computed. The relative difference d_x is located within the [0, 1] interval, where 0 corresponds to 0% and 1 to 100%, in terms of relative difference. These data are then used for the model validation, which is conducted as described in the following section (Ch. 4.3).

4.3 Bayesian Hypothesis Testing

Both types of hypothesis testing (point- and interval-based) can be used, even when the available data sets are drawn for different vehicle model parameter values, e.g., when dealing with real driving data. The choice between these types depends on the application, the level of accuracy required, and the hypotheses themselves. As mentioned in Chapter 2, the point-based approach is stricter, and the null hypothesis (equality to a single value) is harder to be accepted. The abstract model is expected to differ from the implementation model. Thus, applying the point-based hypothesis testing will always deem the abstract model invalid. On the other hand, the interval-based hypothesis testing is most likely a better approach. For that matter, this thesis will focus only on the interval-based hypothesis testing.

The validation of the abstract model using the interval-based method requires the statement of the null (H_0) and at least one alternative hypothesis (H_1), regarding the relative difference d_x (observed data). The hypothesis testing, in this work, is applied on the location parameter of the distribution that describes the data. This distribution is the likelihood $p(y|\theta_x)$, where $y = d_x$ is the relative difference of metric x and θ_x are the unknown parameters of the likelihood distribution. The likelihood is found by comparing the data to a set of well-known distribution families, using the sum of squared errors SSe as the goodness-of-fit measure. SSe is calculated according to Equation 4.2, where y_i is the i^{th}-value of the observed data, $\hat{y}_i$ is the i^{th}-value of the probability density function and df is the statistical degrees of

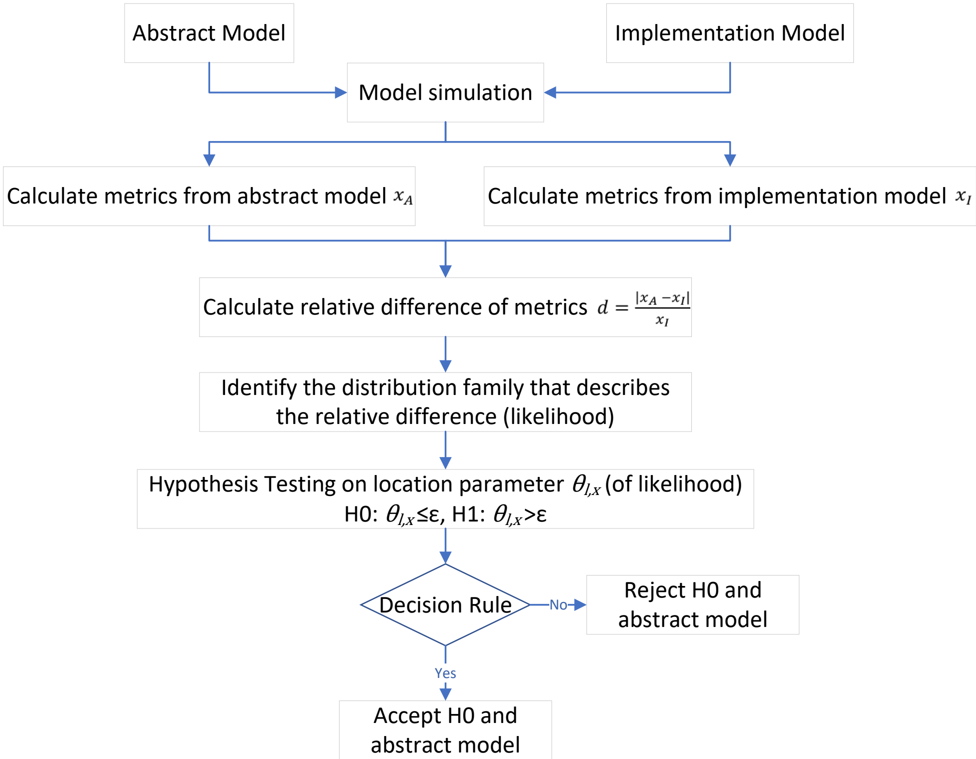


Figure 4.1: Model Validation Process. The metrics x are calculated from the vehicle models simulations, and the relative difference d_x between them is used in the hypothesis testing. Based on that the decision about the abstract model's validity is made. The decision rule depends on the selected hypothesis testing approach.

freedom. The distribution family that provides the lowest SSe describes the data best and thus it is chosen as the likelihood distribution. Please note that the set of distributions evaluated is restricted, by PyMC3 package, which includes a specific number of available distributions (Appendix A.1).

$$SSe = \frac{\sum(\hat{y}_i - y_i)^2}{df} \tag{4.2}$$

Following the formulation described in Chapter 2.2.2, the interval-based hypothesis testing expression for the relative difference d_x between the metrics of each vehicle model (x) is stated in Equation 4.3. The location parameter of the likelihood distribution of metric x is denoted as $\theta_{l,x}$. The null hypothesis ($H_{0,x}$) involves only an upper limit (ϵ_x), since the data are bounded from below (data are greater or equal to 0). It should be noted that in this work the hypothesis testing involves only one statistical parameter (location). However, this formulation can be extended for multivariate cases, with the addition of statements for the rest of the statistical parameters θ_x. The following chapters (Ch. 4.4, 4.5) describe two approaches to the interval-based hypothesis testing.

$$H_{0,x} : \theta_{l,x} \leq \epsilon_x \qquad H_{1,x} : \theta_{l,x} > \epsilon_x \tag{4.3}$$

4.4 Parameter Estimation

The model validation can be performed through hypothesis testing formulated as parameter estimation. In this context, the two hypotheses are not stated directly in the statistical model. More specifically, the relative difference d_x of a metric x is modelled by specifying the likelihood ($p(y|\theta_x)$, $y = d_x$) to describe the data and the priors for the unknown statistical parameters θ_x, without taking into account the hypotheses. After the inference, the posterior distributions ($p(\theta_x|y)$, $y = d_x$) of the parameters are computed. The hypothesis testing is then applied on these posteriors and the decision is made based on the posterior summary (HDI, ROPE). It should be reminded that the data handled are located in the [0, 1] interval.

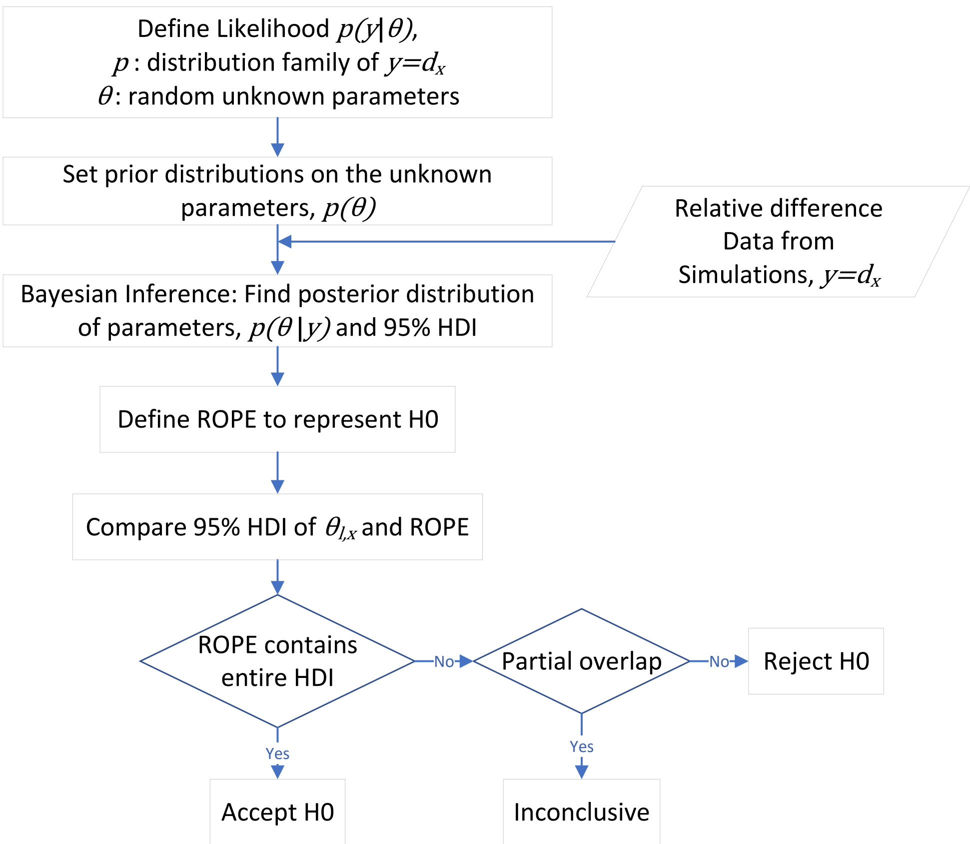


Figure 4.2: Hypothesis Testing as Parameter Estimation. ...

The parameter estimation method begins with structuring the statistical model M_x. The first component of M_x is the likelihood of the observed data $p(y|\theta_x)$, where $y = d_x$ is the relative difference of metric x and θ_x are the parameters of the likelihood distribution. Depending on the distribution family used for the likelihood, the statistical parameters θ_x can be different. The second part of the statistical model is the definition of the prior distributions $p(\theta_x)$ of the parameters θ_x of the likelihood. The statistical parameters θ_x are unknown and the prior distributions of θ_x reflect the available knowledge prior of observing the data. The uniform distribution $U(0, 1)$ is assumed for the location parameter ($\theta_{l,x}$), since the data are located in [0, 1]. The rest of the parameters of the likelihood are either kept constant or non-informative priors are applied reflecting any existing (or lack of) prior beliefs and knowledge.

An example of a statistical model (priors and likelihood) is shown in Equation 4.4, for a Normal likelihood ($p(y|\theta_x) \sim N(\mu_x, \sigma_x)$, $\theta_x = \langle\mu_x, \sigma_x\rangle$)). The standard deviation σ_x is considered constant, and only for the location parameter ($\theta_{l,x} = \mu_x$) a prior is defined. The non-informative prior chosen is a uniform distribution $p(\mu_x) \sim U(0, 1)$. Similar methodology is followed for any other likelihood distribution, that involves different sets of statistical parameters θ_x.

Using the inference engines, NUTS for the continuous and Metropolis for the discrete parameters (Ch. 2.1.6), the posterior distributions of the parameters in the statistical model are found. The hypothesis testing and decision making are done based on these posteriors as follows. In Chapter 2.1.4, where the posterior summary was discussed, two tools were introduced, namely the Highest Density Interval (HDI) and the Region Of Practical Equivalence (ROPE). These tools are used to determine if the location parameter $\theta_{l,x}$ is within the acceptable range. The ROPE represents the null interval ($[0, \epsilon_x]$), while the HDI contains the most credible values of the location parameter. Only when the ROPE contains the entire HDI, one can conclude that all the credible values are within the null interval and therefore the null hypothesis (and the abstract model) can be accepted. In the other cases, either no conclusion can be made (partial overlap of HDI and ROPE), or the null hypothesis is rejected (HDI not included in ROPE), as discussed in Chapter 2.1.4.

$$\begin{aligned} &M_x : \\ &\quad p(y|\theta_x) \sim N(\mu_x, \sigma_x) \\ &\quad p(\mu_x) \sim U(0, 1) \\ &\quad \sigma_x = constant\ (= std(y)) \end{aligned} \tag{4.4}$$

4.5 Model Comparison of Statistical Models

Similar to the parameter estimation approach, a statistical model representing the hypothesis testing formulation must be defined. In the model comparison approach, each hypothesis is modelled as a distinct statistical model $M_{x,m}$. Each model is indicated by the model index parameter m, which is a discrete variable $m \in \{0, 1, \ldots, n\}$, and n is the number of alternative hypotheses. In this work, where only one alternative hypothesis is considered, the model index is either 0, for the null hypothesis ($M_{x,0}$), or 1 for the alternative hypothesis ($M_{x,1}$). Since variable m only takes the values of 0 or 1, then $Pr(m = 1) = 1 - Pr(m = 0)$. As a reminder, the null hypothesis is that the relative difference of a metric x (d_x) is lower than an upper limit ($d_x \leq \epsilon_x$), while the alternative is that d is greater than the limit ($d_x \geq \epsilon_x$) (Eq.

4.3). The distinct statistical models $M_{x,m}$ are combined into one hierarchical model M_x, which has as hyper-parameter (top-level parameter) the model index m.

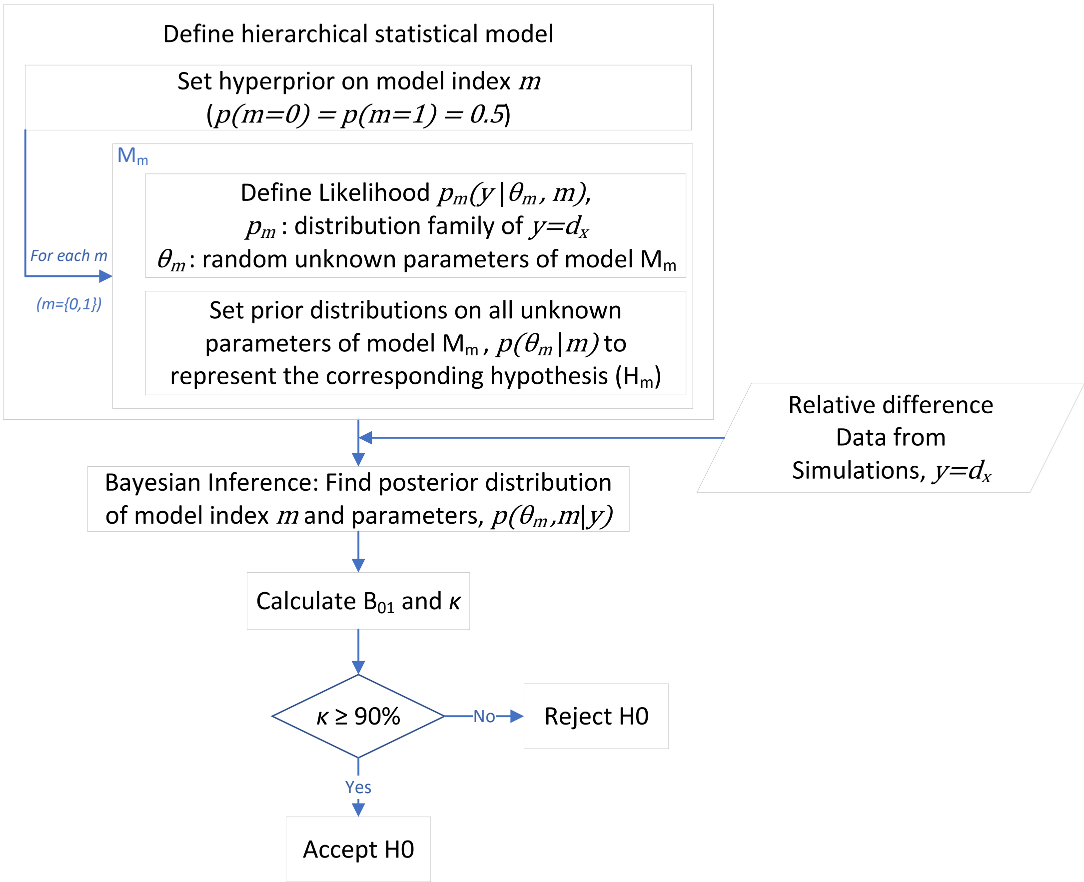


Figure 4.3: Hypothesis Testing as Model Comparison of Statistical Models ....

Figure 4.3 shows the process followed in the model comparison approach. First, the prior for model index $Pr(m)$ is defined. Model index is sampled from a categorical distribution. Since no prior knowledge exists regarding which hypothesis is more probable, equal probabilities are given to the model index values ($Pr(m = 0) = Pr(m = 1) = 0.5$). For each statistical model $M_{x,m}$, the likelihood of the data and the prior distributions of the parameters must be set. All the hypotheses, as they are formulated in this work, correspond to the same likelihood of the data ($p_m(y|\theta_{x,m}, m) = p(y|\theta_{x,m}, m)$, $y = d_x$), in terms of distribution family. Therefore, the hierarchical model includes one likelihood, which is shared between the statistical models $M_{x,m}$. In each $M_{x,m}$, different priors are set on the parameters of the same likelihood, to reflect the hypothesis H_m.

The prior of the location parameter $\theta_{l,x,m}$ of the likelihood $p(y|\theta_{x,m}, m)$ is defined to represent the hypothesis testing formulation. In model $M_{x,0}$, the prior distribution of $\theta_{l,x,0}$ stands for the null interval $[0, \epsilon_x]$, while in model $M_{x,1}$ the prior of $\theta_{l,x,1}$ stands for the alternative interval $(\epsilon_x, 1]$. The non-informative uniform distribution is selected for both priors, since no other knowledge exists. For the rest of the statistical parameters $\theta_{x,m}$, the priors are set up as discussed in Chapter 4.4.

Equation 4.5 shows an example of a hierarchical model M_x, for a Normal likelihood ($p(y|\theta_{x,m}, m) \sim N(\mu_{x,m}, \sigma_{x,m})$, $\theta_{x,m} = \langle\mu_{x,m}, \sigma_{x,m})\rangle$) for the hypothesis testing in Equation 4.3. The null and alternative hypothesis are assumed to be equally probable, so the prior probability of model index is 0.5 for each value $(0, 1)$. The standard deviation is considered constant in both models $M_{x,0}$ and $M_{x,1}$, so priors are defined only for the location parameters ($\mu_{x,0}$, $\mu_{x,1}$). A non-informative uniform prior is chosen, with bounds defined from the hypothesis formulation. It should be noted again that this methodology can be applied for any likelihood distribution, involving different statistical parameters $\theta_{x,m}$.

$$
\begin{aligned}
M_x :& \\
& Pr(m = 0) = Pr(m = 1) = 0.5 \\
M_{x,0} :& \\
& p(y|\theta_{x,0}) \sim N(\mu_{x,0}, \sigma_{x,0}) \\
& p(\mu_{x,0}) \sim U(0, \epsilon_x) \\
& \sigma_{x,0} = constant\ (= std(y)) \\
M_{x,1} :& \\
& p(y|\theta_{x,1}) \sim N(\mu_{x,1}, \sigma_{x,1}) \\
& p(\mu_{x,1}) \sim U(\epsilon_x, 1) \\
& \sigma_{x,1} = constant\ (= std(y))
\end{aligned}
\tag{4.5}
$$

The model index and the parameters of each statistical model $M_{x,m}$ are inferred using the inference engines and the posterior distributions are computed. The hypothesis testing is then concluded by calculating the Bayes factor, using Equation 2.13 in Chapter 2.2.4. As mentioned before, the Bayes factor B_{01} and confidence κ show the strength of the evidence that support or reject the null hypothesis and the level of confidence in that decision. Based on Table 2.1, conclusions about the validity of the abstract model can be made. In this work, the abstract model is considered valid, when the evidence of that are at least strong ($\kappa \geq\sim 90\%$). Naturally, the selected strength of evidence depends on the application. The higher the strength the more confidence exists in accepting the abstract model. In this work, strong evidence are considered sufficient, however for more safety-critical cases very strong or extreme evidence might be preferred.

