UNIVERSITÀ DEGLI STUDI DI NAPOLI
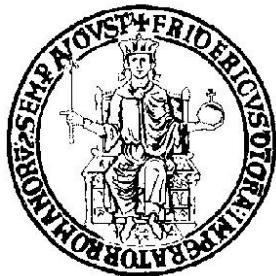# FEDERICO II

Dottorato di ricerca in Ingegneria dei Sistemi Idraulici, di Trasporto e Territoriali

Indirizzo "Infrastrutture Viarie e Sistemi di Trasporto"

XXVI° ciclo

Candidato

Marcello Montanino

UNCERTAINTY MANAGEMENT IN TRAFFIC SIMULATION: METHODOLOGY AND APPLICATIONS

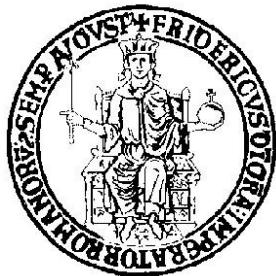**Coordinatore di dottorato:**

Prof. Arch. Elvira Petroncelli

Tutor:

Prof. Ing. Vincenzo Punzo

UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

Doctorate in Hydraulics, Transportation and Land-Use Systems Engineering

Specialization in "Transportation Systems and Infrastructures"

XXVI° period

PhD Candidate

Marcello Montanino

UNCERTAINTY MANAGEMENT IN TRAFFIC SIMULATION:
METHODOLOGY AND APPLICATIONS

School Coordinator:

Prof. Arch. Elvira Petroncelli

Supervisor:

Prof. Ing. Vincenzo Punzo

Uncertainty Management In Traffic Simulation: Methodology and Applications

Marcello Montanino

March 31, 2014

Preface

Scientific inquiry, creativity and problem solving attitudes have always been at the basis of what I thought (Da Vinci's epoch) engineers would have been. And, in this view, the three-years Ph.D. program gave me exactly what I was looking for. To free my mind from sector-based constraints and throw myself headlong in the development of new ideas. I have substantially expanded the range of my knowledge, in computer science, financial modeling, web development, and, of course, traffic flow theory.

Though not related to the field of my research, the (extra-time) work I spent on the development of "Pulse" and "Hangound", the direct involvement in two startups, the management of relationships with investors, and (above all) the great moments that I spent programming on my Dell and sharing thoughts and happiness with my "brother" Francesco, is priceless.

I worked hard, of course, but mostly because I was really enjoying it.

I would like to thank Prof. Vincenzo Punzo for giving me the chance of fulfilling all these complementary activities that have helped me to grow both culturally and scientifically.

However, at the beginning of everything, there was my research. And the generality of the methodologies that I deepened allowed me to think out-of-the-box. When I finished my Master in Hydraulics and Transportation Systems Engineering, almost four years ago, Vincenzo passed me the "Great Beauty" of microscopic traffic flow simulation modeling and, thereafter, it started an abundant endless flow of research, thoughts, tests, and writings without which this thesis simply would not have been possible. I will never be enough grateful to Vincenzo for having thought me how words should be combined

together to be effective, concise, attracting and pondered. Unfortunately, you never stop learning, and this thesis is the proof.

I would like to thank Prof. Ennio Cascetta for the constant feed of innovative and stimulating topics that helped me in extending my range of knowledge in transportation systems engineering. Since the very first day I met him in his office, I was enchanted by his ability to "see-the-future", to understand earlier which research direction would be relevant to sustain future mobility needs.

A heartfelt thank to Biagio Ciuffo for all the precious discussions, double-checks and advices on how to conduct my research and dissemination activities. I would never forget the walk-around in Haifa at the beginning of my Ph.D, where he convinced me about the significance of the research problems under investigation.

During my Ph.D., I had the great opportunity of being involved in the EU COST Action TU0903 (MULTITUDE), from which I benefited a lot in the development of the contents of this dissertation thesis. And I am grateful for all the nice talks in amazing locations that I shared with all the researchers there involved.

Many thanks to my roommate, Valerio, for the great moments we spent together, for the high-speed lunches, and for all "those talks".

Last, but certainly not the least, my family. To my Mamma and Papà: "without your love I would have never become the guy I am today".

Marcello Montanino

March, 2014

Contents

Chapter 1

Introduction

1.1 Background and Motivations

In the last years, the expanding range of new technologies in the field of traffic management and control called for accurate modeling of traffic flows in order to evaluate their potential impact on society and environmental decision-making.

The inner complexity of these applications sought for detailed stochastic traffic simulation tools which could enable their analysis, design and evaluation. In this view, microscopic traffic flow simulation models are increasingly used as cost-effective tools to support these tasks.

They are invaluable in offering a "common ground" for evaluating policies and examining, with a level of objectivity, the inevitable compromises required in practice. For instance, they are widely applied in the evaluation of new traffic control schemes (e.g. actuated/coordinated control, ramp metering, lane management, speed control, prioritarization, route clearance), to predict commuters' behavior in presence of advanced travelers information systems (e.g. variable message signs, route and parking guidance), or to assess travel demand management policy (e.g. congestion charge zones, eco-pricing, tolling system, mobility credits).

Furthermore, results of traffic simulation studies are frequently at the basis of investment decisions of private operators in transport-related markets and, thus, cover a crucial role in supporting the credibility of such proposals.

The enabling potential of micro-simulation software made it very popular among practitioners and triggered for advanced solutions to customize applications in an increasing range of contexts.

However, despite their importance, the use of these tools is far from being trivial. Indeed, the "goodness" of a simulation study does not depend only on the expertise of the analyst/modeler but (mostly) on the "correct" use of such models which, conversely, can be challenging even for specialists (Brackstone et al., 2012).

This could be due to a number of reasons.

Among the others, model "indeterminacy" or "equifinality" can be singled out (Young et al., 1996; Beven and Freer, 2001). Such (apparent) paradox refers to the case in which different models (based on different assumptions and resulting in different mathematical structures) provide results that are all compatible with the same set of traffic observations.

The impossibility to conceive "absolute" or "unique" models mainly stems from the way traffic flow models are derived, which is "quasi-" law-driven in the sense that physical principles (e.g. conservation equation) are mixed with reasonable assumptions, made by the modeler, on the way traffic entities may behave (e.g. safety-distance car-following model assumptions).

For such models, it has been suggested that they cannot be validated or verified, but only empirically confirmed by the non-contradiction between observation and prediction (Oreskes et al., 1994). Indeed, traffic flow models are formulated/calibrated against a limited variability of data, i.e. depicting only a small portion of the wide range of possible traffic phenomena. This often leads to models that are over-parameterized, as to depict also the non-observed traffic phenomena. However, this may be one of the causes for model unreliability, as suggested by the words of Hornberger and Spear (1981): "*[..] most simulation models will be complex, with many parameters, state-variables and non linear relations. Under the best circumstances, such models have many degrees of*

freedom and, with judicious fiddling, can be made to produce virtually any desired behaviour, often with both plausible structure and parameter values."

Apart from modeling uncertainties, unreliability may depends also on different sources of "uncertainty", such as the parametric and the non-parametric model inputs. Indeed, the quality in the characterization of the input uncertainties has the same impact on model results as the credibility of modeling assumptions.

In the field of traffic simulation, modeling input uncertainties is usually referred as model "calibration", and "*basically consists in identifying the parameters values that make the model outputs as close as possible to the reality*" (Punzo, 2014). The need for model calibration is implicit in the nature of traffic flow simulation models. In fact, such models are derived following an approach which is half way between a purely deductive and a purely inductive one (Papageorgiou, 1999), where "*[…] one first develops (via physical reasoning and/or adequate idealizations and/or physical analogies) a basic mathematical modeling structure and then one fits this specific structure (its parameters) to real data*".

Calibration is therefore viewed as a complementary step to model development, being expected to cover both the uncertainty in the modeling assumptions/formulations and the uncertainty in the inputs (Punzo, 2014).

Unfortunately, adequate paradigms for model calibration and validation against suitable observed data, are far from being established in the field of traffic simulation. Indeed, at present, no standardized methods exist, with most of the efforts and resources having been focused on model (and software) development. The importance of the subject is emphasized by the fact that a EU COST Action (MULTITUDE, 2014) was entirely dedicated to it. The Action proposal was driven by the concern that, although modeling is now widespread, we are unsure how much we can trust our results and conclusions. Such issues force into question the trustworthiness of the results, and indeed how well we are using them (MULTITUDE, 2014). Indeed, models encompassing a disproportionate amount of uncertainty turn out to have no practical utility for the transport analyst, as credibility of results is inevitably undermined.

Despite the importance of uncertainty management in scientific modeling, it is a very under investigated issue in the field of traffic flow simulation modeling. Indeed, there is no systematic approach in the literature encompassing the subject with an holistic

methodological framework. As a matter of fact, in the field research uncertainties were managed only indirectly, by (customarily) incorporating them within the parametric inputs and resolved by model parameter calibration.

Common symptoms of neglecting the management of uncertainty in traffic flow simulation modeling may be the (un)repeatability of experiments, the (un)reliability of predictions, and the vulnerability to instrumental or otherwise unethical use of models. Above all, this turns out in the lack of effectiveness, credibility, and transparency of simulation results (Punzo, 2014).

Therefore, the intent of this dissertation thesis is to provide a methodological contribution in the management of different sources of uncertainty in traffic flow simulation modeling. In the remaining of this Chapter we narrowed down the scope of the research presented (Section 1.2). Subsequently, we summarized the research approach (Section 1.3), and listed the main contributions to the state-of-the-art (Section 1.4). The final part of this introduction then briefly outlined which subjects are covered in each chapter of the thesis (Section 1.5).

1.2   Research Objective and Scope

The main objective of this dissertation thesis is to propose and apply a methodological framework for the management of uncertainty in traffic flow simulation modeling.

It is worth noting that, in this thesis, we focused on microscopic traffic flow simulation models only, and specifically on driver behavioral models (car-following and lane-changing models).

However, the methodological approach here proposed is absolute general and applications to other contexts (e.g. public transportation models, pedestrian simulation models, etc.), might be possible with reasonable easiness.

1.3   Research Approach

The methodological approach adopted in this thesis is based on an ensemble of techniques established in the industrial practice and increasingly applied in many modeling fields, including environmental, climate and financial ones, as well as, in system reliability and risk analysis. These techniques were recently combined in an holistic methodological framework for quantitative uncertainty assessment (de Rocquigny et al., 2008). A schematic representation of such framework is drawn in Figure 1.1.
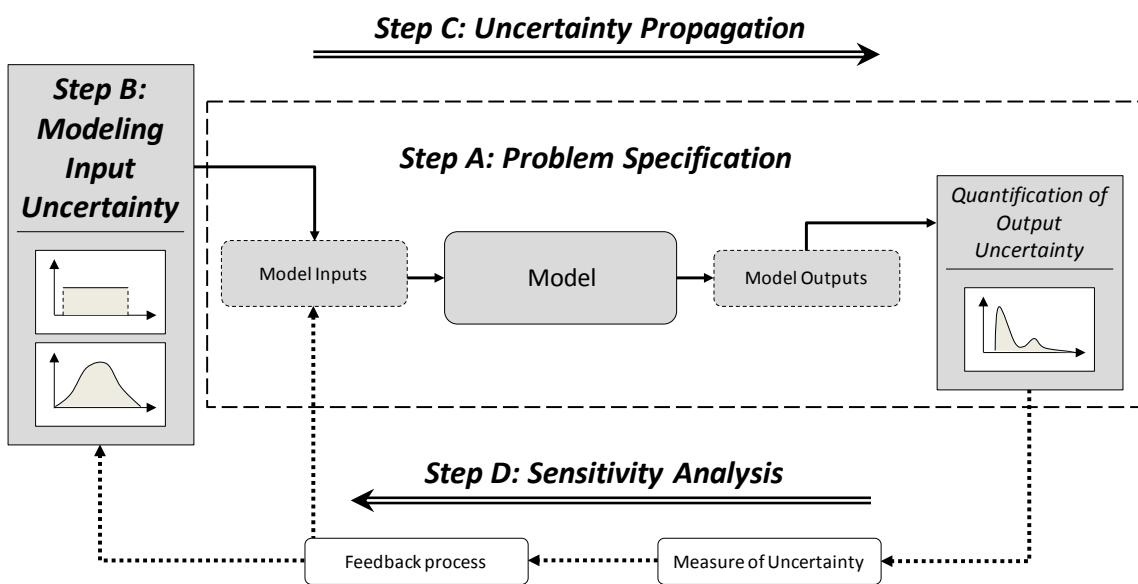
Figure 1.1: Common Conceptual Framework for Uncertainty Management.

In the following, a brief review of the four key steps is provided, indicating for each stage the methodological choices adopted in this thesis.

1.3.1. Problem Specification

This step consists in the specification of the model, the definition of the input and output variables, and the identification of the quantity of interest for measuring the uncertainty in the outputs.

In particular, the model may be viewed as a numerical function linking inputs (uncertain or fixed variables) to outputs (upon which decision criteria are established). Formally, it is sufficient for the model to link the output variables (y) to a number of continuous or

discrete inputs through an analytical function: $y = f(x, u)$, where some inputs (x) are *uncertain* – subject to randomness, lack of knowledge, errors or any other sources of uncertainty – while other inputs (u) are *fixed* – considered to be known.

Regarding the *uncertain* model inputs, also referred as *factors*, the vector x, made of k-components, could gather formally all sources of uncertainty, whatever their nature or type (parametric, model uncertainties, etc.). Some components of x may be continuous, while others could be discrete or branching variables. It could even formally include situations where there is a spatial field of uncertain inputs or even uncertain functions.

Some model inputs may be considered as *fixed* in a specific analysis framework. This is the case for a number of reasons, including the fact that *i)* some model inputs represent variables under full control; *ii)* uncertainties affecting some model inputs are considered to be negligible or of secondary importance with respect to the output variables of interest; and *iii)* for some model inputs, the decision process will conventionally fix the values despite uncertainties.

The methodological framework aims to quantify and (possibly) reduce output uncertainty, with regards to a well-defined *quantity of interest*. Possible measures of the output uncertainty can be the percentages of error or variability of model output(s) (e.g. variance, coefficient of variation); the expected value of model output(s); the confidence intervals of model output(s); quantiles of model output(s); the probabilities of exceeding a threshold or failure frequency; the ranges or simply the maximal value of model output(s).

On this basis, the *measure of uncertainty* is defined as the more complete mathematical distribution function comprehensively representing the output uncertainty (de Rocquigny et al., 2008), i.e. the probability density function of the model output.

As the objective of this research is not focusing on modeling itself, we concentrated on well-known driver behavioral models, widely adopted in the field literature and usually integrated in common micro-simulation packages, such as the Gipps' model (Gipps, 1981) and the Intelligent Driver Model (Treiber, 1999) for car-following, and the MOBIL model (Kesting, 2007) for lane-changing. Further, only driver behavioral model parameters where considered *uncertain*, being the other input sources (e.g. OD flows, path choice model parameters, etc.) considered as *fixed*. More details on this choice are provided in Chapter 2.

1.3.2. Uncertainty Modeling

This step consists in the quantification of the uncertainty sources in the inputs.

To this aim, several approaches may be possible, according to different uncertainty settings (de Rocquigny et al., 2008). However, in the common practice, it seems straightforward to consider input factors as stochastic variables[1]. On this basis, different probabilistic settings are proposed in the literature.

In some cases, when one or more components of the uncertain inputs are known to follow given distribution models, a standard probabilistic approach can be configured. In this setting, the uncertain input factors are distributed as random variables according to given probability density functions (*pdf*) , or more generally, as a random vector with a joint *pdf* (in case of correlation). The other model inputs are generally fixed at conventional deterministic values (within the vector u), so that this setting can in fact be considered a "mixed deterministic – probabilistic" setting. This means generally that no explicit separation has been made between natures of uncertainty; whatever their natures, all sources of uncertainty pertaining to vector x are randomized together.

In some other cases, when the distribution model itself is unknown (or its parameters), more complex settings can be adopted (2-level probabilistic settings). As in the previous setting, the uncertain model inputs are also considered as random variables (level-1 *pdfs*) to represent the uncertainty attached to them. But in addition, the parameters of their *pdfs* are considered sufficiently uncertain to be modeled within a probabilistic setting (level-2 *pdfs*) as well. Apart from the non-parametric methods, the stochastic variable is often modeled with a parametric approach. This means that the shape of the *pdf* is chosen among a list of existing distribution models (e.g. a Gaussian, Weibull or extreme value distribution, etc.), in which few parameters can be fine tuned to achieve a satisfactory description of the uncertainty sources. In a level-2 setting, an uncertainty model on distribution parameters also has to be determined, being it deterministic, probabilistic or non-probabilistic.

---

[1] From a theoretical point of view, a deterministic setting could also be identified. In this context, a set of values (otherwise the variable would be casted into a fixed one) must be chosen to define the range of variation for each of the uncertainty sources represented by the uncertain model inputs x. Although such approach could be used in screening uncertainty assessment, the methodological framework adopted in this work requires the use of probabilistic settings (and more detailed uncertainty models).

In this work, a standard probabilistic setting was adopted, which turned out into the estimation of the empirical joint probability density function (*pdf*) of driver behavioral model parameters.

1.3.3. Uncertainty Propagation

This step is necessary to map the uncertainty in the inputs into the uncertainty measures in the outputs. In a probabilistic setting, this implies estimating the *pdf* of $y = f(x, u)$, knowing the *pdf* of x, being given the values of u, and the numerical model $f(\cdot)$. According to the quantity of interest and the system model characteristics, the propagation may be a more or less difficult numerical step involving a wide variety of methods dependent on the adopted uncertainty setting. In a probabilistic setting, several different sampling schemes have been proposed in the literature, ranging from One-At-Time (OAT) random/stratified sampling to a wide ramification of factorial designs (full factorial design, two-level fractional factorial design, s-level fractional factorial design with h-strata Latin Hypercube scheme), multivariate stratified sampling and Monte Carlo Sampling. For a review of these approaches, please refer to Saltelli et al. (2008).

In this work, we relied on the Monte Carlo Sampling (MCS) framework, as it provides considerable benefits in terms of scalability. Indeed as many values as necessary can be generated, and if more parameters or more simulations are desired, it is a simple matter to generate more rather.

Further, MCS simulation can be also used for *meta-modeling*, that consists in building a mathematical function, which is cheaper from the point of view of computation time, and which approximates the behaviour of the model over the domain of variation of its inputs, starting from a set of selected model simulations in the uncertain input space, according to a specific sampling scheme (also called Design Of Experiment). Many meta-model families may be considered, such as: polynomials, generalized linear models (GLM), splines, interpolating radial functions, kriging, local polynomial kernel estimation, support vector machines, stochastic response surface methods using polynomial chaos expansions, partial least squares, neural networks, regression trees, etc. (McCullagh et al., 1989; Wahba, 1990; Ghanem and Spanos, 1991; Antoniadis et al., 1992; Fan and Gijbels, 1996; Vapnik, 1998; Chilès and Delfiner, 1999; Breiman, 2001; Hastie et al., 2001; Santner et al., 2003; Smola and Scholkopf, 2004; Fang et al., 2006).

For the sake of linearity, applications of meta-modeling techniques to microscopic traffic flow simulation modelling are not reported in this dissertation thesis. The interested reader could refer to Ciuffo et al. (2013).

In this thesis, a quasi-random Monte Carlo simulation framework based on the use of *low-discrepancy* sequences was adopted. Indeed, as shown in the literature, quasi-random low-discrepancy sequences overcome the limitations of traditional pseudo-random samples that tend to have clusters and gaps affecting the reliability of statistical analyses results (where a cluster occurs, function values in that vicinity are overemphasized in statistical analysis; conversely, where a gap arises, function values within that gap are not sampled for statistical analysis; for more details, the reader could refer to Saltelli et al., 2008).