5

Vehicle Models Simulation & Validation Results

This chapter presents the results of the validation framework, implemented on a specific test case for a tractor-semitrailer vehicle. The validity of the abstract model is studied under three cases. The first concerns the calibrated abstract model, while the other two cases regard the incorrect tuning of a parameter of the abstract model. The vehicle models are simulated for a 90 ° turn with constant speed (30 km/h).

5.1 Model Calibration

The parameter values of the abstract model of a tractor-semitrailer combination are calibrated based on the implementation model design. Since the latter model is assumed to correspond to reality, the values of its parameters are used for the abstract model as well. However, due to the discrepancy in the modelling of the physical quantities in each vehicle model, some of the parameters are defined approximately. Table 5.1 presents the parameter values used in the single track model.

Table 5.1: Vehicle parameters values after calibration.

Parameter	**Value**
m_1 [kg]	7318.2
m_2 [kg]	7030.0
$I_{zz,1}$ [kgm^2]	26264
$I_{zz,2}$ [kgm^2]	82661
l_{11} [m]	1.00
l_{12} [m]	2.70
l_{21} [m]	1.36
l_{22} [m]	2.66
l_{23} [m]	3.96
$l_{c,12}$ [m]	2.12
$l_{c,21}$ [m]	5.44
$C_{cy,11}$ [rad^{-1}]	5.32
$C_{cy,12}$ [rad^{-1}]	6.13
$C_{cy,21}$ [rad^{-1}]	6.15
$C_{cy,22}$ [rad^{-1}]	6.15
$C_{cy,23}$ [rad^{-1}]	6.15

5.2 Test Case - 90 degree Turn

To perform the validation of the tractor-semitrailer model, the 90 degree turn maneuver was selected. The following sections describe the general setup of the test case, as well as the test case variations that are studied.

5.2.1 Simulation Setup

The driving context under which the 90 degree turn is performed is city driving. The chosen scenario involves the tractor-semitrailer vehicle to perform a 90° left turn on a flat and dry road (asphalt), without lane markings. The scene does not include other dynamic objects and the environmental conditions are considered for a typical day.

To simulate the vehicle models, the steering input profile and the velocity must be defined. The steering input profile is defined as shown in Figure 5.1. It is generated using Equation 5.1, for time period of 17 s and maximum amplitude (A_δ) equal to 0.07 rad ($\sim 4°$). In Equation 5.1, t_0 is 2 s, before which the steering input is 0. The duration from t_0 to t_1 and from t_2 to t_3 is 2.5 s, while the duration of constant maximum amplitude A_δ is 8 s ($t_1 = 4.5$ s, $t_2 = 12.5$ s, $t_3 = 15$ s). After t_3, the steering input is again 0 for 2 s. As far as the velocity is concerned, it is considered constant at 30 km/h (~ 8.33 m/s).

$$\delta_f = \begin{cases} \frac{3A_\delta}{(t_1-t_0)^2} \cdot t^2 - \frac{2A_\delta}{(t_1-t_0)^3} \cdot t^3 & , \ t \leq t_1 \\ A_\delta & , \ t_1 \leq t \leq t_2 \\ A_\delta - \frac{3A_\delta}{(t_3-t_2)^2} \cdot t^2 + \frac{2A_\delta}{(t_3-t_2)^3} \cdot t^3 & , \ t_2 \leq t \leq t_3 \end{cases} \tag{5.1}$$

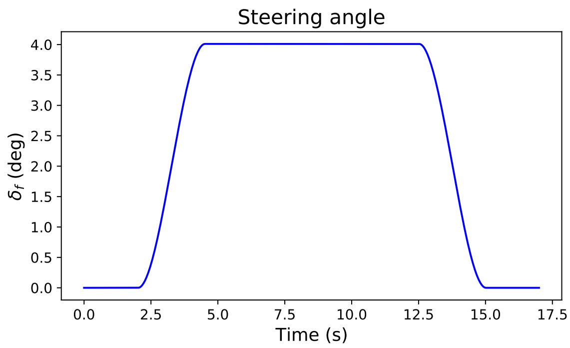


Figure 5.1: Steering angle input profile.

The payload fraction is chosen as the varied vehicle parameter, which ranges between 0 and 1, where 0 and 1 are equivalent to 0 tn and 24 tn, respectively. For the simulations, it is assumed that the payload fraction follows the uniform distribution $payload_F \sim U(0, 1)$ (Figure 5.2). The abstract and the implementation model are simulated simultaneously for each value of the payload. First, the implementation model is simulated. Because of the structure of the model, the velocity is not exactly constant, but it is controlled to be around the reference value 30 km/h, using a simple P controller with gain $K = 10$. The velocity profile, computed from the implementation model, is then used in the abstract model simulation. Moreover, due to changes in normal forces and the load transfer between the tractor and the semitrailer in the implementation model, the aforementioned quantities are also used as input to the abstract model. This is done to avoid additional uncertainties of the abstract model, since it does not account for the inertia change.

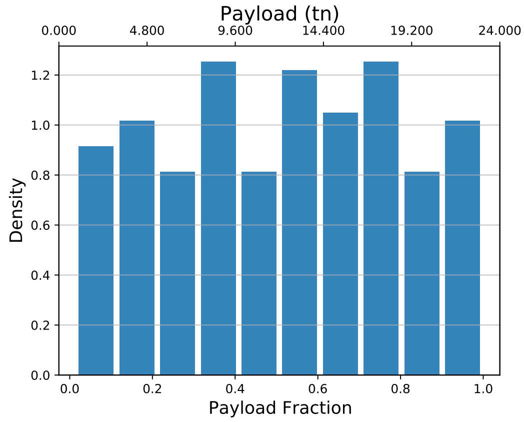
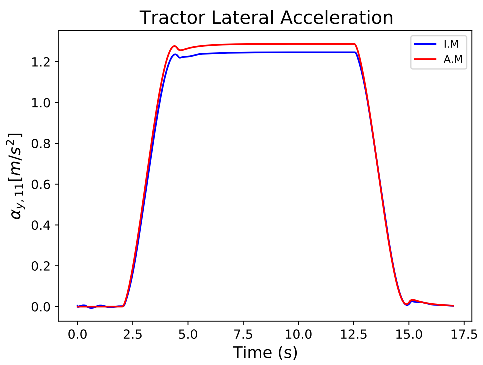


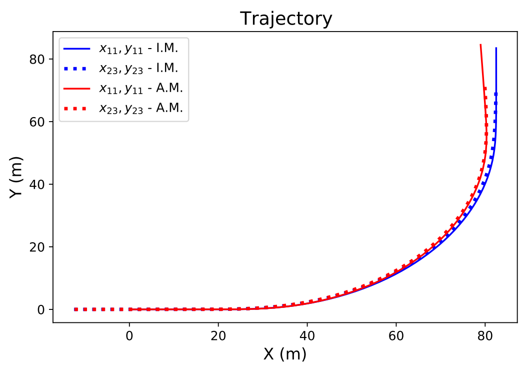
Figure 5.2: Histogram of payload fraction parameter, sampled from uniform distribution ($payload_F \sim U(0, 1)$).

5.2.2 Test Cases Setup & Vehicle Responses

The first case (TC1) investigated is when the abstract model is calibrated correctly, and therefore its parameters have the values presented in Table 5.1. Figure 5.3 illustrates the lateral acceleration response of the tractor and the vehicle trajectories of the abstract and implementation models, for zero payload. The abstract model responses are relatively similar to the ones computed from the implementation model, and the discrepancies between the vehicle models seem reasonable.



(a) Tractor lateral acceleration

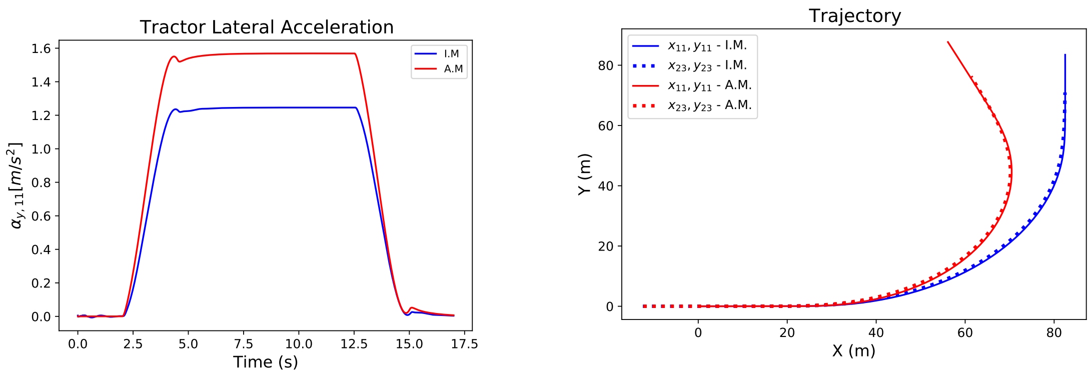


(b) Trajectory

Figure 5.3: TC1 (Calibrated Value) - Lateral acceleration of tractor and trajectories for the abstract (A.M.) and the implementation model (I.M.).

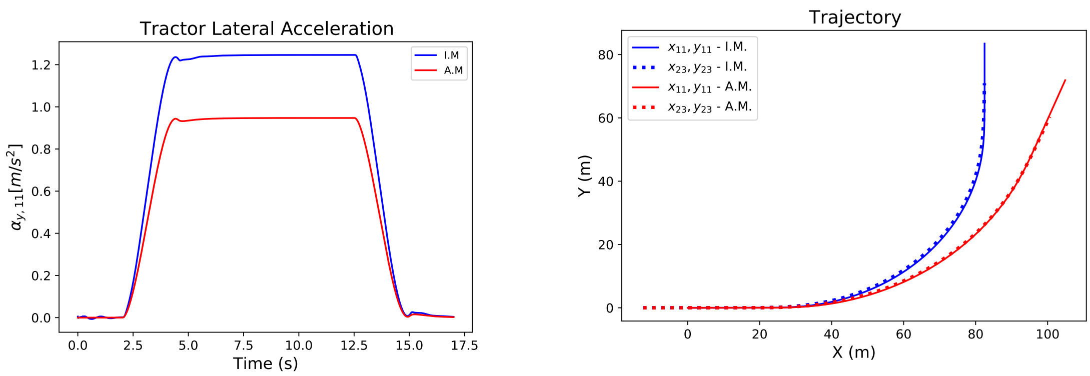
Two more test cases are simulated, both for an erroneous vehicle parameter value of the abstract model. This parameter is the cornering coefficient of the tractor's front axle ($C_{cy,11}$). In the second test case (TC2) the cornering coefficient $C_{cy,11}$ is set as double the calibrated value and equal to $10.64\,\mathrm{rad}^{-1}$. Similarly, in the third case (TC3) the value of $C_{cy,11}$ is tuned to half the calibrated value and equal to $2.66\,\mathrm{rad}^{-1}$. The tractor lateral acceleration and trajectories for zero payload are illustrated in Figures 5.4 and 5.5 for these two cases, respectively. It is evident, that when a vehicle parameter of the abstract model has incorrect value, the vehicle responses greatly differ from those calculated by the implementation model. More specifically, when the cornering coefficient $C_{cy,11}$ equals to double the calibrated value, the abstract model overestimates the responses. On the other hand, when $C_{cy,11}$ is half the calibrated value, the abstract model underestimates the vehicle responses.

The additional vehicle state responses for each test case investigated are shown in Appendix B.1. Table 5.2 summarizes the cornering coefficient values of the tractor's front axle for all three test cases under investigation.



(a) Tractor lateral acceleration **(b)** Trajectory

Figure 5.4: TC2 (Double Value) - Lateral acceleration of tractor and trajectories for the abstract (A.M.) and the implementation model (I.M.).



(a) Tractor lateral acceleration **(b)** Trajectory

Figure 5.5: TC3 (Half Value) - Lateral acceleration of tractor and trajectories for the abstract (A.M.) and the implementation model (I.M.).

Table 5.2: Cornering coefficient values of tractor's front axle for the three test cases.

Test Case	$C_{cy,11}$ [rad^{-1}]
TC1	5.32
TC2	10.64
TC3	2.66

5.3 TC1 - Calibrated Parameter Value

The test case investigated here concerns the calibrated abstract model. The likelihood that describes the relative difference between the vehicle models is first identified and then the two Bayesian hypothesis testing approaches are applied.

5.3.1 TC1 - Likelihood of Data

The histograms of the relative difference d_x (observed data) for each metric x are illustrated in Figure 5.6. The data are compared with a set of applicable distributions (Appendix A.1), using the sum of square error (SSe) as a measure of goodness of fit, to identify which distribution family describes the data best. Table 5.3 presents the likelihood distribution family for the relative difference of each metric.

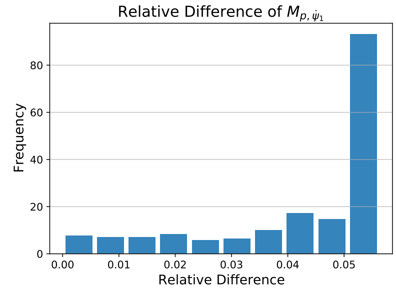
The rearward amplification (RWA_{a_y}) and the rise time of the tractor's lateral acceleration ($t_{r,a_{y1}}$) follow the asymmetric Laplace distribution $p(y|\theta) \sim AL(\mu, b, \kappa^{AL})$. The hypothesis testing (Eq. 4.3) is implemented on the location parameter ($\theta_{l,x} = \mu_x$), where x is the metric under investigation. The scale (b_x) and symmetry (κ_x^{AL}) parameters are also assumed unknown, and are inferred alongside the location parameter. The rest of the metrics studied ($M_{p,\dot{\psi}_1}$, $M_{p,\dot{\psi}_2}$, $M_{p,a_{y1}}$, $M_{p,a_{y2}}$, $t_{r,a_{y2}}$ and $M_{p,LSOT}$) follow the beta distribution $p(y|\theta) \sim Beta(\mu, \sigma)$, where μ is the mean and σ is the standard deviation. The parameter on which the hypothesis testing is applied is the mean ($\theta_{l,x} = \mu_x$), while the standard deviation σ_x is considered constant.

It should be noted that the additional parameters inside the likelihood distributions are not included in the hypothesis testing. They are only inferred to obtain the correct posterior. The upper limit ϵ_x used in the hypothesis testing is shown in Table 5.4 for each metric x.

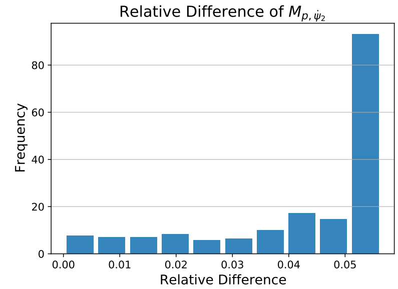
5.3.2 TC1 - Hypothesis Testing as Parameter Estimation (P.E.)

5.3.2.1 TC1 - Statistical Model (P.E.)

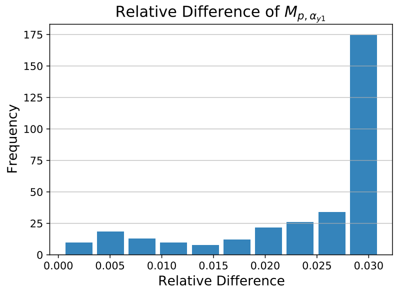
The prior distribution of the location parameter μ_x is assumed uniform, ranging from 0 to 1, for all the evaluated metrics x ($\theta_{l,x} = \mu_x \sim U(0, 1)$). Metrics $M_{p,\dot{\psi}_1}$, $M_{p,\dot{\psi}_2}$, $M_{p,a_{y1}}$, $M_{p,a_{y2}}$, $t_{r,a_{y2}}$ and $M_{p,LSOT}$ follow the beta distribution. The standard deviation σ_x is considered equal to the standard deviation of the relative difference of the metric ($std(d_x)$). Equation 5.2 shows the statistical model for the peak value



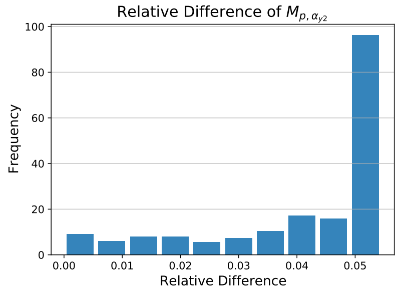
(a)



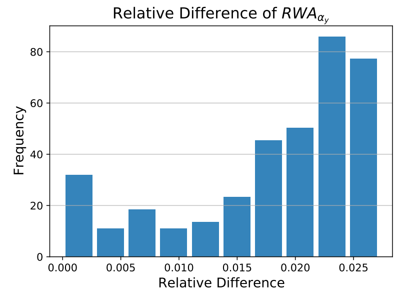
(b)



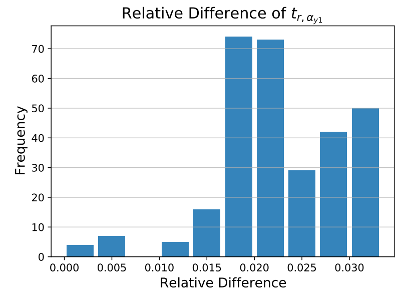
(c)



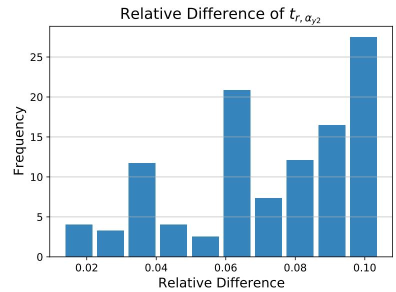
(d)



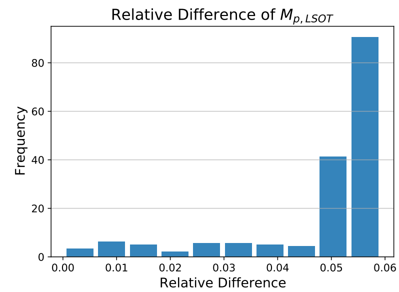
(e)



(f)



(g)



(h)

Figure 5.6: TC1 - Histogram of relative difference d. 5.6a Peak yaw rate of tractor $M_{p,\dot{\psi}_1}$, 5.6b Peak yaw rate of semitrailer $M_{p,\dot{\psi}_2}$, 5.6c Peak lateral acceleration of tractor $M_{p,a_{y1}}$, 5.6d Peak lateral acceleration of semitrailer $M_{p,a_{y2}}$, 5.6e Rearward amplification RWA_{a_y}, 5.6f Rise time of lateral acceleration of tractor $t_{r,a_{y1}}$, 5.6g Rise time of lateral acceleration of semitrailer $t_{r,a_{y2}}$, 5.6h Peak low speed off-tracking $M_{p,LSOT}$

Table 5.3: Likelihood distribution family and parameters of the relative difference of the evaluated metrics (TC1).