In the literature two types of quasi-random low-discrepancy sequences are suggested: the Halton sequence (Halton, 1964; Niederreiter and Harald, 1992; Kuipers and Niederreiter, 2005) and the Sobol' sequences (Sobol, 1967; Sobol and Levitan, 1976), also called LPτ sequences or (t, s) sequences in base 2. See Bratley and Fox (1988) for a review.

In this work, we made use of the Sobol' sequences (Sobol et al., 1992), as fully integrated within the numerical computation scheme for variance-based global sensitivity analyses.

1.3.4. Sensitivity Analysis

It represents the feedback process of the uncertainty management cycle, and aims at understanding "*how uncertainties in the model outputs can be apportioned to different sources of uncertainties in the model inputs*" (Saltelli et al., 2004). In other words, "*the objective of the sensitivity analysis is to instruct the modeler with regards to the relative importance of the uncertain inputs in determining the variable of interest*" (Saltelli et al., 2008).

It generally involves some statistical treatment of the input/output relationship drawn within the Uncertainty Propagation step. Compact measures of the degree of importance of the uncertain factors x in affecting the model output variability are the so-called *sensitivity indices*. A wide variety of approaches can be adopted to compute these

measures, including: graphical methods (e.g. scatter plots, bars, tornado graphs, radar charts, box-and-whisker plots, cobweb plots; for details, see Wegman, 1990; Cooke and Noortwijk, in Saltelli et al., 2000; Saltelli et al., 2008), differential methods (e.g. sigma-normalized derivatives; for details, see Saltelli et al., 2008), screening techniques (e.g. Elementary Effects method; for a review, see Morris, 1991; Saltelli et al., 2008; for an enhanced formulation, see Campolongo et al., 2007, 2011), regression-based techniques (e.g. Pearson, Spearman, Standard Regression Coefficient, Partial Correlation Coefficient, Rank Correlation Coefficient, Standardized Rank Regression Coefficient, Partial Rank Correlation Coefficient; for a review, see McKay, 1997; Saltelli et al. 2008), non-parametric statistics (e.g. Smirnov test, Kruskal-Wallis test; for a review see Law and Kelton, 2000), variance-based decomposition (e.g. ANOVA, reviewed in Law and Kelton, 2000, and Box et al., 2005; Fourier Amplitude Sensitivity Test, reviewed in Cukier et al., 1978, Granger Morgan and Henrion, 1990, Saltelli and Bolado, 1998, Saltelli et al., 1999, Isukapalli, 1999, Frey and Patil, 2002; Correlation ratios, reviewed in de Rocquigny et al., 2008).

Among the others, techniques based on the *Sobol decomposition of variance* are widely accepted as the most versatile and effective among the various available techniques for sensitivity analysis of model output, tovercoming the limitations of the methods listed above (for a discussion, please refer to de Rocquigny et al., 2008). In particular, they allow for *i)* a *global* analysis, model-independent, not conditioned to any base-point in the input space; *ii)* the estimation of both first-order and *interaction* effects among model inputs, with relatively cheap computational cost and weak dependency on the number of model inputs k for the estimation of first order effects, while inevitably expensive and strictly k-dependent for total effect indices; and *iii)* the calculation of synthetic measures (the so-called *sensitivity indices*) to quantify the importance ranking of the uncertain model inputs.

For the above reasons, in this work we applied techniques based on the Sobol' decomposition of variance, in order to perform global sensitivity analyses of driver behavioral models.

1.4   Contributions to the State-of-the-Art

The main contributions to the State-of-the-Art offered in this dissertation thesis are listed below:

1. A methodological framework for the management of uncertainty in microscopic traffic flow simulation modeling (Chapter 2);

2. A robust methodology for "disaggregate" calibration of car-following and lane-changing models (Chapter 3);

3. A general procedure for handling measurement errors in vehicle trajectory data (Chapter 4);

4. A general methodology to simplify microscopic traffic flow simulation models based on global sensitivity analysis of model outputs (Chapter 5);

5. Investigation of the relation between "disaggregate" modeling and "aggregate" microscopic traffic flow simulation (Chapter 6).

1.5   Thesis Outline

A detailed overview of the structure of the main body of the thesis is here outlined.

Chapter 2 presents an introduction to uncertainty management in microscopic traffic flow simulation modeling. In particular, a review of the possible different sources of uncertainty in traffic simulation models is presented. Then, the methodological framework for managing parametric and non-parametric uncertainties is proposed. The Chapter ends with an outline of the contributions to the management cycle that are provided in the remaining chapters of the thesis.

Chapter 3 and 4 focus on the problem of indirect estimation of driver behavioral model parameters (Uncertainty Modeling), evaluating the impacts of the calibration procedure (Chapter 3) and of the measurement errors (Chapter 4) on the quality of calibration results.

Chapter 5 deals with the propagation of the input uncertainty into the modeling errors and present the results of a global sensitivity analysis of model outputs in order to understand the rank of importance of model parameters, with the aim to simplify models

without sensibly affecting their performances in the reproduction of observed traffic phenomena.

Chapter 6 presents the investigation of the relationship between "disaggregate" modeling and "aggregate" simulation. In particular, we evaluated the impact of measurement errors, model simplifications and parameters' correlation on "aggregate" microscopic traffic flow simulation performances.

Chapter 7 provides the main conclusions of this dissertation thesis and outlines directions for future research.

The mathematical review of the variance-based techniques for global sensitivity analysis of a simulation model is presented in Appendix A. Appendix B is dedicated to the analysis of the Gipps car-following model (Gipps, 1981) and to the presentation of an enhanced model formulation which generalizes the "free-flow" model equation. Appendix C deals with the problem of car-following model estimation against time-correlated measurements, proposing the adoption of a new class of goodness of fit functions derived in the frequency domain. Appendix D presents a general framework for the calibration of not-stochastic lane-changing models.

Chapter 2

Uncertainty in Traffic Flow Simulation Models

2.1 Introduction

Traffic is a stochastic highly dynamic non-linear phenomenon, resulting from the actions and interactions of large numbers of travelers, along with various exogenous events (Antoniou et al., 2014). Indeed, drivers perform several decision choices along their path from origin to destination. Initially, they select the departure time and choose among different route strategies to reach the destination. Then, en-route, they may adjust their path as they experiment congestion or delays over the network.

Moreover, drivers' choices may vary in time and across individuals. Different driving behaviors produce vehicles' interactions and are responsible for the decay of network performances (emergence of shockwaves, flow breakdown, capacity drop and so on).

As human beings are involved, traffic systems – as, more in general, transportation systems – are extremely complex, and with an intrinsic source of aleatory.

Traffic flow simulation models aims to reproduce the aleatory of traffic over road networks by stochastically modeling, (more or less) explicitly, both *strategic* (departing time, route choice) and *tactical* (actions aimed to directly control the vehicle in the traffic stream, subject to a number of environmental constraints, such as road rules, traffic lights, surrounding traffic, etc.) decision layers.

They are invaluable in offering a "common ground" for evaluating policies and examining, with a level of objectivity, the inevitable compromises required in practice. For instance, they are widely applied in the evaluation of new traffic control schemes (e.g. actuated/coordinated control, ramp metering, lane management, speed control, prioritarization, route clearance), to predict commuters' behavior in presence of advanced travelers information systems (e.g. variable message signs, route and parking guidance), or to assess travel demand management policy (e.g. congestion charge zones, eco-pricing, tolling system, mobility credits).

However, a large amount of uncertainty is encoded in such models and propagates in the simulation results. If not properly assessed and (possibly) reduced, models would encompass a disproportionate amount of uncertainty, and turn out to have no practical utility for the transport analyst, as credibility of results is inevitably undermined.

For example, ignoring the stochastic nature of a microscopic traffic flow simulation experiment, by performing a one-shot simulation, would not allow to capture the reliability of a management strategy (e.g. a ramp metering algorithm, a dynamic speed control policy, a congestion charge system), or, more in general, of a urban traffic plan, with regards to the variability of model inputs (e.g. demand levels, route choice model and driver behavioral model parameters). Further, in the assessment of different project/plan alternatives, an option that performed the best in a one-shot simulation experiment (deterministic approach), could be easily turn out to be the worst in a probabilistic assessment when propagating the variability (uncertainty) of model inputs into the simulation model. Therefore, as suggested in Antoniou et al. (2014), "*using such an instance for decision making could jeopardize the validity of the results and lead to bad planning*".

However, despite its relevance, uncertainty management is a strongly under investigated issue in traffic flow simulation modeling, and its implications on model applicability, robustness and credibility are frequently neglected by both practitioners and researchers.

As a matter of fact, the topic was traditionally relegated to marginal sections of transportation systems analysis books (Sinha and Labi; 2007; Cascetta, 2009; Willumsen and Ortuzar, 2011). Also in existing international guidelines (DMRB, 2013; HA, 2007; TfL, 2010; FHWA, 2004; VTRC, 2006; AUSTROADS, 2006; AUSTROADS, 2010;

TAC, 2008; JSTE, 2011), there is either a total lack of coverage of the subject, or, when definitions are proposed, they are largely misleading. For more details, see Section 5.2.

Further, no commercial simulation software allow for a quantitative assessment of uncertainty in simulation results and they are not design to let the user perform it. As a consequence, most of practitioners neglect this issue and are unaware of the terrible implications on simulation results for project assessment (Brackstone et al., 2012).

On the other hand, researchers have (more or less) deliberately paid less attention to the problem over the years, as more focused on modeling itself, that is to enhance models by changing their structure or adding more parameters as to reproduce specific observed or desirable phenomena/behaviors. However, the above research lines, in absence of corroborated procedures for model application in presence of uncertainty (including calibration and validation), would probably contribute to model indeterminacy (as clarified in Section 2.2).

Investigating uncertainty in traffic flow simulation models was the core objective of the EU COST Action (MULTITUDE, 2014). It was driven by the concern that, although modeling is now widespread, we are unsure how much we can trust our results and conclusions. Such issues force into question the trustworthiness of the results, and indeed how well we are using them (MULTITUDE, 2014).

For the above reason, the objective of this dissertation thesis is to provide a common methodological framework for the analysis of traffic flow simulation models in presence of uncertainty, that covers the steps of the quantitative uncertainty assessment cycle proposed in de Rocquigny et al. (2008).

It is worth noting that the accent will be on microscopic traffic flow simulation models only, and specifically on driver behavioral models (e.g. car-following and lane-changing models), although generalization to other traffic flow models, or even to more general transportation systems models (e.g. public transportation models, pedestrian simulation models), might be possible with reasonable easiness.

The remaining of this Chapter is organized as follows. Section 2.2 describes the concept of "uncertainty entailment" in scientific modeling. In Section 2.3, a literature review of the applications of uncertainty analysis in traffic flow simulation modeling is provided. In Section 2.4 we provided a discussion of the literature to highlight the motivation for

the work. In Section 2.5 the proposed framework for the analysis of microscopic traffic flow simulation models is described. Section 2.6 presents the outline of the study steps that will be carried out in the following chapters. A brief summary ends the Chapter.

2.2   Uncertainty Entailment in Scientific Modeling

A view of modeling that may help to illustrate the concept of "uncertainty entailment" is offered in Figure 2.1 (taken from the work of biologist Robert Rosen, 1991).
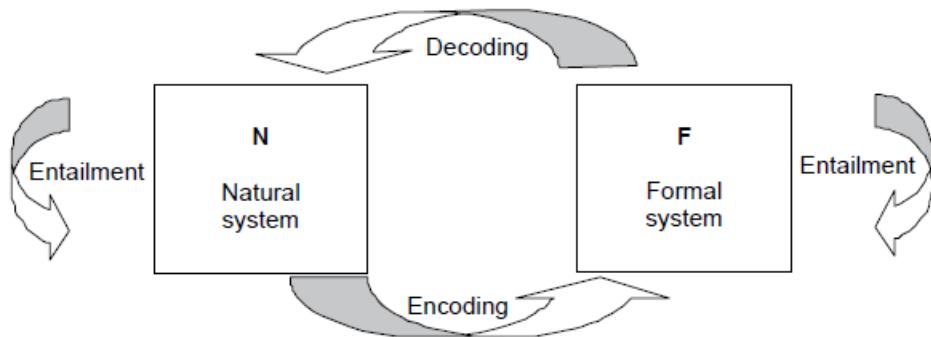
Figure 2.1: Modeling framework (Rosen, 1991).

In Rosen's diagram, the natural system, depicted on the left-hand side, obeys rules that we have the ambition to uncover. To this end we craft a set of structures in a model (depicted on the right-hand side of the figure). For example, a hypothesized set of rules for commuters' travel choices in a urban transportation system are casted into a random utility choice model. While people may keep choosing different travel options day by day, following the forces of its own systemic causality (which we aim to understand), our model can be solved using the rules of mathematical calculus. The intuition of Rosen, as brilliantly explained by Saltelli et al. (2008), was that "*while the world obeys rules, and the model has 'rules' as well, whether formal or mathematical, no 'rule' whatsoever can dictate how one should map the hypothesized rules in the world onto the rules in the model. In other words, while the world and the model are each internally 'entailed', nothing entails the world with the model.*"

In the case of traffic flow simulation models, the (desired) capability to describe the system evolution under unobserved circumstances, on one hand, and the (un)availability

of enough information to formulate "absolute" models, translates into the fact that "*different modelers can generate different nonequivalent descriptions of it, that is, models whose outputs are compatible with the same set of observations but whose structures are not reconcilable with one another*" (Saltelli et al., 2008).

In the literature, it has been coined a word for this paradox: equifinality – meaning that different models can lead to the same end (Beven and Binley, 1992; Beven, 1993, 2001). Others refer to the phenomenon as model indeterminacy (Oreskes et al., 1994).

As pointed out by Saltelli et al. (2008), along with the process of system encoding, a set of sources of uncertainty are intrinsically generated. Among the others, they include the classical categories of: "aleatory phenomena", "lack of data or knowledge", "variability", "measurement errors", etc. These may affect the model in various ways, through uncertain values for model inputs, model errors or even uncertain (or incomplete) structures of the model itself.

2.3   Applications in Traffic Flow Simulation Modeling

Despite of the underlined importance of uncertainty management in scientific modeling, in the field of traffic flow simulation modeling it is a very under investigated issue and, as a consequence, not properly handled by existing guidelines.

From the beginning, it is worth noting that the roles of uncertainty modeling, uncertainty quantification and sensitivity analysis are largely misinterpreted in many applications related to the traffic flow simulation literature.

A common source of misinterpretation is due to confounding *stability* analysis of model parameter estimates with *sensitivity* analysis of model outputs. This is the case, for example, of the application of generalized autoregressive conditional heteroscedasticity (GARCH) time series models for representing the dynamics of traffic flow volatility in Kamarianakis et al. (2005), where "sensitivity" of objective function in the neighbor of the optimal parameter estimates is investigated. Similarly, Ossen and Hoogendoorn (2009), in assessing the reliability of car-following model parameter estimates against vehicle trajectory data in different traffic conditions, provided measures of "sensitivity" of calibration objective function to small parameter changes. Likewise, Hoogendoorn and Hoogendoorn (2010a) performed the "sensitivity" analysis of car-following model

parameter estimations by local perturbation of the MLE log-likelihood at estimated base values. In such cases, stability analysis was performed in place of sensitivity analysis.

Confusion may also arise from the dualisms between *uncertainty* and *sensitivity* analysis, which are very frequently thought as synonymous. However, this is not the case. Such misunderstanding is frequent in the existing guidelines for traffic flow simulation modeling, such as the "Guidelines for Applying Traffic Microsimulation Modeling Software" of the California Department of Transportation (FHWA, 2004), the Austroads Research Report AP-R286/06 (AUSTROADS, 2006), and the Traffic Modelling Guidelines (TfL, 2010).

For instance, the FHWA guidelines (2004) state that *"... a sensitivity analysis is a targeted assessment of the reliability of the microsimulation results, given the uncertainty in the input or assumptions. The analyst identifies certain input or assumptions about which there is some uncertainty and varies them to see what their impact might be on the microsimulation results"*. However, such "sensitivity analysis" was actually a propagation of uncertainty aiming at uncertainty quantification. Indeed, what is mostly missing there is the loopback process associated to the sensitivity analysis which serves for the identification of how uncertainty in model outputs can be apportioned among the different sources of uncertainty in the inputs.

Also the majority of the studies reviewed in the literature fall into such misconception.

In many studies, uncertainty analysis aimed at the prediction, through (analytical or numerical) calculus of partial derivatives, of the changes in a system output to any *small* and *local* variation in the input base values. For example, in the field of traffic demand assignment, this approach was adopted for the analysis of the equilibrium network flows in Tobin and Friesz (1988) and further extended by Yan and Lam (1996) to predict changes in equilibrium link flow pattern, queuing delay, and system objective function in response to any small variation in toll charges. On the same research topic, the approach was implemented by Leurent (1998) to perform the analysis of the dual criteria traffic assignment model, and by Tam and Lam (1999). Lam and Zhou (2000) designed a solution algorithm, based on the derivative information, for some bi-level transportation optimization problems in which the traffic equilibrium problem was taken as the lower-level problem. Yang (1998) performed a sensitivity analysis of the queuing equilibrium network flow, to derive the explicit expressions of the derivatives of

equilibrium link flows and equilibrium queuing times with respect to traffic control parameters, in order to understand the stability of the equilibrium solutions to any *local* combination of changes of traffic control parameters. Later, the same approach was adopted by Chen et al. (2002) to estimate the sensitivity of the network equilibrium flows to the change of arc capacities. Same approach can be found also in Clark and Watling (2000), where an approximation to changes in the equilibrium solution of a probit-based SUE was deduced by the perturbation of its input parameters (specifically origin/destination flows and link cost-flow function parameters).

However, though a derivative-based approach has the attraction of being very efficient in computer time, it is unwarranted when the model input is uncertain and when the model is of unknown linearity. Indeed, this approach is "*illicit and unjustified unless the model under analysis is proved to be linear*" (Saltelli et al., 2006). In other words, derivatives are only informative at the base point where they are computed and do not provide an exploration of the rest of the space of the input factors. To overcome these limitations, sigma-normalized derivatives were introduced. For more details, please refer to Saltelli et al., 2008.

In the context of traffic simulation, sigma-normalized derivatives have been effectively used by Ji and Prevedouros (2005a), together with regression-based techniques and the Fourier Amplitude Sensitivity Test, with the purpose of benchmarking different techniques on the delay model proposed in the HCM (2000). On this topic, the same authors performed also similar studies in order *i)* to address the problem of uncertainty analysis in presence of correlation among model inputs (Ji and Prevedouros, 2005b), *ii)* to evaluate the impact of the a priori knowledge of model input probability distribution on results of uncertainty quantification (Ji and Prevedouros, 2006), and *iii)* to compare several other techniques (Ji and Prevedouros, 2007).

However, apart from these sporadic works, the most common setting used in the literature is the One-At-Time (OAT) uncertainty quantification. However, this approach has a major drawback: as inputs never vary simultaneously, the method completely hides the interaction effects of the parameters on the output, and thus provides unbiased results only for purely additive models.

In the field of microscopic traffic flow simulation, OAT applications can be found since 2000s.

For example, Bloomberg and Dale (2000) investigated the effect of travel demand variability on travel times by using the VISSIM and CORSIM micro-simulation software in a Monte Carlo-based simulation framework with a factorial design. Similar design for OAT uncertainty was applied by Lownes and Machemel (2006) and Mathew and Radhakrishnan (2010) on the VISSIM model, in order to select the model parameters in terms of their effects on some model outputs, such as simulated capacity.

Bartin et al. (2006) and Li and Zhang (2009) focused instead on the PARAMICS model, drawing inference on the variability of the average network travel times as to the variation, in a 2^k factorial design (Bartin et al., 2006) and in a 2^{k-p} fractional factorial design (Li and Zhang, 2009), of a subset of model parameters, selected based on modeler's experience. Similar approach was adopted in Bonsall et al. (2005) on the DRACULA microscopic traffic flow simulation model, where the sensitivity of model predictions, and perhaps policy decisions, to the value of some of the key parameters was studies.