Metric	**Likelihood**	**Parameters**
$M_{p,\dot{\psi}_1}$	Beta	$\theta_{M_{p,\dot{\psi}_1}} = \left\langle \mu_{M_{p,\dot{\psi}_1}},\ \sigma_{M_{p,\dot{\psi}_1}} \right\rangle$
$M_{p,\dot{\psi}_2}$	Beta	$\theta_{M_{p,\dot{\psi}_2}} = \left\langle \mu_{M_{p,\dot{\psi}_2}},\ \sigma_{M_{p,\dot{\psi}_2}} \right\rangle$
$M_{p,a_{y1}}$	Beta	$\theta_{M_{p,a_{y1}}} = \left\langle \mu_{M_{p,a_{y1}}},\ \sigma_{M_{p,a_{y1}}} \right\rangle$
$M_{p,a_{y2}}$	Beta	$\theta_{M_{p,a_{y2}}} = \left\langle \mu_{M_{p,a_{y2}}},\ \sigma_{M_{p,a_{y2}}} \right\rangle$
RWA_{a_y}	Asymmetric Laplace	$\theta_{RWA_{a_y}} = \left\langle \mu_{RWA_{a_y}},\ b_{RWA_{a_y}},\ \kappa^{AL}_{RWA_{a_y}} \right\rangle$
$t_{r,a_{y1}}$	Asymmetric Laplace	$\theta_{a_{y1},t_r} = \left\langle \mu_{t_{r,a_{y1}}},\ b_{t_{r,a_{y1}}},\ \kappa^{AL}_{t_{r,a_{y1}}} \right\rangle$
$t_{r,a_{y2}}$	Beta	$\theta_{t_{r,a_{y2}}} = \left\langle \mu_{t_{r,a_{y2}}},\ \sigma_{t_{r,a_{y2}}} \right\rangle$
$M_{p,LSOT}$	Beta	$\theta_{M_{p,LSOT}} = \left\langle \mu_{M_{p,LSOT}},\ \sigma_{M_{p,LSOT}} \right\rangle$

Table 5.4: Maximum allowed limit of relative difference of studied metrics.

Metric	**Limit** ϵ_x
$M_{p,\dot{\psi}_1}$	6%
$M_{p,\dot{\psi}_2}$	6%
$M_{p,a_{y1}}$	3%
$M_{p,a_{y2}}$	6%
RWA_{a_y}	3%
$t_{r,a_{y1}}$	3%
$t_{r,a_{y2}}$	8%
$M_{p,LSOT}$	6%

of tractor's yaw rate ($M_{p,\dot{\psi}_1}$). The statistical models of the other five metrics follow exactly the same formulation.

The statistical model of the rearward amplification (RWA_{a_y}) is shown in Equation 5.3. The scale ($b_{RWA_{a_y}}$) and symmetry ($\kappa^{AL}_{RWA_{a_y}}$) parameters are both assumed to follow a uniform distribution, since no prior knowledge is available. This formulation applies also to the rise time of the tractor's lateral acceleration ($t_{r,a_{y1}}$).

$$
\begin{aligned}
&M_{M_{p,\dot{\psi}_1}}: \\
&\qquad p(y|\theta_{M_{p,\dot{\psi}_1}}) \sim Beta(\mu_{M_{p,\dot{\psi}_1}}, \sigma_{M_{p,\dot{\psi}_1}}) \\
&\qquad p(\mu_{M_{p,\dot{\psi}_1}}) \sim U(0,1) \\
&\qquad \sigma_{M_{p,\dot{\psi}_1}} = constant\ (= std(d_{M_{p,\dot{\psi}_1}}))
\end{aligned}
\tag{5.2}
$$

$$
\begin{aligned}
&M_{RWA_{a_y}}: \\
&\qquad p(y|\theta_{RWA_{a_y}}) \sim AL(\mu_{RWA_{a_y}}, b_{RWA_{a_y}}, \kappa^{AL}_{RWA_{a_y}}) \\
&\qquad p(\mu_{RWA_{a_y}}) \sim U(0,1) \\
&\qquad p(b_{RWA_{a_y}}) \sim U(0,5) \\
&\qquad p(\kappa^{AL}_{RWA_{a_y}}) \sim U(0,50)
\end{aligned}
\tag{5.3}
$$

5.3.2.2 TC1 - Posterior Predictive Checks

Figure 5.7 illustrates the posterior predictive checks for each evaluated metric x, and Table 5.5 shows the Bayesian p-values. In each plot in Figure 5.7, the black line is a KDE (Kernel Density Estimate) of the observed data (d_x). A KDE plot is another visualization method of a dataset's distribution, comparable to a histogram. The semitransparent blue lines are KDE plots computed from the posterior predictive samples. These curves depict the uncertainty in the distribution of the predicted data [33]. The dashed blue line is the mean posterior predictive curve. Using these figures, the statistical model can be evaluated in terms of how well it can reproduce the observed data.

The statistical models in Figure 5.7 cannot predict all the aspects of the observed data. For example, in Figure 5.7a the posterior predictive KDEs cannot predict the high density of the observed data (relative difference of the peak tractor's yaw rate). The Bayesian p-value is 0.44, which indicates a small bias in the statistical modelling. Since the p-value is less than 0.5, the observed values are more extreme than the predicted ones. However, it is still close to 0.5, and therefore the statistical modelling is accepted. On the other hand, the relative difference of the rearward amplification has a p-value of 0.15. This indicates that the statistical model cannot predict the mean of the actual data. Figure 5.7e also shows that the statistical model fails to capture other aspects, such as the multiple peaks, of the observed data.

Nevertheless, as far as the evaluation of the location (mean) of the distribution is concerned, the predictive curves can roughly predict the observed data, at least for most of the metrics investigated. The metrics, for which the statistical model is rejected, are excluded from the decision-making process to avoid incorrect conclusions about the anstract model's validity. Therefore, the current statistical modelling is accepted for this particular use, expect for the rearward amplification RWA_{a_y}. Please note that accepting the statistical model does not extend to validating the abstract model. It only means that the statistical modelling is considered good enough in the context of the study, and thus any decision based on that is satisfactory.

Table 5.5: Bayesian p-values (TC1).

Metric	$p - value$
$M_{p,\dot{\psi}_1}$	0.44
$M_{p,\dot{\psi}_2}$	0.44
$M_{p,a_{y1}}$	0.57
$M_{p,a_{y2}}$	0.43
RWA_{a_y}	0.15
$t_{r,a_{y1}}$	0.40
$t_{r,a_{y2}}$	0.51
$M_{p,LSOT}$	0.45

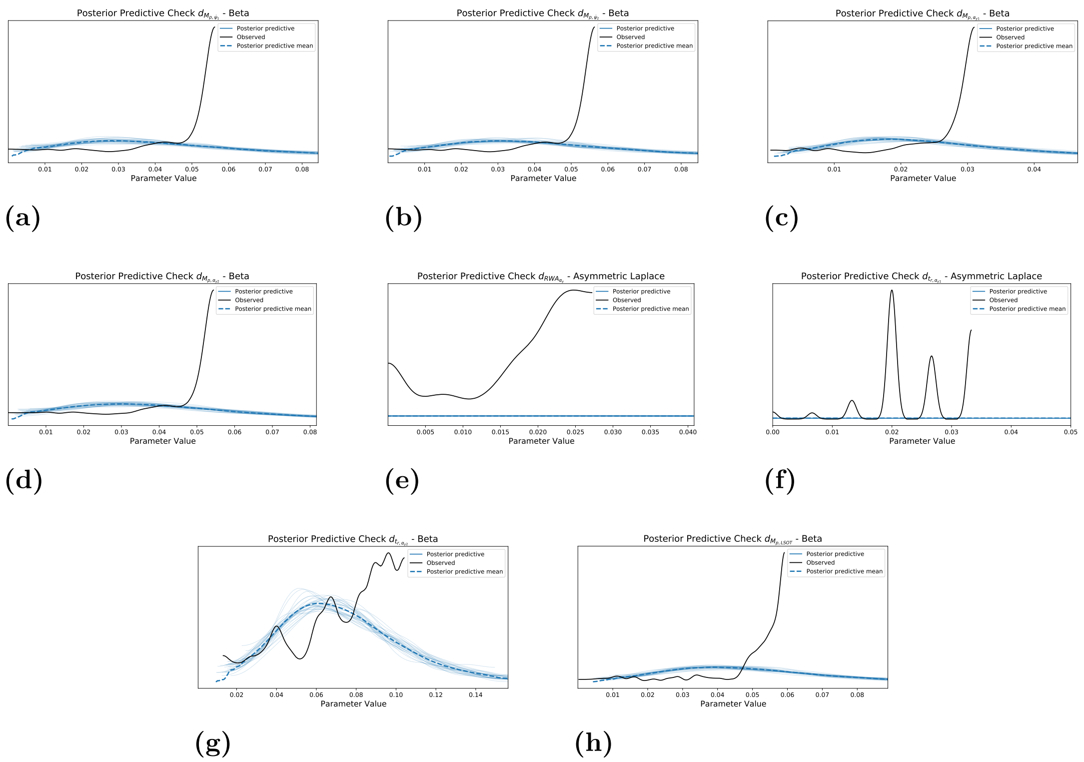
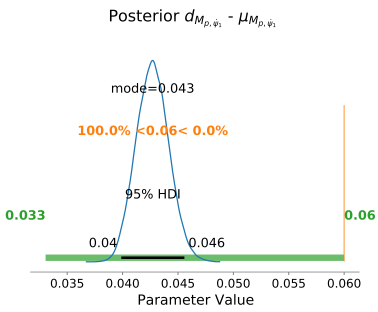


Figure 5.7: TC1 - Parameter Estimation - Posterior predictive check of relative difference d_x. 5.7a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, 5.7b Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, 5.7c Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, 5.7d Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, 5.7e Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, 5.7f Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$, 5.7g Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, 5.7h Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.

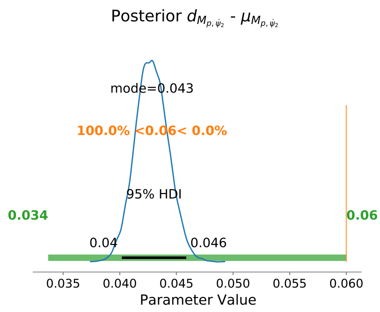
5.3.2.3 TC1 - Posterior Distributions and Decision (P.E.)

Figure 5.8 illustrates the posterior distributions of the location parameter ($\theta_{l,x}$) of the relative difference d_x of the studied metrics x, using as maximum allowed values those given in Table 5.4. Here, the results for all metrics are presented, even the ones that have biased statistical modelling (rearward amplification RWA_{a_y}). However, the best practice is to exclude the biased statistical models, in order to ensure the trustworthiness of the decision regarding the validity of the abstract model.

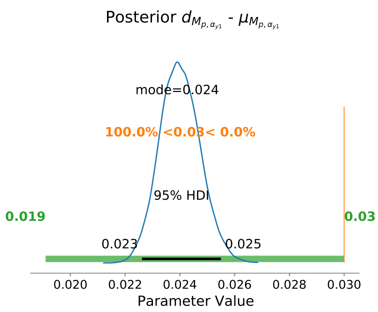
According to Figure 5.8, the 95% HDI is entirely within the ROPE interval, for all the evaluated metrics. It should be reminded that the ROPE interval reflects the range of allowed values of the relative difference (null interval). This means that for all metrics, the 95% most credible values are within the accepted range. Therefore, the abstract model can be considered valid (Table 5.6).



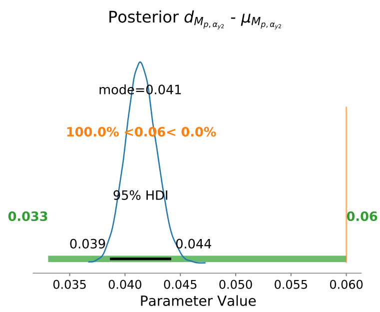
(a)



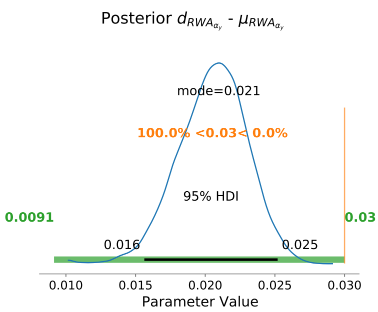
(b)



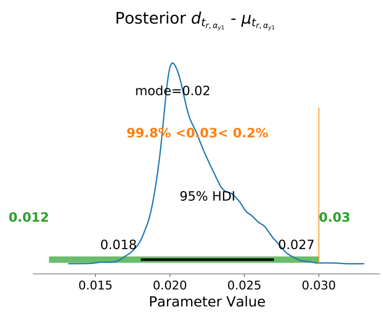
(c)



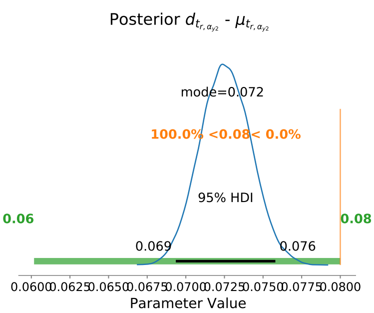
(d)



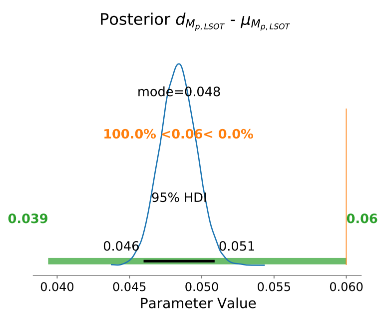
(e)



(f)



(g)



(h)

Figure 5.8: TC1 - Parameter Estimation - Posterior of location parameter ($\theta_{l,x}$) of relative difference d_x. 5.8a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, 5.8b Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, 5.8c Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, 5.8d Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, 5.8e Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, 5.8f Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$, 5.8g Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, 5.8h Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.

Table 5.6: Decision of abstract model's acceptance for each evaluated model following the Bayesian Hypothesis testing formulated as Parameter Estimation (TC1).

Metric	Decision
$M_{p,\dot{\psi}_1}$	Valid
$M_{p,\dot{\psi}_2}$	Valid
$M_{p,a_{y1}}$	Valid
$M_{p,a_{y2}}$	Valid
RWA_{a_y}	Valid
$t_{r,a_{y1}}$	Valid
$t_{r,a_{y2}}$	Valid
$M_{p,LSOT}$	Valid
Final Decision	Accepted

5.3.3 TC1 - Hypothesis Testing as Model Comparison of Statistical Models (M.D.)

5.3.3.1 TC1 - Statistical Model (M.D.)

The difference between the model comparison (of statistical models) and the parameter estimation approach lies on the structure of the statistical model, and especially on the prior distribution of the parameter on which the hypothesis testing is implemented (location parameter $\theta_{l,x}$). As mentioned in Chapter 2.2.4, the statistical model M_x in the model comparison approach is a hierarchical model, where the model index m_x is the hyper-parameter (top-level parameter). Similarly to the previous section, Equations 5.4 and 5.5 illustrate the statistical models for the relative difference of tractor's peak yaw rate ($M_{p,\dot{\psi}_1}$) and rearward amplification (RWA_{a_y}), respectively.

$$\begin{aligned}
&M_{M_{p,\dot{\psi}_1}}: \\
&\qquad Pr(m_{M_{p,\dot{\psi}_1}} = 0) = Pr(m_{M_{p,\dot{\psi}_1}} = 1) = 0.5 \\
&M_{M_{p,\dot{\psi}_1},0}: \\
&\qquad p(y|\theta_{M_{p,\dot{\psi}_1},0}) \sim Beta(\mu_{M_{p,\dot{\psi}_1},0}, \sigma_{M_{p,\dot{\psi}_1},0}) \\
&\qquad p(\mu_{M_{p,\dot{\psi}_1},0}) \sim U(0, \epsilon_{M_{p,\dot{\psi}_1}}) \\
&\qquad \sigma_{M_{p,\dot{\psi}_1},0} = constant\ (= std(d_{M_{p,\dot{\psi}_1}})) \\
&M_{M_{p,\dot{\psi}_1},1}: \\
&\qquad p(y|\theta_{M_{p,\dot{\psi}_1},1}) \sim Beta(\mu_{M_{p,\dot{\psi}_1},1}, \sigma_{M_{p,\dot{\psi}_1},1}) \\
&\qquad p(\mu_{M_{p,\dot{\psi}_1},1}) \sim U(\epsilon_{M_{p,\dot{\psi}_1}}, 1) \\
&\qquad \sigma_{M_{p,\dot{\psi}_1},1} = constant\ (= std(d_{M_{p,\dot{\psi}_1}}))
\end{aligned} \tag{5.4}$$

$$
\begin{aligned}
&M_{RWA_{a_y}}: \\
&\qquad Pr(m_{RWA_{a_y}} = 0) = Pr(m_{RWA_{a_y}} = 1) = 0.5 \\
&M_{RWA_{a_y},0}: \\
&\qquad p(y|\theta_{RWA_{a_y},0}) \sim AL(\mu_{RWA_{a_y},0}, b_{RWA_{a_y},0}, \kappa^{AL}_{RWA_{a_y},0}) \\
&\qquad p(\mu_{RWA_{a_y},0}) \sim U(0, \epsilon_{RWA_{a_y}}) \\
&\qquad p(b_{RWA_{a_y},0}) \sim U(0, 5) \\
&\qquad p(\kappa^{AL}_{RWA_{a_y},0}) \sim U(0, 50) \\
&M_{RWA_{a_y},1}: \\
&\qquad p(y|\theta_{RWA_{a_y},1}) \sim AL(\mu_{RWA_{a_y},1}, b_{RWA_{a_y},1}, \kappa^{AL}_{RWA_{a_y},1}) \\
&\qquad p(\mu_{RWA_{a_y},1}) \sim U(\epsilon_{RWA_{a_y}}, 1) \\
&\qquad p(b_{RWA_{a_y},1}) \sim U(0, 5) \\
&\qquad p(\kappa^{AL}_{RWA_{a_y},1}) \sim U(0, 50)
\end{aligned}
\tag{5.5}
$$

5.3.3.2 TC1 - Posterior Distributions and Decision (M.D.)

Figure 5.9 illustrates the prior and posterior distributions of the model index of the relative difference of the peak tractor yaw rate $M_{p,\dot{\psi}_1}$. The figure depicts how the prior belief of equal probability of the two hypotheses to be true is updated when the data (simulated relative difference) are observed. Table 5.7 shows the Bayes factor B_{01}, confidence κ and decision regarding the validity of the abstract model for each evaluated metric. It should be reminded that the abstract vehicle model is considered valid when the Bayesian confidence κ is greater or equal to 90%.