Pel et al. (2010) performed an uncertainty quantification of the network accumulation from the macroscopic evacuation traffic simulation model EVAQ, in which input factors related to trip generation and departure rate, route choice, road capacities, and maximum speeds were systematically varied in a Monte Carlo-based simulation.

Focusing on car-following models, Kesting and Treiber (2008) applied a Monte Carlo approach to get additional insight on the sensitivity of calibrated parameter values for two car-following models. However, such approach has more affinities with the stability analysis described at the beginning of this section, rather than with a model sensitivity analysis.

An OAT approach was applied also in Patel et al. (2003), for the assessment of uncertainty in input data for the CAL3QHC roadway emission model, and select the most sensitive parameters with regards to the model simulated carbon monoxide concentrations. Attempts in this research direction can be recently found also in Song et al. (2012), where a local variation of eight parameter of traffic simulation model was performed as to determine their effects on simulated vehicle-specific power distributions.

Moreover, applications of OAT uncertainty analysis were not strictly limited to traffic flow simulation modeling, but can be found also in the wider context of transportation systems analysis.

In travel time prediction and reliability, the approach was adopted by Hellinga (2001) and Sun et al. (2007) to investigate the impact of several key parameters on loop/video-based and AVI-base identification algorithms for vehicle travel time estimation.

In the field of Intelligent Transportation Systems (ITS), such analysis was performed by Lawe et al. (2009) to assess the stability of the TRANSIMS model results (traffic volumes and average speeds) to changes in the random seed number and in the pre-timed signals of actuated controllers. Similarly, Peng and Beimborn (2001) and Sadek and Baah (2003) applied a OAT uncertainty propagation to assess the variability, though called sensitivity therein, of cost-effectiveness of an ITS deployment system with regards to the choice of different values for some of the model's parameters. Riemann et al. (2012) evaluated the variability in performances of certain cooperative systems within micro-simulation scenarios.

In pedestrian simulation and modeling, Wan and Rouphail (2004) performed an OAT analysis of pedestrian delay on a (customarily) selection of control parameters, including the vehicle demand, splitter island holding spaces, pedestrian crosswalk width, and pedestrian walking speed.

In transportation planning, Melkote and Daskin (2001) explored the tradeoff between investment and operating costs in network location problem for public and private facilities, though an OAT analysis of key model inputs. Rodier and Johnston (2002) analyzed plausible errors in projected trends for population, employment, fuel price, and income, as to the variability of the travel demand and emissions models. Schuster et al. (2005) evaluated the cost-effectiveness of a commuter-based car-sharing model, in Monte Carlo-based simulation where economic decision variables were varied one-at-time.

In transit planning, Fu and Liu (2003) applied sensitivity analysis to identify the conditions under which the proposed operating strategy for dynamic scheduling of transit operations were more advantageous. Chien et al. (2001) performed an OAT analysis for the estimation of cost-effectiveness of different bus systems, while Mesbah et al. (2010) applied it for the assessment of transit policy for lane priority.

As clarified earlier, applications based on an OAT variation of the uncertain input factors are likely to produce biased results. On the contrary, experimental designs that allow varying all the inputs simultaneously are needed to explore the whole input space and to account for the interactions of the various inputs (like e.g. demand, parameters and the network). Conversely, very computationally expensive design (multi-level factorial designs) are needed to allow for the estimation of second and higher-order interaction effects among model inputs on the variable of interest.

Traditionally, factorial designed ANOVA was adopted for the purpose. However, very few application of these method could be found in the field literature. A two-step analysis was adopted in Beegala et al. (2005) for the assessment of ramp control strategies in freeway management: a first screening OAT analysis was followed by an ANOVA with three level factorial design with the purpose of model parameters prioritarization. In Ciuffo et al. (2007) and in Punzo and Ciuffo (2009) second order interactions effects among model parameters were investigated for the AIMSUN model.

2.4   Towards a New Approach

What emerges from the review of the literature given in Section 2.3 is that uncertainty quantification and sensitivity analysis have covered only marginal roles in traffic flow simulation studies, and, more in general, in transportation systems analyses. Indeed, the main focus of the analyses was on model description and application, while a (frequently misleading) sensitivity analysis was performed, at last, to show how *small* and *local* variation in model inputs affected the reliability of results.

Such traditional approaches tangle with the need for model credibility and robustness. Indeed, as brilliantly explained by Ni et al. (2004), "*to successfully apply a simulation model, the correctness or credibility of the model is crucial, and some testing processes have to be resorted to in order to ensure the quality of the model through model validation, a critical testing process that compares the model output with real-world system behaviour*".

However, quality and robustness of results mainly depend on two components: *i)* the credibility of the model, and *ii)* its correct use.

With regards to the first point, Papageorgiou (1998) observed that available traffic flow models are derived following an approach which is half way between a purely deductive approach and a purely inductive one. Such an approach neither consists in deriving models from invariable basic principles – like in Newtonian physics and related mechanical models (purely deductive) – nor in fitting a generic mathematical structure to observed data (purely inductive). Rather, "*one first develops (via physical reasoning and/or adequate idealizations and/or physical analogies) a basic mathematical modeling structure and then one fits this specific structure (its parameters) to real data*" (Papageorgiou, 1998). In fact, at the basis of traffic flow modeling, there could be physical principles (e.g. conservation equation), common-sense assumptions (e.g. safety-distance car-following model), or (mostly) a mix of both types. However, for the purpose of model applicability, they are generally limited only to those system behaviors that we attempt to reproduce or, in other cases, even self-tailored on the application case-study. Further, in some cases, common-sense assumptions may be contested, as not general enough to depict the range of phenomena which we attempt to reproduce. For example, at the basis of most popular car-following models, such as the Gipps' model (1981), the OVM model (Bando et al., 1995), the Intelligent Driver Model (Treiber et al., 2000), there are hypothesis regarding the dependence of driving response from the leader stimulus in the current lane, but there could be situations in which the response may either depend also on other surrounding vehicles (e.g. the anticipation and relaxation behaviors of the vehicle in response to an imminent lane-changing; Ahn et al., 2013), or depend on exogenous factors not modeled at all (e.g. rubbernecking behavior at work zones).

As a consequence, different (type of) assumptions, often equally reasonable, have led to alternative model formulations which compete concerning their ability of describing and interpreting traffic phenomena. Model *indeterminacy* or *equifinality* are two terms applied to define such a condition – common to many disciplines – that results in having more than one model compatible with the same set of data or evidence. The implicit understanding behind such concepts is that the models being used are approximations of reality and that their outputs are necessarily uncertain, not only because of the lack of knowledge in the inputs, but also due the modeling process itself (Antoniou et al., 2014).

The above considerations substantiated the second point related to the correct use of such models to avoid (potential) vulnerability to instrumental or otherwise unethical scopes.

Indeed, traffic flow simulation models are often over-parameterized. According to Hornberger and Spear (1981), "*most simulation models will be complex, with many parameters, state-variables and non-linear relations. Under the best circumstances, such models have many degrees of freedom and, with judicious fiddling, can be made to produce virtually any desired behaviour, often with both plausible structure and parameter values.*"

This poses serious questions on the reliability of the results of a simulation study as well as on the transparency of the study itself (i.e. (un)-repeatability of experiments). In fact, it is reasonable to claim that results of a study are mostly driven by the way in which model inputs (parameters and OD flows) are estimated. However, their estimation is a complex non-linear problem, with a very large number of unknowns, and it is hard to find a solution that is reliable and robust.

As Ge and Menendez pointed out (2012), "*commercial traffic simulators usually contain a huge number of parameters to cover various kinds of simulators (e.g. vehicles, public transport, pedestrians). As an example, VISSIM has 192 parameters …, and this figure will most likely continue to grow with new updates*". This raised questions on the practical affordability of software calibration, which, de facto, is limited only to a (customarily) selected subset of parameters. Also, it generally happens that just a subset of the input parameters drives the overall variability of the outputs: in most of the cases, complex high-dimensional models present a strong asymmetry in the way the inputs influence their outputs. Therefore, the identification of these inputs is crucial to simplify the problem and to make it tractable and affordable as well.

The above considerations clarified that the uncertainty management in traffic flow simulation modeling is not a complementary topic, but rather strictly connected with the issues of modeling, calibration, application and validation.

Indeed, Bayarri et al. (2004) recognized that uncertainty modeling and quantification should be part of an integrated procedure for calibration and validation of traffic simulation models. However, the formalization of the idea was not provided therein. On

the other hand, the emergence of new techniques in many research fields (e.g. environmental modeling, system reliability and risk analysis) that allow for model simplification, development and validation, could open new research perspectives also in the field of traffic flow simulation modeling.

Therefore, in the following section, we present an integrated framework for the analysis of microscopic traffic flow simulation models in presence of uncertainty.

2.5   Proposed Framework for Uncertainty Management in Microscopic Traffic Flow Simulation Models

Based on the conceptual framework provided in de Rocquigny et al. (2008), in this section we discussed the issues related to its specification to the context of microscopic traffic flow simulation models.

The section is organized as follows. A general, goal-oriented, overview of microscopic traffic flow simulation models is given in Section 2.5.1, while Section 2.5.2 discusses the different sources of uncertainty generally entailed in the modeling process. Then, Section 2.5.3 describes the proposed framework for handling uncertainty in model simulation and the principles for its implementation (Section 2.5.4).

2.5.1 Microscopic traffic flow simulation model

Traffic flow simulation models may be distinguished with regards to several factors, ranging from the level of representation of time (static, dynamic) and space (discrete, continuous), to the flow structure (continuous fluid, individual vehicles) and the level of representation of traveler responses (e.g. driving behavior, aggregate link performance functions, pre-trip and en-route path choices, departure time choices, etc.).

In the present work, we will focus on microscopic traffic flow simulation models only. These models can be ascribed to the class of driven multi-particle models, where the flow representation is vehicle-based (Treiber and Kesting, 2013). They describe the reaction of each single driver to the surrounding traffic (e.g. acceleration and braking responses, lane-changing, merging) and traffic control systems (e.g. tolling systems, control plans at signalized intersections, dynamic speed limits, ramp-metering).