It is evident that the abstract model is valid for each metric. The Bayes factors approach infinity. This means that the evidence of the null hypothesis being more probable than the alternative are extreme. Therefore, the confidence in accepting the abstract model is 100%. These results agree with the ones from the parameter estimation approach in Chapter 5.3.2.3.

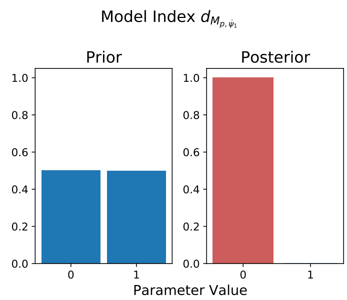


Figure 5.9: TC1 - Model Comparison of Statistical Models- Prior and Posterior distributions of model index of relative difference d_x of Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon\ = 6\%$.

Table 5.7: Bayes factor B_{01}, confidence κ and final decision about abstract model's validity for each evaluated model, following the Bayesian Hypothesis testing formulated as Model Comparison (TC1).

Metric	B_{01}	**Confidence** κ	**Decision**
$M_{p,\dot{\psi}_1}$	∞	100%	Valid
$M_{p,\dot{\psi}_2}$	∞	100%	Valid
$M_{p,a_{y1}}$	∞	100%	Valid
$M_{p,a_{y2}}$	∞	100%	Valid
RWA_{a_y}	∞	100%	Valid
$t_{r,a_{y1}}$	∞	100%	Valid
$t_{r,a_{y2}}$	∞	100%	Valid
$M_{p,LSOT}$	∞	100%	Valid
Final Decision			Accepted

5.4 TC2 - Double Parameter Value

In this test case, the cornering coefficient of the tractor's front axle ($C_{cy,11}$) is set to double the calibrated value. Similar to test case TC1, the likelihood the describes the data (relative difference of metrics) is first identified and then the hypothesis testing is implemented. In the following sections only the most important findings are presented, while the additional figures are provided in Appendix B.3.

5.4.1 TC2 - Likelihood of Data

Table 5.8 illustrates the likelihood distribution families for the relative difference of each metric. The families differ from the previous test case (TC1) for almost all of the metrics, except for $M_{p,a_{y1}}$ and $RWAa_y$. The peak yaw rate of the tractor ($M_{p,\dot{\psi}_1}$) and the semitrailer ($M_{p,\dot{\psi}_2}$), as well as the peak lateral acceleration of the semitrailer ($M_{p,a_{y2}}$) follow the Pareto distribution $p(y|\theta) \sim Pareto(\alpha^P, m)$. This particular distribution contains only a shape (α^P) and a scale (m) parameter. To perform the hypothesis testing, an algebraic equation that relates the central location ($\theta_{l,x}$) to one of the distribution's parameters is required. The central location is described by the mean of the Pareto distribution $\theta_{l,x} = \mu$, x is the metric under study. Using Equation 5.6, the shape parameter α^P is correlated to the distribution mean μ. The scale parameter m is considered unknown.

$$\alpha^P = \frac{\mu}{\mu - m} \tag{5.6}$$

The likelihood of the rise time of the tractor's lateral acceleration ($t_{r,a_{y1}}$) is a skew Normal distribution ($p(y|\theta) \sim SN(\mu, \sigma, \alpha^{SN})$). The location parameter, on which the hypothesis testing is applied, is μ. The other two parameters are also assumed unknown. The rise time of the semitrailer's lateral acceleration $t_{r,a_{y2}}$ is described by a triangular distribution ($p(y|\theta) \sim Triang(c, lower, upper)$). Here, the location parameter is the mode c, and the parameters $lower$ and $upper$ are considered constant and equal to the lower and upper bound of the relative difference ($lower = 0$, *upper* = 1). The peak low speed off-tracking $M_{p,LSOT}$ is described by the Student's t distribution ($p(y|\theta) \sim t(\mu, \sigma, \nu)$). The location parameter used in hypothesis testing is μ and the rest of the statistical parameters are assumed unknown. Finally, the peak tractor lateral acceleration ($M_{p,a_{y2}}$) data is described by the Beta distribution and the rearward amplification (RWA_{a_y}) data by the asymmetric Laplace. These distributions were described in test case TC1 (Ch. 5.3.1) and therefore they will not be explained here.

Table 5.8: Likelihood distribution family and parameters of the relative difference of the evaluated metrics (TC2).

Metric	Likelihood	Parameters
$M_{p,\dot{\psi}_1}$	Pareto	$\theta_{M_{p,\dot{\psi}_1}} = \left\langle \alpha^P_{M_{p,\dot{\psi}_1}},\ m_{M_{p,\dot{\psi}_1}} \right\rangle$
$M_{p,\dot{\psi}_2}$	Pareto	$\theta_{M_{p,\dot{\psi}_2}} = \left\langle \alpha^P_{M_{p,\dot{\psi}_2}},\ m_{M_{p,\dot{\psi}_2}} \right\rangle$
$M_{p,a_{y1}}$	Beta	$\theta_{M_{p,a_{y1}}} = \left\langle \mu_{M_{p,a_{y1}}},\ \sigma_{M_{p,a_{y1}}} \right\rangle$
$M_{p,a_{y2}}$	Pareto	$\theta_{M_{p,a_{y2}}} = \left\langle \alpha^P_{M_{p,a_{y2}}},\ m_{M_{p,a_{y2}}} \right\rangle$
RWA_{a_y}	Asymmetric Laplace	$\theta_{RWA_{a_y}} = \left\langle \mu_{RWA_{a_y}},\ b_{RWA_{a_y}},\ \kappa^{AL}_{RWA_{a_y}} \right\rangle$
$t_{r,a_{y1}}$	Skew Normal	$\theta_{a_{y1},t_r} = \left\langle \mu_{t_{r,a_{y1}}},\ \sigma_{t_{r,a_{y1}}},\ \alpha^{SN}_{t_{r,a_{y1}}} \right\rangle$
$t_{r,a_{y2}}$	Triangular	$\theta_{t_{r,a_{y2}}} = \left\langle c_{t_{r,a_{y2}}},\ lower_{t_{r,a_{y2}}},\ upper_{t_{r,a_{y2}}} \right\rangle$
$M_{p,LSOT}$	Student's t	$\theta_{M_{p,LSOT}} = \left\langle \mu_{M_{p,LSOT}},\ \sigma_{M_{p,LSOT}},\ \nu_{M_{p,LSOT}} \right\rangle$

5.4.2 TC2 - Hypothesis Testing as Parameter Estimation (P.E.)

5.4.2.1 TC2 - Statistical Model (P.E.)

The statistical modelling follows the same logic as in test case TC1 (Ch. 5.3.2.1). The prior distribution of the location parameter is assumed uniform ($\theta_{l,x} \sim U(0,1)$). The additional parameters have non-informative priors, and specifically uniform prior distributions, since no prior knowledge is available.

5.4.2.2 TC2 - Posterior Predictive Checks

The posterior predictive check plots are shown in Figure 5.10 and in Table 5.9, the Bayesian p-values are presented. The Pareto likelihood ($d_{M_{p,\dot{\psi}_1}}$, $d_{M_{p,\dot{\psi}_2}}$, $d_{M_{p,a_{y2}}}$) has Bayesian p-values around 0.8, indicating slight bias, where the predictions are more extreme than the observations. Nevertheless, the predictive check plots (Fig. 5.10a, 5.10b and 5.10d) illustrate that the predictions follow the curve of the observed data.

Similar remarks can be made for RWA_{a_y}, $t_{r,a_{y1}}$, $t_{r,a_{y2}}$ and $M_{p,LSOT}$. However, the p-values for these metrics are close to either 0 or 1, suggesting extreme bias. High Bayesian p-values mean that the predictions are more extreme than the observations. Figure 5.10e shows that the density of the relative difference of rearward amplification ($d_{RWA_{a_y}}$) is higher than the predictions, and the predictive KDEs cannot follow the observed data curve. Additionally, in Figure 5.10f the density of the predictive KDEs are quite close to the density of $d_{t_{r,a_{y1}}}$. However, they fail to capture the multiple peaks of the observed data.

The hypothesis testing on the location parameter can still be performed, even when the statistical modelling is biased. This will require though to take into account that the modelling predictions are more extreme than the observed data or vice versa.

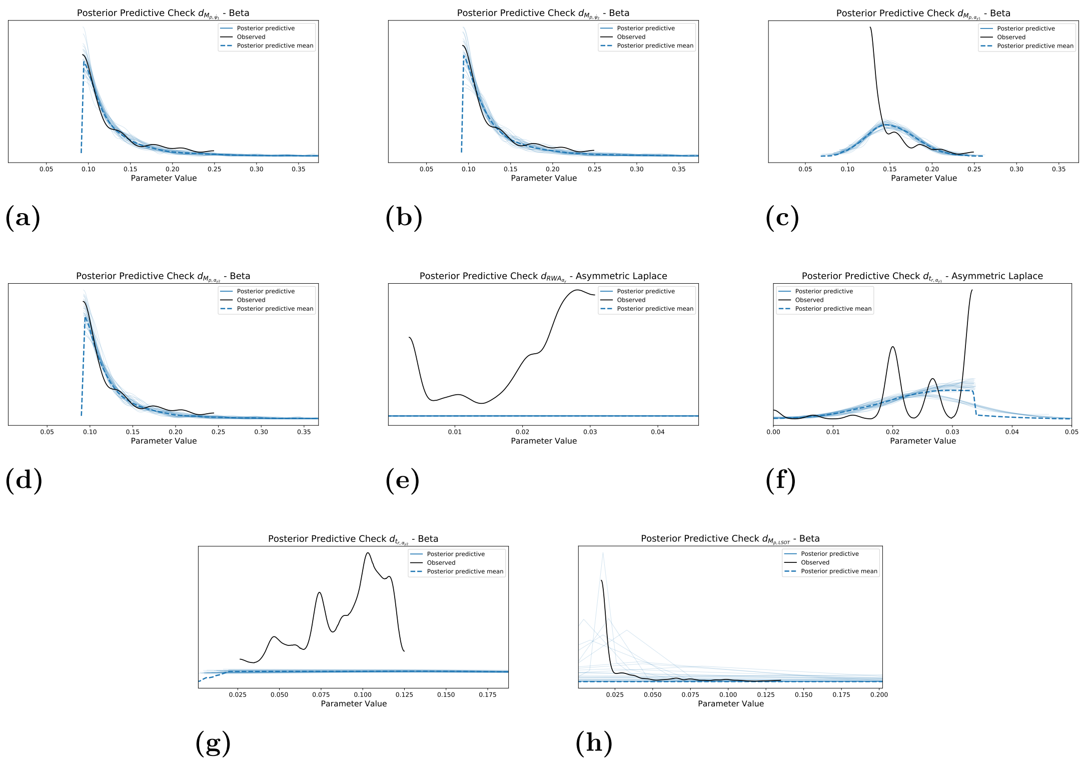


Figure 5.10: TC2 - Parameter Estimation - Posterior predictive check of relative difference d_x. 5.10a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, 5.10b Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, 5.10c Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, 5.10d Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, 5.10e Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, 5.10f Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$, 5.10g Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, 5.10h Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.

5.4.2.3 TC2 - Posterior Distributions and Decision (P.E.)

The evaluation of the validity of the abstract model is based on the relative position of the HDI and ROPE in the posterior distribution plot of the location parameter of each evaluated metric. The same maximum allowed limits are used as in test case TC1 (Table 5.4).

Table 5.9: Bayesian p-values (TC2).

Metric	$p-value$
$M_{p,\dot{\psi}_1}$	0.80
$M_{p,\dot{\psi}_2}$	0.80
$M_{p,a_{y1}}$	0.61
$M_{p,a_{y2}}$	0.84
RWA_{a_y}	0.12
$t_{r,a_{y1}}$	0.22
$t_{r,a_{y2}}$	1.00
$M_{p,LSOT}$	0.00

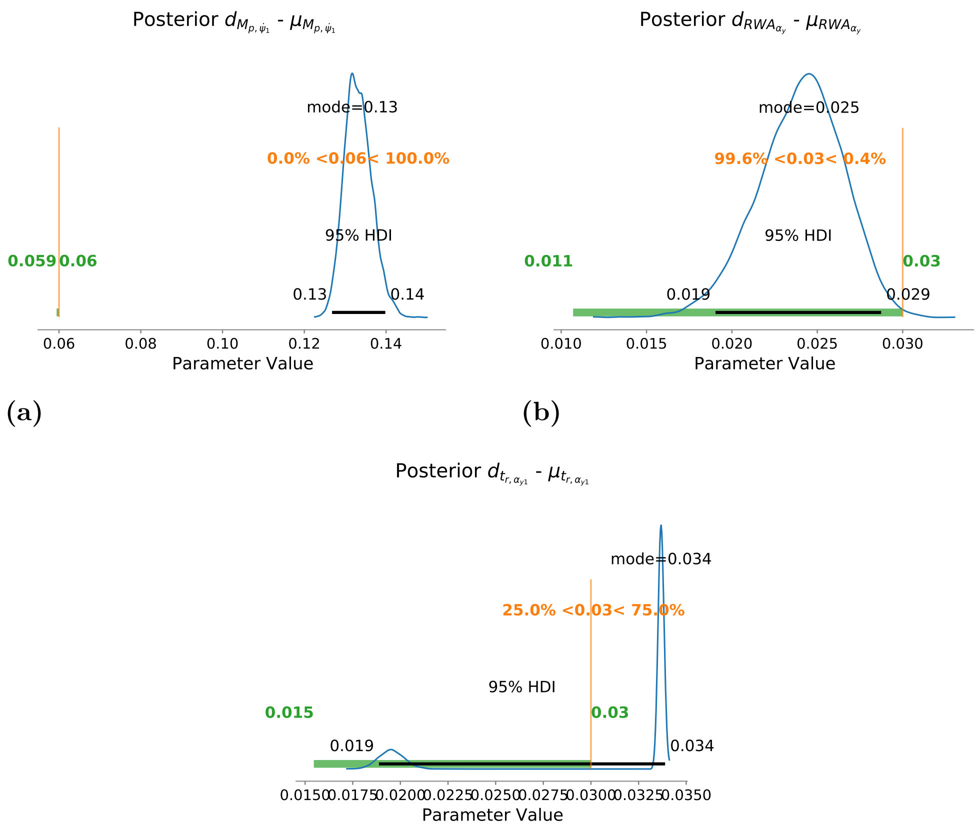


Figure 5.11: TC2 - Parameter Estimation - Posterior of location parameter ($\theta_{l,x}$) of relative difference d. 5.11a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, 5.11b Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, 5.11c Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$.

The abstract model fails to accurately calculate the peak yaw rate and lateral acceleration of tractor and semitrailer. Figure 5.11a shows that the 95% HDI is outside ROPE, where the 95% most credible values are quite greater than the maximum allowed limit (upper limit of ROPE interval).

On the other hand, the 95% most credible values of the relative difference of

$RWAa_y$, $t_{r,a_{y2}}$ and $M_{p,LSOT}$ are within the ROPE interval. Figure 5.11b depicts the posterior distribution of the location parameter of the rearward amplification Moreover, no decision can be made for the rise time of tractor's lateral acceleration ($t_{r,a_{y1}}$), since the 95% HDI partially overlaps with the defined ROPE (Figure 5.11c). It should be reminded that the statistical models of these four metrics are extremely biased, thus no decision should be based on them.

Note that the abstract model's overall validity is determined by the metrics that are important to the intended application. If the abstract model is invalid for at least one of them, then the model is rejected. Table 5.10 illustrates the final decision based on the posterior distributions and summary. Failure of computing the peak lateral acceleration and yaw rate deems the abstract model invalid, when the tractor's front axle cornering coefficient is equal to double the calibrated value.

Table 5.10: Decision of abstract model's acceptance for each evaluated model following the Bayesian Hypothesis testing formulated as Parameter Estimation (TC2).

Metric	**Decision**
$M_{p,\dot{\psi}_1}$	Invalid
$M_{p,\dot{\psi}_2}$	Invalid
$M_{p,a_{y1}}$	Invalid
$M_{p,a_{y2}}$	Invalid
RWA_{a_y}	Valid
$t_{r,a_{y1}}$	Inconclusive
$t_{r,a_{y2}}$	Valid
$M_{p,LSOT}$	Valid
Final Decision	Rejected

5.4.3 TC2 - Hypothesis Testing as Model Comparison of Statistical Models (M.D.)

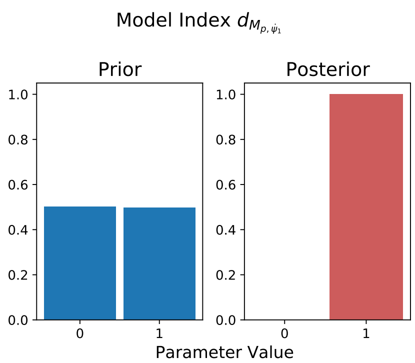
The statistical modelling of the hypothesis testing as model comparison is similar to the one described in TC1 (Ch. 5.3.3.1) and therefore is omitted. The next section presents the decision regarding the validity of the abstract model using the M.D. approach.