Dynamic state variables are vehicles' (longitudinal and lateral) position, speed and acceleration at each simulation step. Traditionally, each driving behavior is modeled separately from the other (e.g. car-following, lane-changing), while efforts towards joint modeling were just recently produced in the literature.

In order to reproduce the stochastic nature of traffic, in terms, for instance, of population variability of driving behaviors, path choices/en-route adjustments, departure time intervals, and so on, microscopic traffic flow simulation models use random variables and sample from random distributions to represent decisions made by the agents simulated in the models (e.g. route or lane choice decisions). Therefore, model parameters are inputted as random variables, following a (generally Gaussian) distribution model of given parameters. The drawback of this approach is that multiple simulation runs are needed to obtain reliable results.

These models are largely used in many applications where heterogeneity of traffic, driving behaviors and interactions play important roles, including the assessment of Intelligent Transportation Systems such as Advanced Driver Assistant Systems (e.g. Adaptive Cruise Control, Infrastructure- and Vehicle-to-vehicle communication systems), the assessment of control and management strategy (e.g. variable speed limits, ramp-metering, freeway lane management, integrated corridor management) and so on.

From the beginning, it is worth noting that the focus of this dissertation thesis is on driver behavioral models only, and specifically on car-following and lane-changing models.

However, in commercial simulation packages (e.g. AIMSUN (2012), VISSIM (2011), PARAMICS (2003)), driver behavioral models are only "components" (together with, for instance, the demand model, the route choice model, the assignment algorithm and the traffic control models) of the simulation software.

For the sake of clarity, in this dissertation thesis, we will refer to the "components", i.e. driver behavioral models, as "disaggregate" models, while to the micro-simulation software as the "aggregate" model.

It is clear that (model and parameter) sources of uncertainty are present in each "component" of the simulation software, as well as in their mutual interactions (e.g. in the assignment, between the route choice model and the driver behavioral models). The

common sources of uncertainty entailed in traffic flow simulation models are reviewed in the following section. Thereafter, the proposed framework for uncertainty management is presented.

2.5.2 Sources of uncertainty

To understand how uncertainty enters traffic modeling is useful to make some reasoning on the sources and the nature of uncertainty in traffic systems/models.

The trajectories of vehicles, which fully depict the evolution of traffic over a road network, are the outcome of a number of human choices and actions. For the sake of simplicity, it is generally acknowledged that a number of decisions/choices like the time to depart or the route to follow belongs to a driver "strategic" decision level, while a "tactical" level comprehends decisions and actions aimed to directly control the vehicle in the traffic stream, subject to a number of environmental constraints (e.g. road rules, traffic lights, surrounding traffic) and according to the driver strategic plans and motivations.

Traffic simulation aims to reproduce traffic over road networks by more or less explicitly modeling these strategic and tactical decision layers.

It is straightforward that this composite modeling process involves a number of uncertainty sources of different nature, often mixed in a complex way. Part of this uncertainty can be directly imputed to the (in)adequacy of the models to the reality, while another part depends on the (uncertain) model inputs. In the literature, it is generally referred to the former as *model* uncertainty, while to the latter as *parameter* uncertainty (see, for example, de Rocquigny et al., 2008).

Uncertainty due to the inadequacy of models, i.e. *model* uncertainty, arises from a number of sources like the modeling basic assumptions, the structural equations, the level of discretization, the numerical resolution method, and so on. Such uncertainty can be reduced by "improving" the model concerning one or more of these aspects. As the cost of reducing such uncertainties often results in the increase of computational time, the choice of the most appropriate modeling framework depends on the specific application (e.g. on-line vs. off-line simulation) and stems from a tradeoff between model adequacy and computing time.

As regards to *parameter* uncertainty, instead, we must distinguish between those inputs which are observable and those which are not. Such distinction is crucial as it affects the possibility, or the cost, of reducing the uncertainty they are responsible for:

- As observable, we intend those model inputs which have a measurable equivalent in the reality. Thus they can be *directly estimated* and used to fed the models. In a microscopic traffic flow simulation model, examples are: the network characteristics, the traffic lights timing, the traffic composition, the distribution of vehicles size, etc.

- Unobservable inputs are those which either are hardly measurable[1], like the OD demand, or have not an actual equivalent in the reality. Concerning the latter case, the most of traffic model parameters, for example, either do just not have a physical interpretation, i.e. they are simply model constants, or they are deliberately considered uncertain by the modeler. In facts, as traffic models are necessarily only coarse representations of the real system, considering modeling parameters as uncertain inputs is commonly taken to cover both the epistemic uncertainty regarding the un-modeled details of the phenomena and the aleatory uncertainty not predicted by the average models[2] (e.g. the variability in time of driver's behaviour). Such parameters can be therefore only *indirectly estimated* by means of inverse analysis (see Section 5.4.3)

Although direct measurement of observable inputs could seem the most straightforward approach to model their uncertainties, it is not necessary the case. In facts, as models are only proxy of the real system, to consider modeling parameters as uncertain inputs is commonly taken to cover both the epistemic uncertainty regarding the un-modeled details of the phenomena, and the aleatory uncertainty not predicted by the average models. In this view, although *some* model inputs could be physically measured, their role in the model might not be guaranteed to be equal to that in the real system. In other words, the inadequacy of the model to reality inherently make *some* model parameters

---

[1] In this context, the immeasurability is intended from practical rather than theoretical point of view. Indeed, some quantities may not be measurable because of operational or economic constraints.

[2] Epistemic, or reducible uncertainty, refers to types of uncertainty which can be directly reduced by an increase in available data. Aleatory, or irreducible uncertainty, refers to events which remain unpredictable whatever the amount of data available.

lose their physical meaning. For this reason they should be deliberately considered uncertain by the modeler, and therefore only indirectly estimated by means of inverse analysis (i.e. model calibration).

Table 2.1 aims to give some examples of uncertainty sources in traffic modeling, classified according to their nature. The distinction is made on the practical notion of reducibility rather than on theoretical distinctions, like epistemic vs. aleatory (see, for instance, Granger Morgan and Henrion, 1990; Patè-Cornell, 1996; Saltelli et al., 2008). It is worth noting that the table gives general indications but, depending on the model and the application context, a source might be classified in different ways. A typical example is whether or not including the model uncertainty together with the uncertainty in parametric inputs.

Table 2.1: Examples of uncertainty sources in traffic modeling and their nature

UNCERTAINTY SOURCE	*UNCERTAINTY NATURE*		
	Mostly reducible	*Mostly irreducible*	*Mixed natures*
Model	Time and space discretization	Model time-invariance	Basic modeling assumptions (e.g. assignment algorithm, model structures)
Observable inputs	Road characteristics and functions; traffic control states; traffic composition; point to point demand (e.g. on freeway network), etc.		Vehicle sizes, free-flow speeds, etc., and variability in population/space
Unobservable inputs — *Hardly measurable*	Stationary OD matrices	Individual departing time	Time-varying OD matrices
Unobservable inputs — *Unphysical parameters*	Aggregate model constants		Disaggregate model constants
Unobservable inputs — *Uncertain parameters*	Fundamental diagram parameters (jam density, etc.), cost coefficients, etc.	Intra-driver variability of parameters	Reaction times; maximum acceleration/decelerations; desired speeds, etc.; and variability in population (pdfs modeling)

2.5.3 *Proposed approach*

Figure 2.2. presents the specification of the conceptual framework for quantitative uncertainty assessment to microscopic traffic flow simulation models. The framework relies on the following steps:

- Step 1: *Problem Specification* (PS)

- Step 2: *Uncertainty Modeling* (UM)

- Step 3A: *Uncertainty Propagation* (UP);

- Step 3B: *Uncertainty Quantification* (UQ);

- Step 4: *Sensitivity Analysis* (SA).
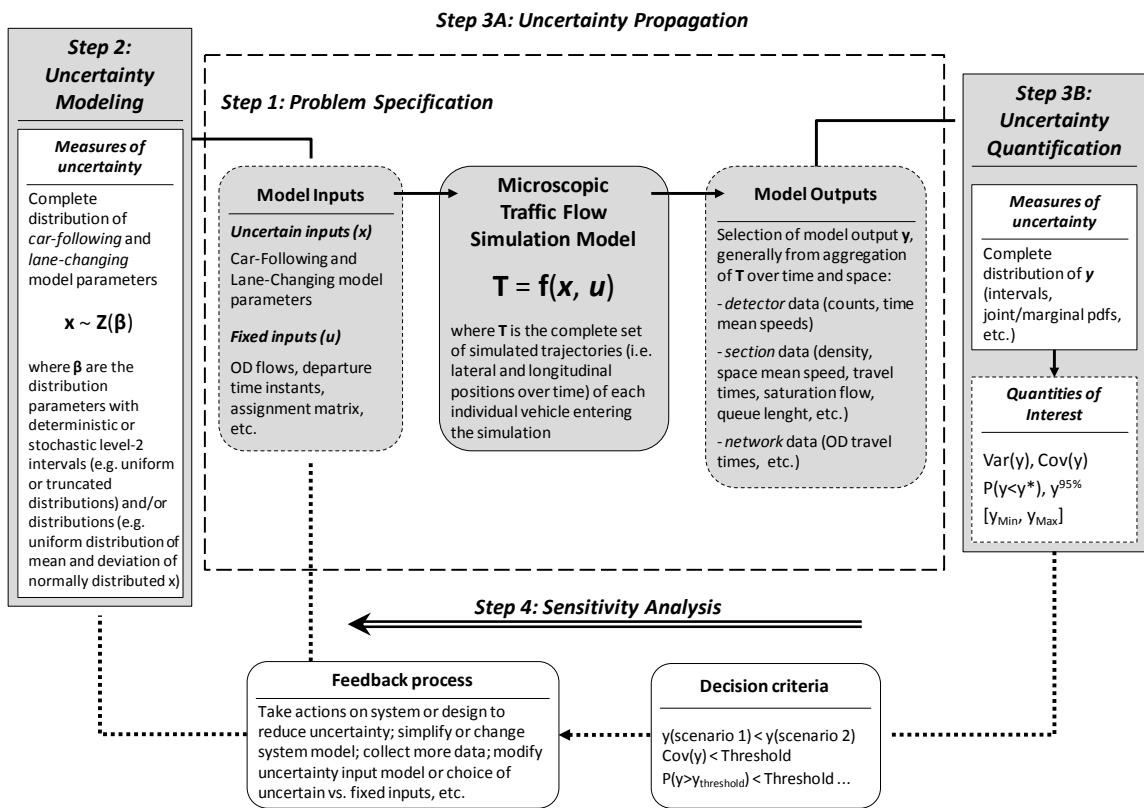
Figure 2.2: Conceptual Framework for Uncertainty Management in Microscopic Traffic Flow Simulation.

Regarding the Problem Specification, in this work we made some assumptions.

First of all, as the objective of the study is not focused on modeling itself, we focused on well-known driver behavioral models, widely adopted in the literature and usually integrated in common simulation packages, such as the Gipps' model (Gipps, 1981) and the Intelligent Driver Model (Treiber, 1999) for car-following, and the MOBIL model (Kesting, 2007) for lane-changing.

Further, as already clarified in the previous sections, for the study purposes we considered as *uncertain* only the driver behavioral model parameters, and assumed the other model inputs, such as OD flows, departure time instants, route choice model parameters, etc., as *fixed*, according to the definitions given in Chapter 1.

Moving to step 2, the objective of Uncertainty Modeling would be that of first sketch out the different sources of model and parameter uncertainty and then, if the case, provide a probabilistic modeling representation.

It is worth noting that in the proposed framework both the model and parametric uncertainties could be modeled. Indeed, in case of model uncertainties, such as time and space discretization levels, the assignment algorithm, etc., we could consider the different modeling assumptions (e.g. different simulation time intervals, different assignment algorithm) as discrete categorical random variables uniformly distributed as to cover all the possibility that we have in coding a model. However, such approach would be feasible only when the computation time for a single model run is minimal. Further, things are complicated by the fact that commercial packages usually do not allow such deep level of user customization through APIs.

Therefore, whatever the complexity of a mathematical model, it is still a simplified representation of the real traffic phenomenon. Thus, it is common practice to consider the model uncertainty alongside the parametric inputs. The estimation of the uncertain model parameters against real observed outputs thus allows covering at once the uncertainty in the system/phenomenon and the inaccuracies of the model, as well as the errors in the data. To give an example, estimating the probability distribution function of the parameters of a car-following model is needed both to account for the variability of driving behaviour within the population (i.e. *inter-driver variability*), to compensate the modeling errors such as the (un)variability in time of driving behaviors (i.e. *intra-driver variability*), but also to cover the neglected uncertainty in those model inputs considered as *fixed* by the modeler (e.g. the OD flows).

However, the estimation of all model parameters of a micro-simulation software (i.e. driver behavioral model parameters, OD flows, route choice model parameters, etc.) may result in an unaffordable task. Figure 2.3 presented a "black-box" approach to the estimation problem in a simulation-based framework. Two main issues arise from the proposed setting:

- computation feasibility, and

- uncertainty entailed in the problem setup.
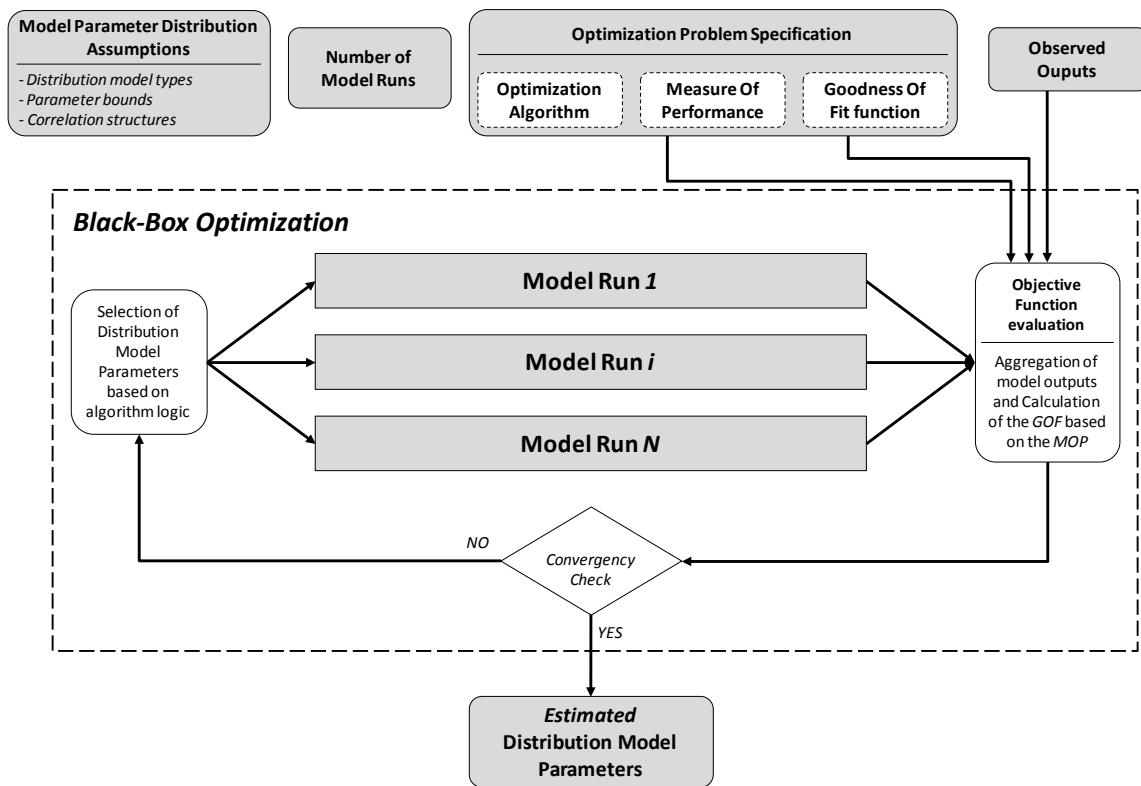
Figure 2.3: Black-box approach for the joint estimation of all parameters of a commercial microscopic traffic flow simulation software.

First of all, the most contingent drawback regards the computation effort related to the problem at hand, which becomes unaffordable with the increase in the number of parameters to be estimated, and the simulation time for a single model run. Indeed, as traffic simulation models are stochastic, multiple model runs are needed each time the objective function has to be calculated. However, if the cost of a simulation run is not

negligible (i.e. greater than few seconds, e.g. 1-2 seconds), the optimization algorithm might not reach convergence in a reasonable amount of time.

Secondly, the quality and robustness of the estimated parameters could be sensibly affected by the assumptions made on the problem setup, including, for example, the number of model runs chosen by the modeler to account for model stochasticity, the hypothesis on the distribution of model parameters (e.g. distribution model type, correlation structures, bounds, for both driver behavioral model parameters and OD flows), the correctness of the optimization problem specification (e.g. choice of the optimization algorithm, measure of performance, goodness of fit function). In addition, apart from the problem setup, several other factors may affect the quality of the solution, such as errors in the observed data or the way in which observed variability is handled and compared with model stochasticity.

Moving to step 3, the propagation of input uncertainties through the simulation model (Uncertainty Propagation) allows for the ex-post estimation of the empirical distribution of model outputs (Uncertainty Quantification). This step is actually preparatory for the sub-sequent Sensitivity Analysis.

Indeed, in Figure 2.2, Sensitivity Analysis (step 4) plays a key role, as it serves to a number of useful purposes, depending on the application setting. The importance ranking of the inputs with regards to their influence on the output uncertainty is the most common function of sensitivity analysis (*factor prioritization* setting). Sensitivity analysis can also be applied to identify which input parameters really need to be calibrated (*factor fixing* setting) and which are the observations that are really sensitive to the inputs and therefore useful for the estimation. Reducing the number of parameters to calibrate may contribute in reducing the complexity of the estimation problem presented in Figure 2.3, while the definition of the most appropriate observations is crucial to guide in the allocation of resources for the collection of new data. Sensitivity analysis may be useful also to identify the elements of the modeling process (inputs, assumptions, etc.) or the regions of the inputs which are most responsible for model realizations in an acceptable region or, at the contrary, which cause the exceeding of specific thresholds (*i/o mapping* setting). A review of the possible settings for Sensitivity Analysis is offered in Saltelli et al. (2008).

However, as the both the results of Uncertainty Quantification and Sensitivity Analysis critically depends on the level of exploration of model input uncertainties (through Uncertainty Propagation), computation feasibility issues may also arise at these steps, undermining the presented framework as the computing time for each simulation run increases.

To solve these drawbacks, it is common practice in the literature to focus on "disaggregate" models (the "components"), rather than on the "aggregate" model (the micro-simulation software). Such approach is generally adopted to study specific properties of the "disaggregate" model (e.g. linear stability of car-following models), to verify its modeling assumptions (e.g. quasi-stationary generation profile for demand models), to calibrate model parameters, and to perform model benchmarking.

However, in the literature, there is no significant contribution in understanding which are the impacts of the results of analyses conducted at the "disaggregate" level on the performances of the "aggregate" microscopic traffic flow simulation model.

Therefore, in the following section, we first specialized the methodological framework presented in Figure 2.2 to the "disaggregate" analysis of driver behavioral models. Successively, we assessed the impact of analyses results (e.g. model simplifications, model calibration) at the "aggregate" level.

2.5.4 *From traffic flow models to sub-models and vice versa*

As clarified in the previous section, focusing on a sub-model would generally allow to reduce the complexity of the methodological framework for uncertainty modeling (i.e. indirect model estimation), as well as to study more accurately some of its properties.

For instance, the analysis of car-following models have been largely undertaken in the literature for several different scopes, ranging from the studies of the population variability of driving behaviors (e.g. Kim and Mahmassani, 2011) and its implication on linear stability properties (e.g. Wilson, 2008; Ward and Wilson, 2011) to more modeling issues such as the multi-anticipative behaviors (e.g. Hoogendoorn at al., 2006).

However, the underlining hypothesis in these studies is that the inputs uncertainties of the "aggregate" simulation model can be decomposed into its sub-models' uncertain inputs.