5.4.3.1 TC2 - Posterior Distributions and Decision (M.D.)

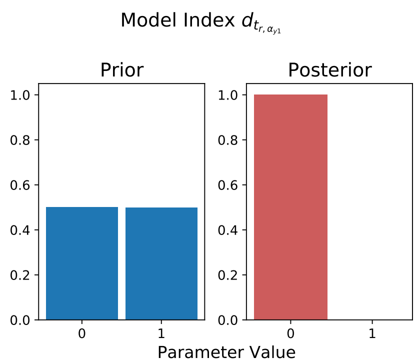
Table 5.11 shows the Bayes factor B_{01}, confidence κ and final decision concerning the abstract model's validity. The model comparison of statistical models approach also rejects the abstract model's computation of the peak lateral acceleration and yaw rate. For instance, Figure 5.12a clearly shows that the model index probability of accepting the null hypothesis is 0. On the other hand, the null hypothesis for the other metrics (RWA_{a_y}, $t_{r,a_{y1}}$, $t_{r,a_{y2}}$, $M_{p,LSOT}$) is accepted. Specifically for the relative difference of $t_{r,a_{y1}}$, the model comparison approach is accepting with 100% confidence the null hypothesis, whereas the parameter estimation approach was inconclusive (Fig. 5.12b). Finally, since the abstract model is invalid for at least one of the metrics, the vehicle model is rejected.

Table 5.11: Bayes factor B_{01}, confidence κ and final decision about abstract model's validity for each evaluated model, following the Bayesian Hypothesis testing formulated as Model Comparison (TC2).

Metric	B_{01}	Confidence κ	Decision
$M_{p,\dot{\psi}_1}$	0.00	0%	Invalid
$M_{p,\dot{\psi}_2}$	0.00	0%	Invalid
$M_{p,a_{y1}}$	0.00	0%	Invalid
$M_{p,a_{y2}}$	0.00	0%	Invalid
RWA_{a_y}	∞	100%	Valid
$t_{r,a_{y1}}$	∞	100%	Valid
$t_{r,a_{y2}}$	∞	100%	Valid
$M_{p,LSOT}$	∞	100%	Valid
Final Decision			Rejected



(a)



(b)

Figure 5.12: TC2 - Model Comparison of Statistical Models - Prior and Posterior of model index of relative difference d_x. 5.12a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, 5.12b Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$.

5.5 TC3 - Half Parameter Value

The final case studied concerns the tractor's front axle cornering coefficient being set to half the calibrated value. The same methodology as in the previous sections is followed. First the likelihood distributions describing the data are identified, followed by the implementation of the two Bayesian hypothesis testing approaches.

5.5.1 TC3 - Likelihood of Data

The likelihood distribution families for the relative difference d_x of each metric x are depicted in Table 5.12. Compared to test case TC1, the likelihood distributions are the same for most of the metrics. The beta distribution ($p(y|\theta) \sim Beta(\mu, \sigma)$) is used to describe the metrics $M_{p,\dot{\psi}_1}$, $M_{p,\dot{\psi}_2}$, $M_{p,a_{y1}}$, $M_{p,a_{y2}}$ and $M_{p,LSOT}$, and the asymmetric Laplace ($p(y|\theta) \sim AL(\mu, b, \kappa^{AL})$) to describe RWA_{a_y} and $t_{r,a_{y1}}$. However, the rise time of semitrailer's lateral acceleration is described by the skew Normal distribution ($p(y|\theta) \sim SN(\mu, \sigma, \alpha^{SN})$).

Table 5.12: Likelihood distribution family and parameters of the relative difference of the evaluated metrics (TC3).

Metric	Likelihood	Parameters
$M_{p,\dot{\psi}_1}$	Beta	$\theta_{M_{p,\dot{\psi}_1}} = \left\langle \mu_{M_{p,\dot{\psi}_1}},\ \sigma_{M_{p,\dot{\psi}_1}} \right\rangle$
$M_{p,\dot{\psi}_2}$	Beta	$\theta_{M_{p,\dot{\psi}_2}} = \left\langle \mu_{M_{p,\dot{\psi}_2}},\ \sigma_{M_{p,\dot{\psi}_2}} \right\rangle$
$M_{p,a_{y1}}$	Beta	$\theta_{M_{p,a_{y1}}} = \left\langle \mu_{M_{p,a_{y1}}},\ \sigma_{M_{p,a_{y1}}} \right\rangle$
$M_{p,a_{y2}}$	Beta	$\theta_{M_{p,a_{y2}}} = \left\langle \mu_{M_{p,a_{y2}}},\ \sigma_{M_{p,a_{y2}}} \right\rangle$
RWA_{a_y}	Asymmetric Laplace	$\theta_{RWA_{a_y}} = \left\langle \mu_{RWA_{a_y}},\ b_{RWA_{a_y}},\ \kappa^{AL}_{RWA_{a_y}} \right\rangle$
$t_{r,a_{y1}}$	Asymmetric Laplace	$\theta_{a_{y1,t_r}} = \left\langle \mu_{t_{r,a_{y1}}},\ b_{t_{r,a_{y1}}},\ \kappa^{AL}_{t_{r,a_{y1}}} \right\rangle$
$t_{r,a_{y2}}$	Skew Normal	$\theta_{t_{r,a_{y2}}} = \left\langle \mu_{t_{r,a_{y2}}},\ \sigma_{t_{r,a_{y2}}},\ \alpha^{SN}_{t_{r,a_{y2}}} \right\rangle$
$M_{p,LSOT}$	Beta	$\theta_{M_{p,LSOT}} = \left\langle \mu_{M_{p,LSOT}},\ \sigma_{M_{p,LSOT}} \right\rangle$

5.5.2 TC3 - Hypothesis Testing as Parameter Estimation (P.E.)

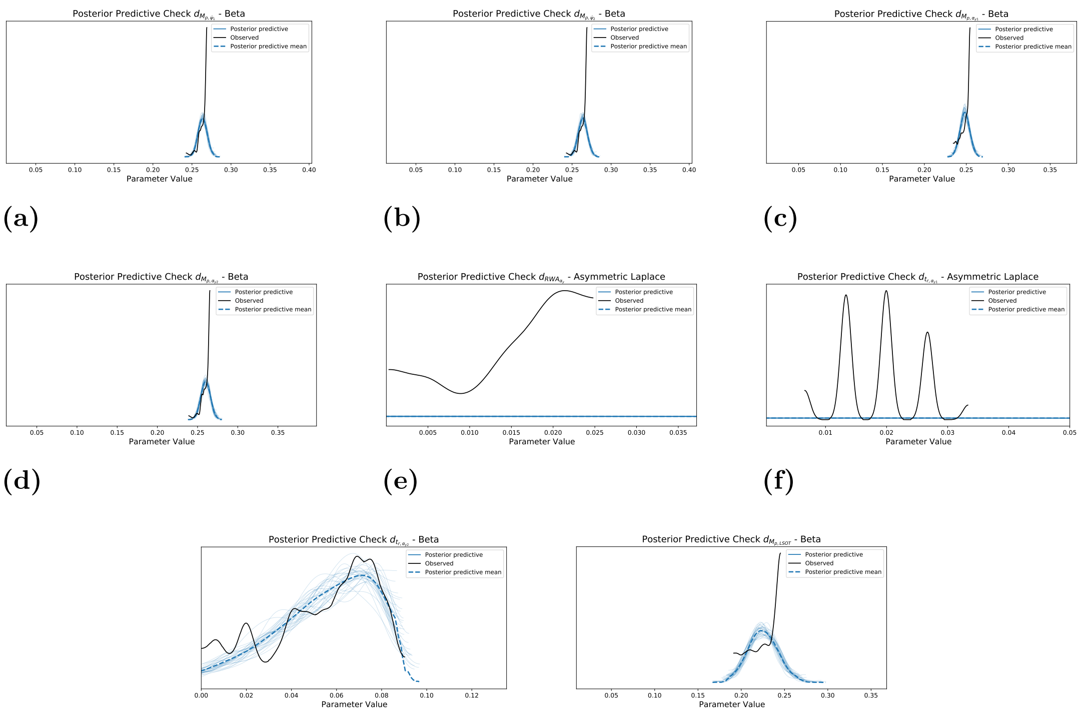
The statistical models are omitted, since they are structured in a similar manner to the ones presented in Chapter 5.3.2.1. The prior distribution of the location parameter is assumed uniform ($\theta_{l,x} = \mu_x \sim U(0, 1)$) and the additional parameters have non-informative priors (uniform distributions).

5.5.2.1 TC3 - Posterior Predictive Checks

Figure 5.13 illustrates the posterior predictive check plots for each statistical model and Table 5.13 shows the corresponding Bayesian p-values. In this case, the statistical modelling is a good fit for most of the metrics studied. The posterior predictive check plot and the Bayesian p-value agree that the statistical modelling is biased for the rearward amplification (RWA_{a_y}) and the rise time of the tractor's lateral acceleration $t_{r,a_{y1}}$. Therefore, a different distribution, other than the asymmetric Laplace, should be used to describe these relative differences.

Table 5.13: Bayesian p-values (TC3).

Metric	$p - value$
$M_{p,\dot{\psi}_1}$	0.49
$M_{p,\dot{\psi}_2}$	0.52
$M_{p,a_{y1}}$	0.52
$M_{p,a_{y2}}$	0.49
RWA_{a_y}	0.15
$t_{r,a_{y1}}$	0.31
$t_{r,a_{y2}}$	0.53
$M_{p,LSOT}$	0.48



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5.13: TC3 - Parameter Estimation - Posterior predictive check of relative difference d_x. 5.13a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, 5.13b Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, 5.13c Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, 5.13d Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, 5.13e Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, 5.13f Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$, 5.13g Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, 5.13h Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.

5.5.2.2 TC3 - Posterior Distributions and Decision (P.E.)

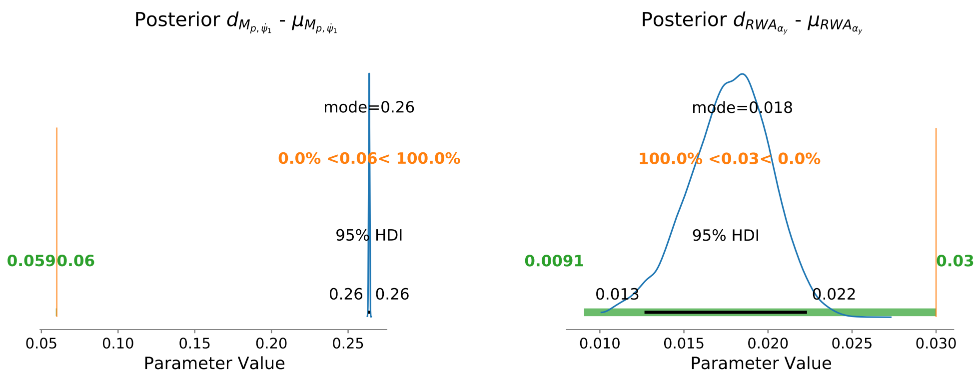
Similar to test case TC2, the maximum allowed limits given in Table 5.4 are considered. The final decisions regarding the abstract model's validity for the evaluated metrics are presented in Table 5.14. Figure 5.14 illustrates the posterior distributions of the location parameter of the relative difference of peak tractor yaw rate and rearward amplification ($d_{M_{p,\dot{\psi}_1}}$, $d_{RWA_{a_y}}$).

According to Table 5.4, the abstract model in this test case is invalid for almost all metrics apart from the rearward amplification and the rise time of tractor's lateral acceleration. Only for these metrics, the 95% most credible values are located with the ROPE. However, as discussed in Chapter 5.5.2.1, these two metrics have poor statistical modelling, and thus no decision should be based on them.

The decision about the overall validity of the abstract model depends on the application. For each application, some metrics are more important than others. The abstract model is considered invalid if it cannot accurately calculate at least one of them. Here, the abstract model fails for almost all metrics. Therefore, the vehicle model is rejected.

Table 5.14: Decision of abstract model's acceptance for each evaluated model following the Bayesian Hypothesis testing formulated as Parameter Estimation (TC3).

Metric	Decision
$M_{p,\dot{\psi}_1}$	Invalid
$M_{p,\dot{\psi}_2}$	Invalid
$M_{p,a_{y1}}$	Invalid
$M_{p,a_{y2}}$	Invalid
RWA_{a_y}	Valid
$t_{r,a_{y1}}$	Valid
$t_{r,a_{y2}}$	Invalid
$M_{p,LSOT}$	Invalid
Final Decision	Rejected



(a) (b)

Figure 5.14: TC3 - Parameter Estimation - Posterior of location parameter ($\theta_{l,x}$) of relative difference d_x. 5.14a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, 5.14b Rearward amplification $RWAa_y$ for $\epsilon = 3\%$.

5.5.3 TC3 - Hypothesis Testing as Model Comparison of Statistical Models (M.D.)

In this section only the results of the analysis and the decision of the abstract model's validity are shown since the statistical modelling is done in a similar way as before (Ch. 5.3.3.1).

5.5.3.1 TC3 - Posterior Distributions and Decision (M.D.)

Table 5.11 shows the Bayes factors B_{01}, confidence κ and final decisions regarding the validity of the abstract model. According to Table 5.11, the abstract model is invalid for all the metrics except for the rearward amplification and the rise time of the tractor's lateral acceleration. As mentioned before, the statistical models of these two metrics are biased, and thus they should be excluded from the decision-making process.

Table 5.15: Bayes factor B_{01}, confidence κ and final decision about abstract model's validity for each evaluated model, following the Bayesian Hypothesis testing formulated as Model Comparison (TC3).

Metric	B_{01}	**Confidence** κ	**Decision**
$M_{p,\dot{\psi}_1}$	0.0	0%	Invalid
$M_{p,\dot{\psi}_2}$	0.0	0%	Invalid
$M_{p,a_{y1}}$	0.0	0%	Invalid
$M_{p,a_{y2}}$	0.0	0%	Invalid
RWA_{a_y}	∞	100%	Valid
$t_{r,a_{y1}}$	∞	100%	Valid
$t_{r,a_{y2}}$	0.0	0%	Invalid
$M_{p,LSOT}$	0.0	0%	Invalid
Final Decision			Rejected

6

Conclusions and Future Work

This thesis proposes a methodology for vehicle model validation using Bayesian statistics, with a specific interest in Bayesian hypothesis testing. Two approaches of performing the hypothesis testing are implemented. The proposed framework is applied on a 90 ° turn maneuver on a tractor-semitrailer combination model as proof of concept. This chapter discusses the conclusions of this work and presents the future work.

6.1 Discussion & Conclusions

The purpose of this thesis was to structure a validation methodology, which compares the vehicle model (abstract) to an implementation model and determines if the former is applicable for a specific application. For that matter, the methodology was applied on a tractor-semitrailer combination vehicle model and different metrics were used to quantify its behavior. The test case selected to evaluate the abstract model was a left turn under city driving. Three variations of this test case were investigated.

The first case concerned the calibrated abstract model. By inspecting the vehicle responses of the abstract and implementation models for zero payload, it was expected that the former was valid for this particular case. The validation framework was able to confirm that and it deemed the abstract model valid for each evaluated metric. The other two cases concerned the incorrect tuning of the abstract model. The vehicle responses under these cases made apparent that the abstract model could not predict the vehicle behavior correctly. Using the proposed validation framework, the abstract model was invalidated for most of the metrics, which is in accordance with what expected.

The statistical modelling used in the Bayesian hypothesis testing is one of the most critical components of the analysis. The evaluation of these models was performed using posterior predictive checks. The combination of the posterior predictive check plots and the Bayesian p-values gives an insight as to whether the statistical model is adequate or not. In the three cases studied, some of the statistical models were biased. That means that the predicted values were more extreme than the data or vice versa. For example, in TC2, the rearward amplification observed values were greater than the predictions. That could lead to accepting the abstract model even though it is invalid. On the other hand, the peak tractor's yaw rate predictions were more extreme than the observations, which could cause invalidation of a valid vehicle model. To avoid that, different statistical models must be constructed. That could include using different priors on the statistical parameters or distributions for the likelihood. However, this is a time-consuming process, and thus the analysis was continued with the biased statistical models.

Both approaches in hypothesis testing have their strengths and weaknesses. The first approach, called the parameter estimation, calculates the posterior distributions of the parameters of the distribution that describes the data. The data here refer to the relative difference of the metrics. However, the difference between the models is a way of describing the uncertainty of the abstract model relative to the implementation model. Therefore, the parameter estimation approach models the uncertainty of the abstract model. This is quite useful and a lot of valuable information can be extracted. Additional, the posterior distributions can be used as a prior belief in another similar study and consequently update the belief regarding the abstract model. On the other hand, the parameter estimation approach cannot always reach a conclusion. As it was shown in the second test case (TC2), if the credible interval (HDI) is not entirely within or outside the region of practical equivalence (ROPE), no decision can be made.

The hypothesis testing as model comparison approach utilizes the Bayes factors. This approach provides a numerical value which quantifies the validity of the abstract model. Additionally, the Bayes factors can be evidence of not only accepting the null hypothesis but also of accepting the alternative hypothesis. Therefore, if a hypothesis testing includes more than one alternative hypotheses, this approach can determine which of them are more plausible. However, model comparison has also limitations. Bayes factors are sensitive to priors. In contrast with the parameter estimation approach which is relatively insensitive to different priors, in model comparison using different priors can lead to different Bayes factors.

6.2 Summary

Concluding, a methodology for vehicle model validation is proposed in this thesis to answer the high demand of valid vehicle models which are used for various automation applications. The validation framework presented is comprised of three main stages. The third stage is the validation process, which utilizes Bayesian hypothesis testing to compare the abstract to an implementation model. Two approaches of performing Bayesian hypothesis testing are introduced. The validation framework is implemented on a tractor-semitrailer combination model as proof of concept for a 90° turn test case, and three cases regarding the tuning of the abstract model are investigated. The performance of the framework proved to be quite good under all three cases. The methodology can validate the abstract model, when it is correctly calibrated, and produce posterior distributions of the statistical parameters that describe the vehicle performance differences. The two hypothesis testing methods agree with each other, and the results highlight their advantages and disadvantages.

6.3 Future Work

This thesis proposes a validation framework and sets its building blocks. The methodology can be applied for multiple applications. However, it is still at early stages and further development is required. This section includes recommendations to enhance the framework's performance and broaden its application, as well as to set the path for an automated validation process.

As far as the vehicle model is concerned, the calibration in this work was done manually. However, it is proposed that the calibration is done by using sensitivity analysis and then Bayesian calibration. The Bayesian calibration can be done by constructing an emulator, using Gaussian processes, which emulates the abstract model at arbitrary input and parameter settings. Then using Bayesian statistics the vehicle parameters can be calibrated to their real values.

The validation framework handles the comparison of two models, but requires parallel simulation. This means that the comparison of the abstract model to already collected data (e.g., real driving data) cannot be done. So far, the hypothesis testing is done on the mean of the difference of the two models. However, in case no parallel simulation can be done, the hypothesis testing should be applied on the difference of the mean of each model. In other words, the null hypothesis will be the equality of the abstract mean and the implementation mean.

The hypothesis testing formulation should also be revised. In this work, the hypothesis testing was applied on the location parameter of the likelihood. A better approach would be to use the expected value of the distribution. Moreover, a multivariate hypothesis testing is recommended to also include in the hypothesis the dispersion and the skewness.

Another development that is important, is the choice of the likelihood distribution to describe the data. It is obvious that the distributions used in this thesis, fail to fit the observed data well. For that matter, two improvements are proposed. First, the process of fitting a distribution to the data should be changed, in terms of the process itself and the used goodness-of-fit measure. In the Bayesian framework, the model comparison could be used as the fitting process, where each model refers to each available distribution. However, this might not be successful, since a single distribution might be unable to describe the data. Therefore, the second improvement would be to use combinations of distributions, called mixture models.

Finally, the framework should be tested on multiple test cases, including different maneuvers to generalize the results.

Bibliography

[1] P. Lima, "Optimization-based motion planning and model predictive control for autonomous driving: With experimental evaluation on a heavy-duty construction truck," Ph.D. dissertation, Sep. 2018.

[2] D. Asljung, J. Nilsson, and J. Fredriksson, "Using Extreme Value Theory for Vehicle Level Safety Validation and Implications for Autonomous Vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 4, pp. 288–297, 2017, ISSN: 23798858. DOI: 10.1109/TIV.2017.2768219.

[3] H. Sarin, M. Kokkolaras, G. Hulbert, P. Papalambros, S. Barbat, and R. J. Yang, "A comprehensive metric for comparing time histories in validation of simulation models with emphasis on vehicle safety applications," *Proceedings of the ASME Design Engineering Technical Conference*, vol. 1, no. PARTS A AND B, pp. 1275–1286, 2008. DOI: 10.1115/DETC2008-49669.

[4] O. Polach and A. Böttcher, "A new approach to define criteria for rail vehicle model validation," *Vehicle System Dynamics*, vol. 52, no. SUPPL. 1, pp. 125–141, 2014, ISSN: 17445159. DOI: 10.1080/00423114.2014.881515. [Online]. Available: https://doi.org/10.1080/00423114.2014.881515.

[5] V. Patil, "Generic and complete vehicle dynamic models for open-source platforms," *Department of Mechanics Sciences, Vehicle Engineering and Autonomous Systems, CHALMERS University of Technology, Gothenburg, Sweden*, 2017.

[6] M. Gäfvert and O. Lindgärde, "A 9-DOF tractor-semitrailer dynamic handling model for advanced chassis control studies," *Vehicle System Dynamics*, vol. 41, no. 1, pp. 51–82, 2004, ISSN: 00423114. DOI: 10.1076/vesd.41.1.51.23408.

[7] D. J. Cole and D. Cebon, "Validation of an Articulated Vehicle Simulation," *Vehicle System Dynamics*, vol. 21, no. 1, pp. 197–223, 1992, ISSN: 17445159. DOI: 10.1080/00423119208969009.

[8] X. Y. Lu and J. K. Hedrick, "Heavy-duty vehicle modelling and longitudinal control," *Vehicle System Dynamics*, vol. 43, no. 9, pp. 653–669, 2005, ISSN: 00423114. DOI: 10.1080/00423110412331282931.

[9] C. J. Hoel and P. Falcone, "Low speed maneuvering assistance for long vehicle combinations," *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 598–604, 2013. DOI: 10.1109/IVS.2013.6629532.

[10] C. Chen and M. Tomizuka, "Dynamic Modeling of Tractor-Semitrailer Vehicles in Automated Highway Systems," *Mechanical Engineering*, no. July, 1995.

[11] B. Li and S. Rakheja, "Yaw Stability Enhancement of Articulated Commercial Vehicles via Gain-Scheduling Optimal Control Approach," *SAE International Journal of Commercial Vehicles*, vol. 10, no. 1, pp. 275–282, 2017, ISSN: 19463928. DOI: 10.4271/2017-01-0437.

[12] J. Klemmer, J. Lauer, V. Formanski, R. Fontaine, P. Kilian, S. Sinsel, A. Erbes, and J. Zäpf, "Definition and Application of a Standard Verification and Validation Process for Dynamic Vehicle Simulation Models," *SAE International Journal of Materials and Manufacturing*, vol. 4, no. 1, pp. 743–758, 2011, ISSN: 19463979. DOI: 10.4271/2011-01-0519.

[13] E. Kutluay and H. Winner, "Validation of vehicle dynamics simulation models - A review," *Vehicle System Dynamics*, vol. 52, no. 2, pp. 186–200, 2014, ISSN: 00423114. DOI: 10.1080/00423114.2013.868500.

[14] K. Popper, "Logik der Forschung," *10th ed. Tübingen: Mohr Siebeck*, 2005.

[15] R. G. Sargent, "Verification and validation of simulation models," *Proceedings - Winter Simulation Conference*, no. January 2011, pp. 166–183, 2010, ISSN: 08917736. DOI: 10.1109/WSC.2010.5679166.

[16] M. Takkoush, "Formal Model Validation by Reachset Conformance between Low and High Order Tractor Semitrailer Vehicle Models," 2020.

[17] G. J. Heydinger, W. R. Garrott, J. P. Chrstos, and D. A. Guenther, "A methodology for validating vehicle dynamics simulations," *SAE Technical Papers*, 1990, ISSN: 26883627. DOI: 10.4271/900128.

[18] T. Buranathiti, J. Cao, W. Chen, L. Baghdasaryan, and Z. C. Xia, "Approaches for model validation: Methodology and illustration on a sheet metal flanging process," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 128, no. 2, pp. 588–597, 2006, ISSN: 10871357. DOI: 10.1115/1.1807852.

[19] H. Pan, M. Kokkolaras, G. Hulbert, M. Castanier, and D. Lamb, "2012 Ndia Ground Vehicle Systems Engineering and Technology Symposium Modeling & Simulation, Testing and Validation (Mstv) Mini-Symposium August 14-16, Michigan Model Validation for Simulations of Vehicle Systems," 2012.

[20] M. Carlsson, *Methods for Early Model Validation.* 2013, pp. 1–61, ISBN: 9789175196275. [Online]. Available: http://www.diva-portal.org/smash/get/diva2:616769/FULLTEXT02.pdf.

[21] F. Zeng, Q. Zong, X. Zhang, and M. You, "Sensitivity analysis based model validation for Hypersonic Vehicle model," *Proceedings of the 33rd Chinese Control Conference, CCC 2014*, pp. 6287–6292, 2014, ISSN: 21612927. DOI: 10.1109/ChiCC.2014.6896022.

[22] B. Danquah, S. Riedmaier, and M. Lienkamp, "Potential of statistical model verification, validation and uncertainty quantification in automotive vehicle dynamics simulations: a review," *Vehicle System Dynamics*, 2020, ISSN: 17445159. DOI: 10.1080/00423114.2020.1854317.

[23] Z. Zhan, J. Yang, Y. Fu, R. J. Yang, S. Barbat, and L. Zheng, "Research on Validation Metrics for Multiple Dynamic Response Comparison under Uncertainty," *SAE International Journal of Materials and Manufacturing*, vol. 8, no. 2, pp. 300–308, 2015, ISSN: 19463987. DOI: 10.4271/2015-01-0443.

[24] P. Honarmandi and R. Arróyave, *Uncertainty Quantification and Propagation in Computational Materials Science and Simulation-Assisted Materials Design*, 1. Springer International Publishing, 2020, vol. 9, pp. 103–143, ISBN: 0123456789. DOI: 10.1007/s40192-020-00168-2. [Online]. Available: https://doi.org/10.1007/s40192-020-00168-2.

[25] Y. Ling and S. Mahadevan, "Quantitative model validation techniques: New insights," *Reliability Engineering and System Safety*, vol. 111, pp. 217–231, 2013, ISSN: 09518320. DOI: 10.1016/j.ress.2012.11.011. arXiv: 1206.5014. [Online]. Available: http://dx.doi.org/10.1016/j.ress.2012.11.011.

[26] R. Rebba, S. Huang, Y. Liu, and S. Mahadevan, "Statistical validation of simulation models," *International Journal of Materials and Product Technology*, vol. 25, no. 1-3, pp. 164–181, 2006, ISSN: 02681900. DOI: 10.1504/ijmpt.2006.008280.

[27] X. Jiang and S. Mahadevan, "Bayesian inference method for model validation and confidence extrapolation," *Journal of Applied Statistics*, vol. 36, no. 6, pp. 659–677, 2009, ISSN: 02664763. DOI: 10.1080/02664760802499295.

[28] X. Jiang, R. J. Yang, S. Barbat, and P. Weerappuli, "Bayesian probabilistic PCA approach for model validation of dynamic systems," *SAE International Journal of Materials and Manufacturing*, vol. 2, no. 1, pp. 555–563, 2009, ISSN: 19463979. DOI: 10.4271/2009-01-1404.

[29] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006, ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207006000239.

[30] T. Archdeacon, *Correlation and regression analysis : a historian's guide*, en. University of Wisconsin Press, 1994, p. 164.

[31] S. Pandey and C. L. Bright, "What are degrees of freedom?" *Social Work Research*, vol. 32, no. 2, pp. 119–128, 2008, ISSN: 10705309, 15456838. [Online]. Available: http://www.jstor.org/stable/42659677.

[32] A. Gelman, *Bayesian data analysis Gelman*, 9. 2013, vol. 53, pp. 1689–1699, ISBN: 9788578110796. arXiv: arXiv:1011.1669v3.

[33] O. Martin, *Bayesian Analysis with Python: Second Edition*. 2018, p. 336, ISBN: 978-1-78588-380-4.

[34] D. Montgomery, *Applied statistics and probability for engineers*, en, 2013.

[35] J. K. Kruschke, *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, second edition*. 2014, pp. 1–759, ISBN: 9780124058880. DOI: 10.1016/B978-0-12-405888-0.09999-2.

[36] J. Kruschke and T. Liddell, "The Bayesian New Statistics (Public)," pp. 1–28, 2017. DOI: 10.31234/osf.io/wntsa.

[37] J. Salvatier, T. Wiecki, and C. Fonnesbeck, "Probabilistic programming in python using pymc3," *PeerJ Computer Science*, vol. 2:e55 DOI, 10 7717 – 55, 2016.

[38] M. Hoffman and A. Gelman, "The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo," *Journal of Machine Learning Research*, 2014.

[39] G. Privitera, *Statistics for the behavioral sciences*, Third. 2018, p. 816, ISBN: 9781506386249.

[40] S. Ross, *Introductory Statistics: Fourth Edition*, en. Elsevier, 2017.

[41] R. Rebba, "Model validation and design under uncertainty," dissertation, University of Vanderbilt, 2005. [Online]. Available: https://etd.library.vanderbilt.edu/etd-11222005-184433.

[42] S. Mahadevan and R. Rebba, "Validation of reliability computational models using bayes networks," *Reliability Engineering and System Safety*, vol. 87, no. 2, pp. 223–232, 2005, ISSN: 09518320. DOI: 10.1016/j.ress.2004.05.001.

[43] S. Shikano, "Hypothesis Testing in the Bayesian Framework," *Swiss Political Science Review*, vol. 25, no. 3, pp. 288–299, 2019, ISSN: 16626370. DOI: 10.1111/spsr.12375.

[44] H. Hoijtink and I. Klugkist, "Comparison of hypothesis testing and Bayesian model selection," *Quality and Quantity*, vol. 41, no. 1, pp. 73–91, 2007, ISSN: 00335177. DOI: 10.1007/s11135-005-6224-6.

[45] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, "Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2015-Octob, pp. 982–988, 2015. DOI: 10.1109/ITSC.2015.164.

[46] S. Geyer, M. Baltzer, B. Franz, S. Hakuli, M. Kauer, M. Kienle, S. Meier, T. Weigerber, K. Bengler, R. Bruder, F. Flemisch, and H. Winner, "Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance," *IET Intelligent Transport Systems*, vol. 8, no. 3, pp. 183–189, 2014, ISSN: 1751956X. DOI: 10.1049/iet-its.2012.0188.

[47] T. Menzel, G. Bagschik, Maurer, and Markus, "Scenarios for Development, Test and Validation of Automated Vehicles," *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2018-June, no. Iv, pp. 1821–1827, 2018. DOI: 10.1109/IVS.2018.8500406. arXiv: 1801.08598.

[48] H. B. Pacejka, "Chapter 1 - tire characteristics and vehicle handling and stability," in *Tire and Vehicle Dynamics (Third Edition)*, H. B. Pacejka, Ed., Third Edition, Oxford: Butterworth-Heinemann, 2012, pp. 1–58, ISBN: 978-0-08-097016-5. DOI: https://doi.org/10.1016/B978-0-08-097016-5.00001-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780080970165000012.

[49] G. Rill, *Road vehicle dynamics: fundamentals and modeling*. CRC Press, 2012, ISBN: 9781439897447. [Online]. Available: http://www.crcnetbase.com/isbn/9781439897447..

[50] J. P. Pauwelussen, *Essentials of vehicle dynamics*. 2014, pp. 1–309, ISBN: 9780081000588. DOI: 10.1016/C2014-0-00503-5.

[51] B. Jacobson, *Vehicle Dynamics Compendium*, en. Gothenburg: Chalmers University of Technology, 2018.

[52] A. Boström, *Rigid body dynamics*, pt, 2012.

[53] K. Ogata, *Modern control engineering*, it. Upper Saddle River, N.J: Prentice Hall, 1997.

[54] M. Kati, "Definitions of performance based characteristics for long heavy vehicle combinations," en, Technical report, 2013.

[55] J. Aurell and T. Wadman, "Vehicle combinations based on the modular concept"," en, Tech. Rep., 2007, Nordiska Vägtekniska Förbundet, report no 1/2007.

[56] D. Frisk, *A Chalmers University of Technology Master's thesis template for LaTeX. Unpublished*, 2018.

A

Distributions

A.1 Distribution Families

The set of distributions that are compared to the evaluated data are given below:

1. Uniform
2. Normal
3. Truncated Normal
4. Half Normal
5. Skewed Normal
6. Beta
7. Exponential
8. Laplace
9. Laplace Asymmetric
10. Student-t
11. Cauchy
12. Gamma
13. Inverse Gamma
14. Weibull
15. Log Normal
16. Chi-sqaured
17. Wald
18. Pareto
19. Exponential Normal
20. Von-Mises
21. Triangular
22. Gumbel
23. Rice
24. Logistic

A.2 Explanation of Continuous Probability Distributions Used in Test Cases

A.2.1 Asymmetric Laplace

The asymmetric Laplace likelihood is written as $p(y|\theta) \sim AL(\mu, b, \kappa)$. This distribution has three parameters. The first parameter (μ) is the location parameter, the second (b) is the scale parameter and the last one (κ) is the symmetry parameter. Both scale and symmetry parameters are positive real numbers (> 0). The distribution supports $x \in (-\infty, +\infty)$ [37].

The probability density function of the asymmetric Laplace distribution is [37]:

$$f\left(x \mid b, \kappa, \mu\right) = \left(\frac{b}{\kappa + 1/\kappa}\right) e^{-(x-\mu)bs\kappa^{s}}$$
$$s = sgn(x - \mu)$$

Mean of distribution [37]:

$$\text{mean} = \mu - \frac{\kappa - 1/\kappa}{b}$$

Variance of distribution [37]:

$$\text{variance} = \frac{1 + \kappa^4}{b^2 \kappa^2}$$

A.2.2 Beta

The Beta likelihood is written as $p(y|\theta) \sim Beta(\mu, \sigma)$, where μ is the mean and σ is the standard deviation. It supports only parameter values between 0 and 1 ($x \in (0, 1)$). The two parameters are bounded as follows [37]:

$$0 < \mu < 1$$
$$0 < \sigma < \sqrt{\mu\left(1 - \mu\right)}$$

The beta distribution can be parameterized in terms of alpha and beta, instead of the mean and standard deviation. The probability density function and the link between the parameters are the following:

$$f\left(x \mid \alpha, \beta\right) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$
$$\text{where } B(\alpha, \beta)\text{: the beta function (normalizing constant)}$$
$$\alpha = \mu\kappa$$
$$\beta = (1 - \mu)\kappa$$
$$\text{where } \kappa = \frac{\mu(1-\mu)}{\sigma^2} - 1$$

A.2.3 Pareto

The Pareto likelihood is written as $p(y|\theta) \sim Pareto(\alpha^P, m)$. This distribution is a function of two parameters; the shape α and the scale m parameter. Both parameters are positive real numbers (> 0). Pareto distribution supports $x \in [m, +\infty)$. The probability density function is the following [37]:

$$f\left(x \mid \alpha, m\right) = \frac{\alpha m^{\alpha}}{x^{\alpha+1}}$$

Mean of distribution [37]:

$$\text{mean} = \frac{\alpha m}{\alpha - 1} \ , \text{ for } \alpha \geq 1$$

Variance of distribution [37]:

$$\text{variance} = \frac{m\alpha}{(\alpha - 1)^2 (\alpha - 2)} \ , \text{ for } \alpha > 2$$

A.2.4 Skew Normal

The skew Normal likelihood is written as $p(y|\theta) \sim SN(\mu, \sigma, \alpha)$. The location parameter is μ, σ is the scale parameter and α is the skewness parameter. The location and skewness parameters are real, while the scale parameter is positive real (> 0). The distribution supports $x \in (-\infty, +\infty)$. When $\alpha = 0$, the skew Normal becomes a Normal distribution, where μ is the mean and σ is the standard deviation. When $\alpha \to \pm\infty$, the skew Normal gives a half Normal distribution When alpha=0 we recover the Normal distribution and mu becomes the mean, tau the precision and sigma the standard deviation. In the limit of alpha approaching plus/minus infinite we get a half-normal distribution. The probability distribution function is the following [37]:

$$f\left(x \mid \mu, \sigma, \alpha\right) = 2\Phi\left((x - \mu)\sqrt{\frac{1}{\sigma^2}}\alpha\right)\phi\left(x, \mu, \frac{1}{\sigma^2}\right)$$

where $\Phi(x)$: the cumulative distribution function and
$\phi(x)$: the standard Normal probability density function

Mean of distribution [37]:

$$\text{mean} = \mu + \sigma\sqrt{\frac{1}{\pi}}\frac{\alpha}{\sqrt{1 + \alpha^2}}$$

Variance of distribution [37]:

$$\text{variance} = \sigma^2\left(1 - \frac{2\alpha^2}{(\alpha^2 + 1)\pi}\right)$$

A.2.5 Student's t

The Student's t likelihood is written as $p(y|\theta) \sim t(\mu, \sigma, \nu)$. Here, μ is the location parameter, σ is the scale parameter and ν is the degrees of freedom, also known as normality parameter. The location parameter is real, while the scale and normality parameters are positive real numbers (> 0). As the degrees of freedom increase, σ converges to the standard deviation. Student's t distribution supports $x \in (-\infty, +\infty)$. The probability distribution function is the following [37]:

$$f(x \mid \mu, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{\pi\sigma^2\nu}\right)^{\frac{1}{2}} \left[1 + \frac{(x-\mu)^2}{\sigma^2\nu}\right]^{-\frac{\nu+1}{2}}$$
$$s = sgn(x - \mu)$$

A.2.6 Triangular

The triangular likelihood is written as $p(y|\theta) \sim Triang(c, lower, upper)$. Here, c is the mode, and the parameters *lower* and *upper* are the lower and upper limits of the variable x ($x \in [lower, upper]$), respectively. The probability distribution function of the Triangular distribution is the following:

$$f(x \mid lower, upper, c) = \begin{cases} 0\ , \text{ for } x < lower, \\ \frac{2(x-lower)}{(upper-lower)(c-lower)}\ , \text{ for } lower \leq x < c, \\ \frac{2}{upper-lower}\ , \text{ for } x = c, \\ \frac{2(upper-x)}{(upper-lower)(upper-c)}\ , \text{ for } c < x \leq lower, \\ 0\ , \text{ for } upper < x \end{cases}$$

Mean of distribution [37]:

$$\text{mean} = \frac{lower + upper + c}{3}$$

Variance of distribution [37]:

$$\text{variance} = \frac{uuper^2 + lower^2 + c^2 - lower \cdot upper - lower \cdot c - upper \cdot c}{18}$$

B

Additional Figures

B.1 Test Cases - Vehicle Responses Time Series

The following figures illustrate the vehicle responses under the three evaluated test cases (TC1, TC2 & TC3).

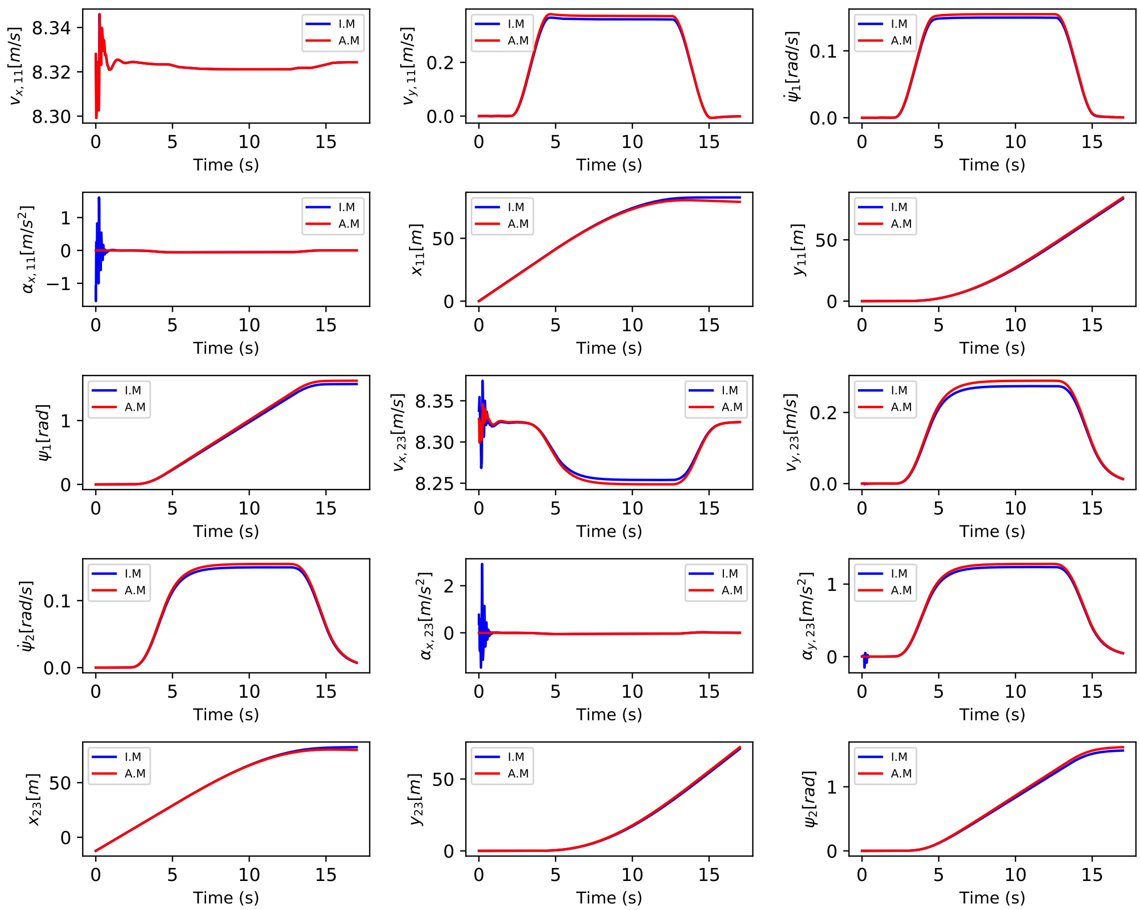


Figure B.1: TC1 - Calibrated parameter value. Vehicle responses for zero payload

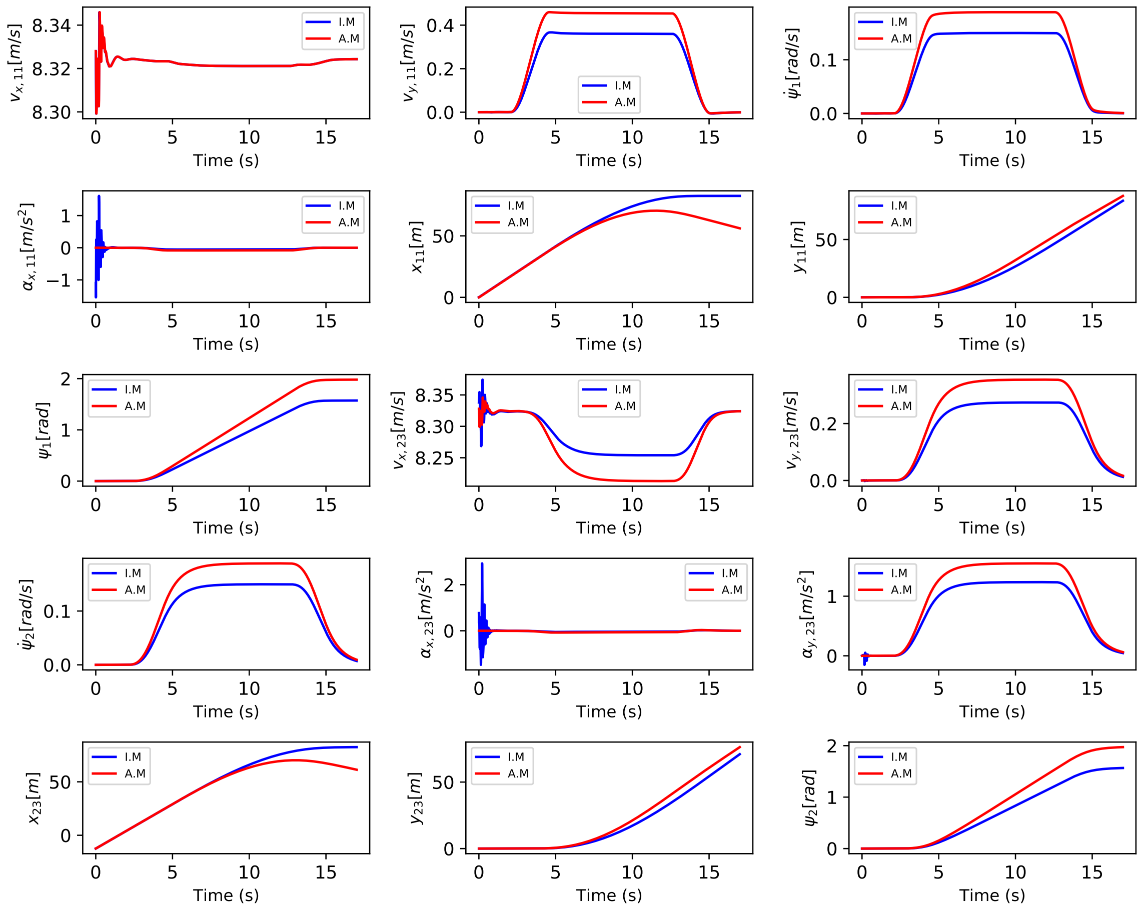


Figure B.2: TC2 - Double parameter value.

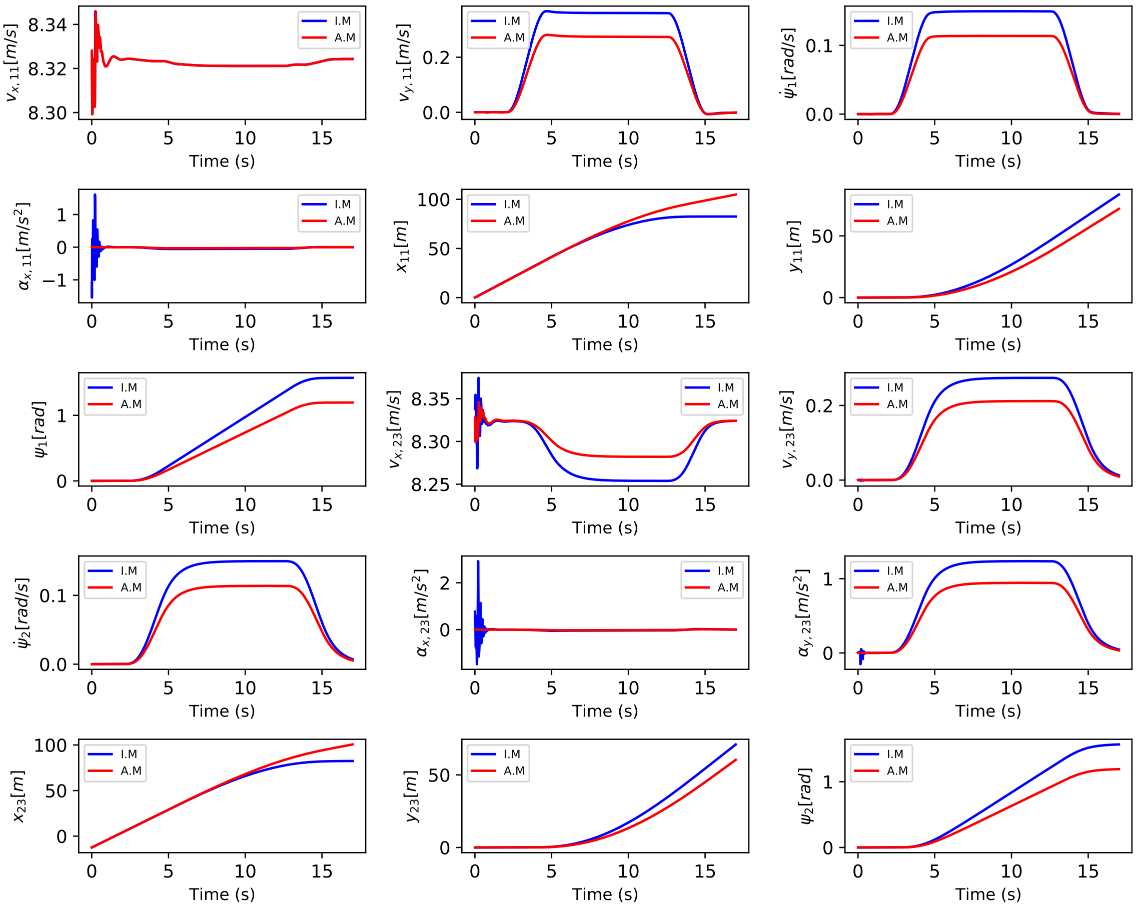
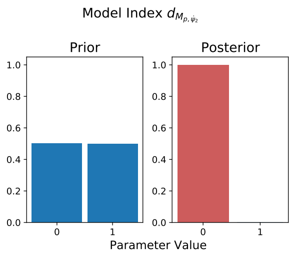


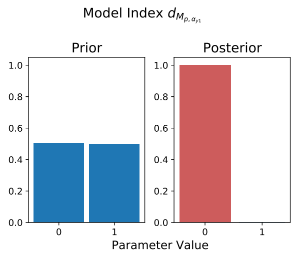
Figure B.3: TC3 - Half parameter value.

B.2 Test Case 1 - Calibrated Parameter Value

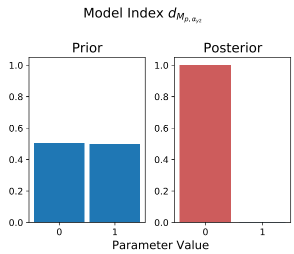
The following figure illustrate the prior and posterior distributions of the model index in the hypothesis testing as model comparison of statistical models approach, under the evaluated test case TC1.



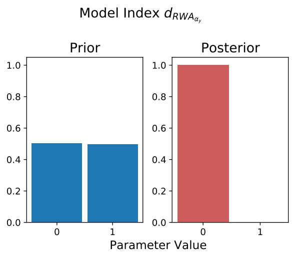
(a)



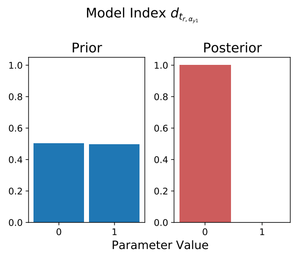
(b)



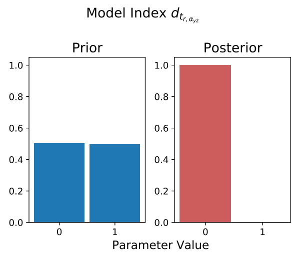
(c)



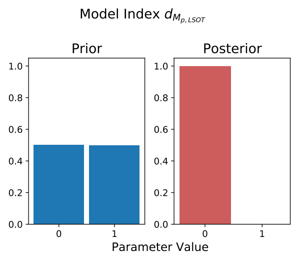
(d)



(e)



(f)

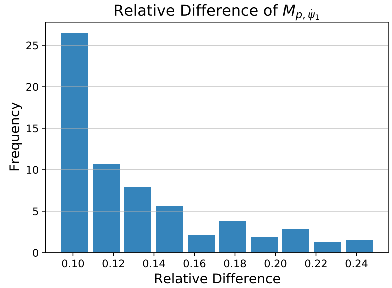


(g)

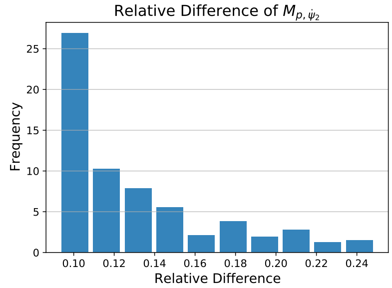
Figure B.4: TC1 - Model Comparison of Statistical Models - Prior and Posterior of model index of relative difference d_x. B.4a Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, B.4b Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, B.4c Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, B.4d Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, B.4e Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$, B.4f Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, B.4g Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.

B.3 Test Case 2 - Double Parameter Value

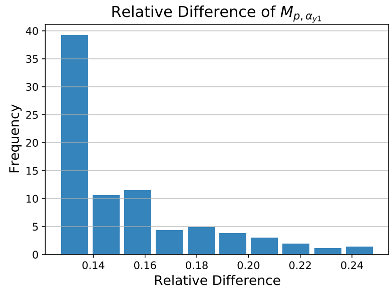
The following figures illustrate the under the evaluated test case TC2.



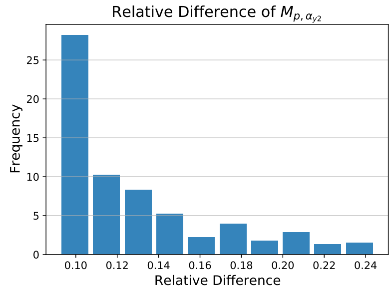
(a)



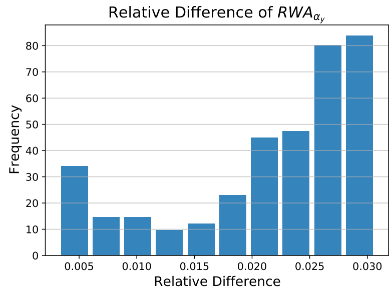
(b)



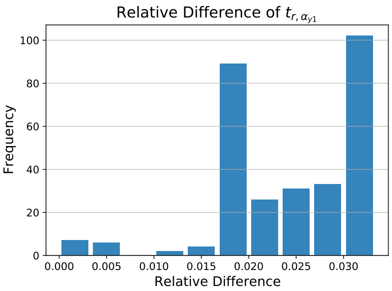
(c)



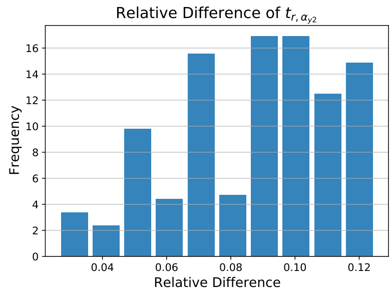
(d)



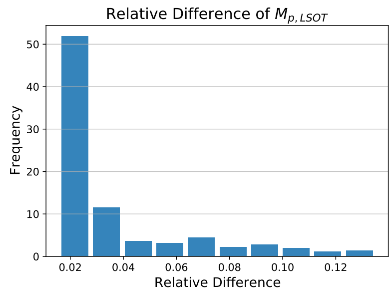
(e)



(f)

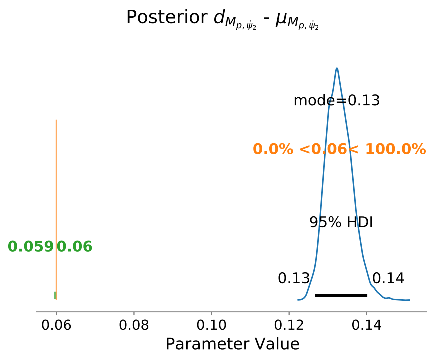


(g)

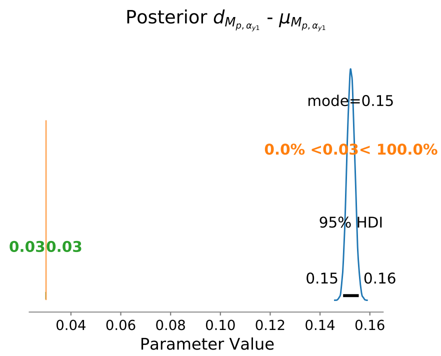


(h)

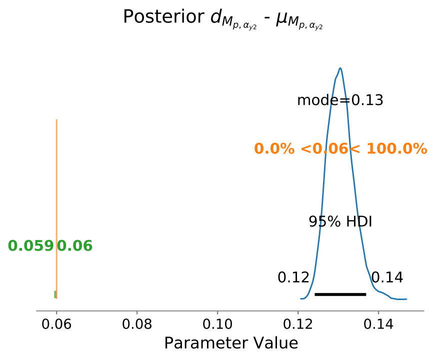
Figure B.5: TC1 - Histogram of relative difference d. B.5a Peak yaw rate of tractor $M_{p,\dot{\psi}_1}$, B.5b Peak yaw rate of semitrailer $M_{p,\dot{\psi}_2}$, B.5c Peak lateral acceleration of tractor $M_{p,a_{y1}}$, B.5d Peak lateral acceleration of semitrailer $M_{p,a_{y2}}$, B.5e Rearward amplification RWA_{a_y}, B.5f Rise time of lateral acceleration of tractor $t_{r,a_{y1}}$, B.5g Rise time of lateral acceleration of semitrailer $t_{r,a_{y2}}$, B.5h Peak low speed off-tracking $M_{p,LSOT}$



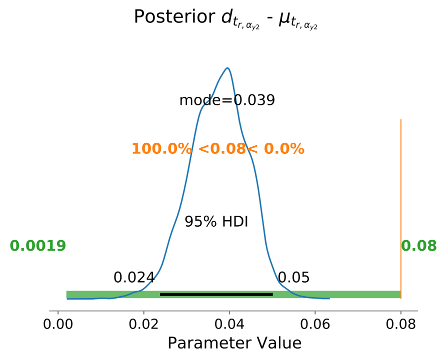
(a)



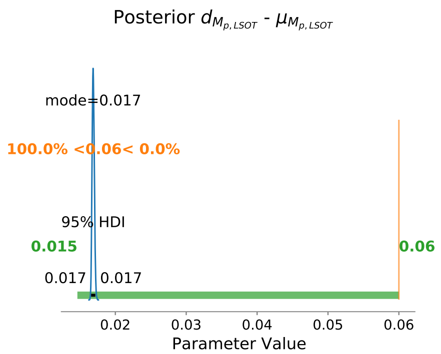
(b)



(c)

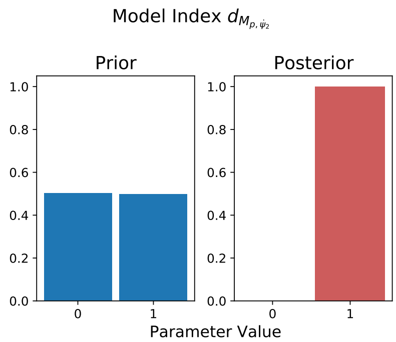


(d)

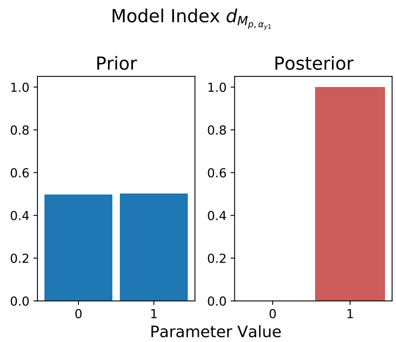


(e)

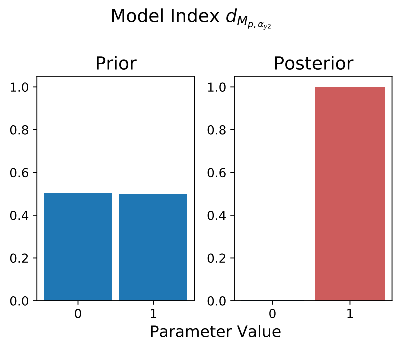
Figure B.6: TC2 - Parameter Estimation - Posterior of location parameter $(\theta_{l,x})$ of relative difference d. B.6a Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, B.6b Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, B.6c Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, B.6d Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, B.6e Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.



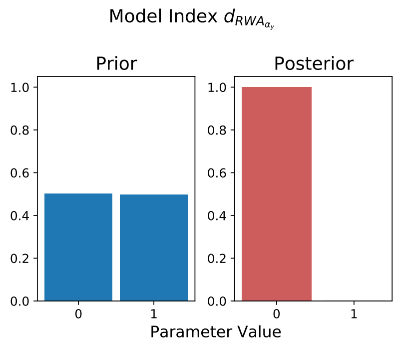
(a)



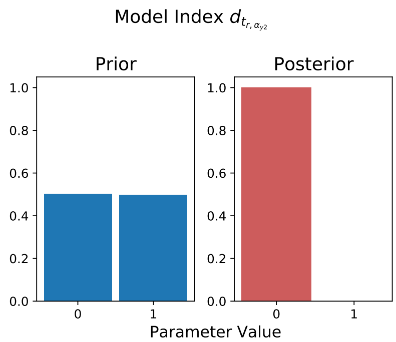
(b)



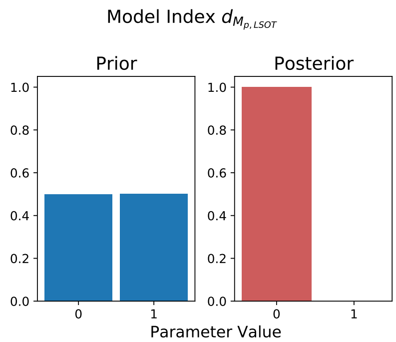
(c)



(d)



(e)

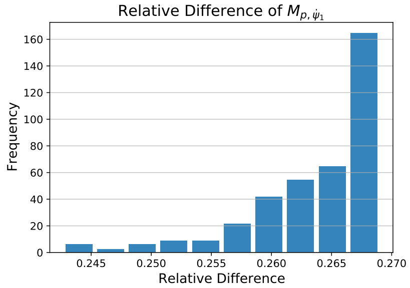


(f)

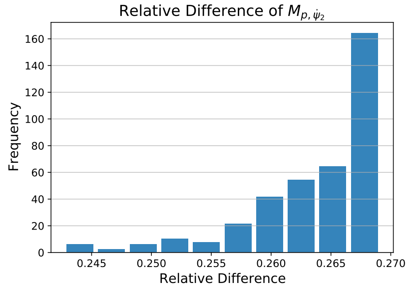
Figure B.7: TC2 - Model Comparison of Statistical Models - Prior and Posterior of model index of relative difference d_x. B.7a Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, B.7b Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, B.7c Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, B.7d Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, B.7e Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, B.7f Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.

B.4 Test Case 3 - Half Parameter Value

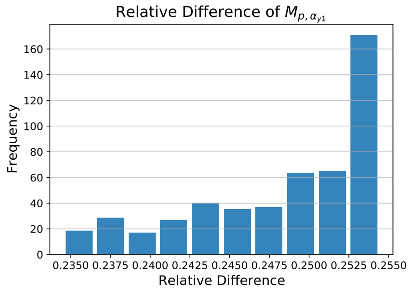
The following figures illustrate the under the evaluated test case TC3.



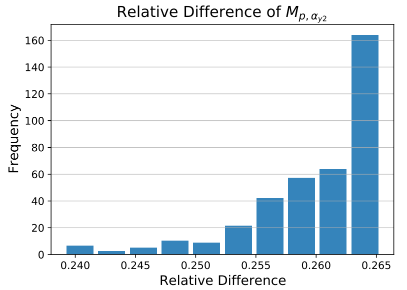
(a)



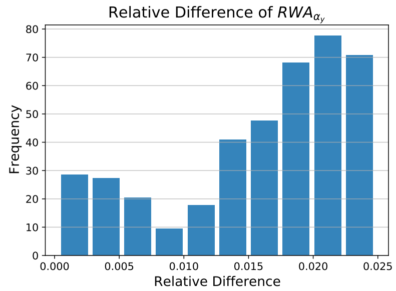
(b)



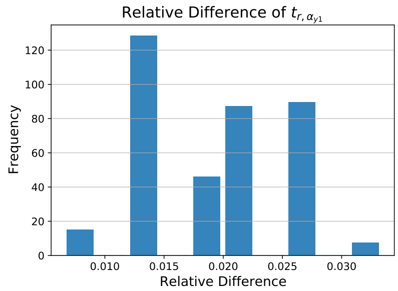
(c)



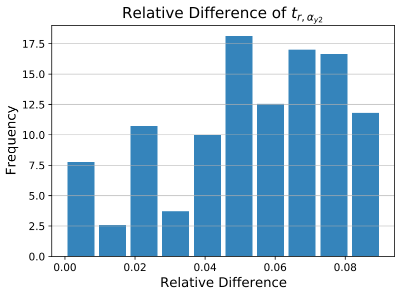
(d)



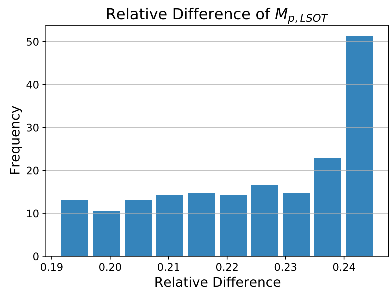
(e)



(f)



(g)



(h)

Figure B.8: TC1 - Histogram of relative difference d. B.8a Peak yaw rate of tractor $M_{p,\dot{\psi}_1}$, B.8b Peak yaw rate of semitrailer $M_{p,\dot{\psi}_2}$, B.8c Peak lateral acceleration of tractor $M_{p,a_{y1}}$, B.8d Peak lateral acceleration of semitrailer $M_{p,a_{y2}}$, B.8e Rearward amplification RWA_{a_y}, B.8f Rise time of lateral acceleration of tractor $t_{r,a_{y1}}$, B.8g Rise time of lateral acceleration of semitrailer $t_{r,a_{y2}}$, B.8h Peak low speed off-tracking $M_{p,LSOT}$

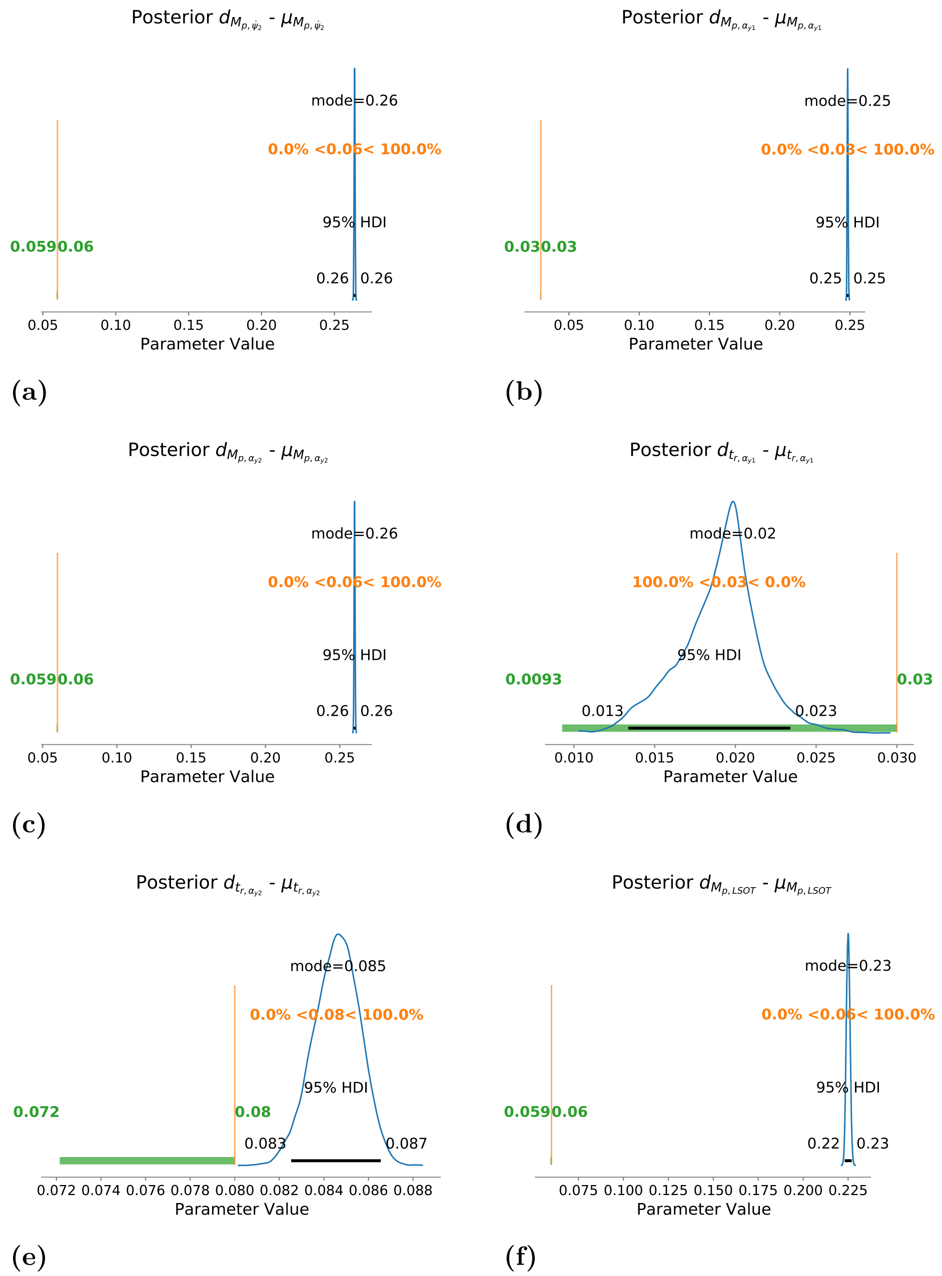


Figure B.9: TC3 - Parameter Estimation - Posterior of location parameter ($\theta_{l,x}$) of relative difference d_x. B.9a Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, B.9b Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, B.9c Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, B.9d Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$, B.9e Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, B.9f Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.

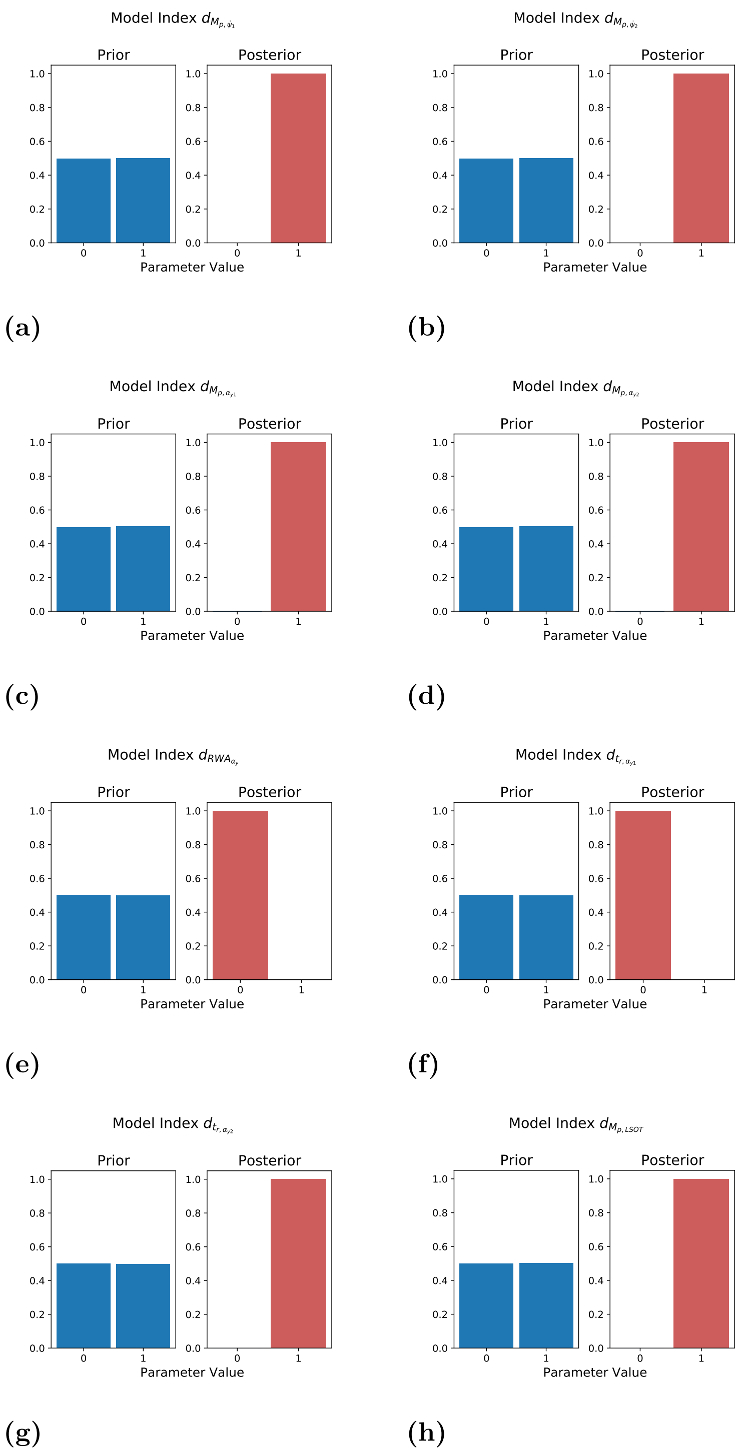


Figure B.10: TC3 - Model Comparison of Statistical Models - Prior and Posterior of model index of relative difference d_x. B.10a Peak tractor yaw rate $M_{p,\dot{\psi}_1}$ for $\epsilon = 6\%$, B.10b Peak semitrailer yaw rate $M_{p,\dot{\psi}_2}$ for $\epsilon = 6\%$, B.10c Peak tractor lateral acceleration $M_{p,a_{y1}}$ for $\epsilon = 3\%$, B.10d Peak semitrailer lateral acceleration $M_{p,a_{y2}}$ for $\epsilon = 6\%$, B.10e Rearward amplification $RWAa_y$ for $\epsilon = 3\%$, B.10f Rise time of tractor lateral acceleration $t_{r,a_{y1}}$ for $\epsilon = 3\%$, B.10g Rise time of semitrailer lateral acceleration $t_{r,a_{y2}}$ for $\epsilon = 8\%$, B.10h Peak low speed off-tracking $M_{p,LSOT}$ for $\epsilon = 6\%$.