When focusing on the "disaggregate" models, the UM issue mainly deal with the following factors:

- the scarceness, incompleteness or inconsistency of data as to the model complexity;

- the data measurement errors;

- the computational complexity of the estimation process;

- its (proper) set up as to the nature of the specific problem, and

- the asymmetry in the importance of uncertain inputs.

The scarceness, incompleteness or inconsistency of data with respect to the complexity of a model may lead either to ill-posed inverse problems – such as the case of the static OD matrix estimation problem (Marzano et al. 2009) – or to biased or not robust estimates of the parameters' pdf. The latter effect also arises in presence of measurement errors (Ossen and Hoogendoorn, 2008a). In addition, a high number of parameters can make computationally unfeasible the analysis. For example, in case of least-square "black-box" calibration of model parameters, the computational complexity is exponential in the number of parameters, making the search for a global optimum generally unfeasible even for a relatively small number of parameters (Ciuffo et al., 2008).

Further, the quality of the solution found, i.e. the chance of finding a global optimum or at least a stable solution, could be then affected by the problem setup, including the choices of the algorithm, the measure of performance and the goodness of fit function. In addition most of the models present a pronounced asymmetry in the influence of the parametric inputs on the model outputs, with a small subset of parameters accounting for most of the output uncertainty and the others playing little or no role. The calibration of model parameters with scarce influence on the outputs (i.e. flat objective functions) is a hard challenge for any optimization algorithm.

Finally, once investigated these issues on sub-models (e.g. car-following and lane-changing models), one could question about how scalable are the findings when applied for "aggregate" microscopic traffic flow simulation. For instance, what is the impact of measurement errors on the estimated model parameters when used for an "aggregate" micro-simulation study? Or what is the impact of adopting different correlation

structures among sub-model parameters on "aggregate" micro-simulation results? Or what is the effect of model simplification on the performances of the "aggregate" micro-simulation model?

In the following section, a review of the study steps undertook to accomplish the analyses at both the "disaggregate" and "aggregate" levels will be presented.

2.6   From Theory To Practice: Outline of Study Steps

Following the previous discussion, we first focused on the management of uncertainty at the "disaggregate" level. Therefore, we applied the framework to the analysis of car-following and lane-changing models, separately, facing the issues of uncertainty modeling (step 2), propagation, quantification and sensitivity analysis (step 3 + step 4). Figure 2.4 specialized the framework to the analysis of car-following models. Same considerations hold also for lane-changing models.
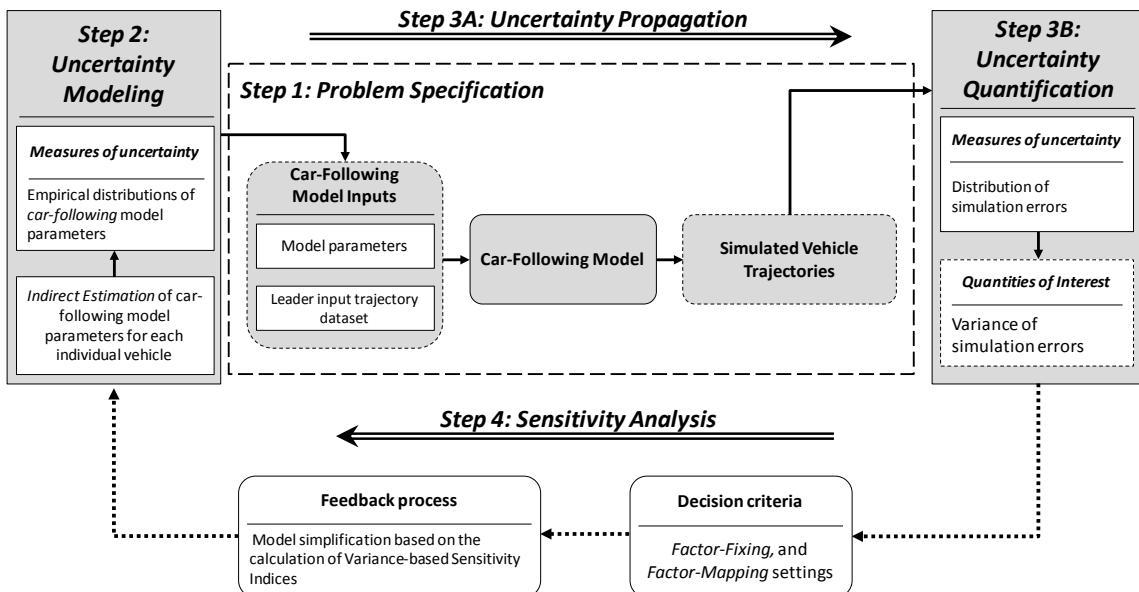
Figure 2.4: General framework adopted in this thesis for the uncertainty management in the analysis of car-following models.

For the study purpose, we used data from the NGSIM I80 project (2005), which provides a complete set of all the individual vehicle trajectories in a 500 meters x 15 minutes

space-time domain on an extra-urban road. This allowed us to conduct a two-stage analysis:

i.   First, we performed an uncertainty assessment of driver behavioral models, using individual vehicle trajectory data to accomplish for uncertainty modeling, quantification and sensitivity analysis, as illustrated in Figure 2.4.

ii.   Secondly, we evaluated the performances of the "aggregate" microscopic traffic flow model simulation, in a trace-driven environment[3], as conditioned to the estimated driver behavioral model parameters and model simplifications.

In the following, a summary of the topics investigated on driver behavioral models is first presented. They relate to the uncertainty modeling (Section 2.6.1), and to the uncertainty propagation and sensitivity analysis (Section 2.6.2). Successively, a summary of the analysis conducted on the "aggregate" microscopic traffic flow simulation model is reported (Section 2.6.3).

2.6.1 Uncertainty Modeling

As pointed out in the previous sections, the objective of this step is the definition of the joint *pdf* of car-following and lane-changing model parameters, via model calibration.

Figure 2.5 presents the general approach for the estimation of car-following model parameters of a given vehicle *i* in the traffic stream. A similar scheme could be adopted for lane-changing model parameter calibration.

In the figure, inputs of the "black-box" optimization process are:

•   the optimization problem specification

•   the measured vehicle trajectories;

Therefore, on one hand, we analyzed and quantified the uncertainty entailed in the calibration process itself, with reference to the methodological choices concerning the

---

[3] In a trace-driven simulation, both the simulated and the real systems have exactly the same inputs. Indeed, in the microscopic traffic flow simulation, each individual vehicle enters (if possible) the simulation at the same instant and in the same lane as in the observed trajectory data. More details are provided in Chapter 6.

optimization algorithm, the measure of performance and the goodness of fit function. This analysis is described in Chapter 3.

On the other hand, we studied the impact of measurement errors in trajectory data on the results of model calibrations, as to investigate whether measurement errors propagate in the estimation results. This analysis is described in Chapter 4.
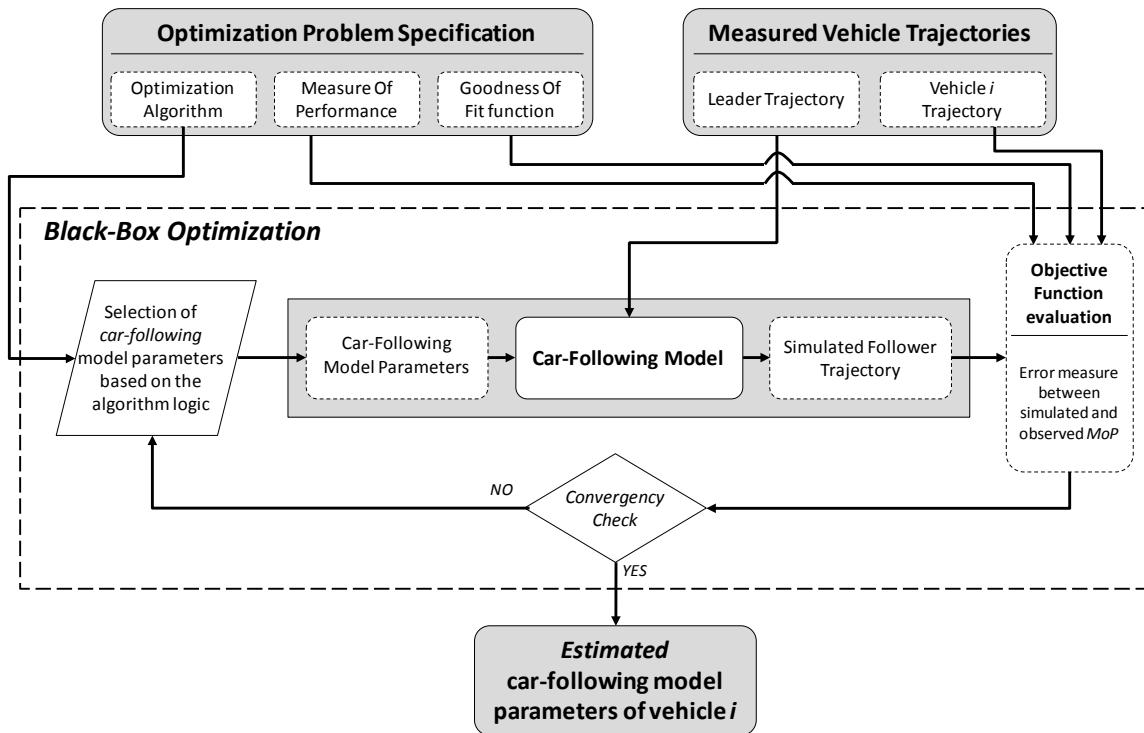
Figure 2.5: Black-box optimization for the calibration of car-following model parameters of each individual vehicle.

2.6.2 *Propagation, Quantification and Sensitivity Analysis*

Based on the UM obtained from the analyses performed in Chapters 3 and 4, a variance-based global sensitivity analysis was performed on the car-following model with the objective of simplifying it through the *factor prioritarization* and *factor fixing* settings. The output of the analysis allowed us to identify *i)* which are the most influential parameters the estimation process should focus on, and *ii)* which one could be fixed at whatever value in their range of variability without considerably affecting the model performances. This analysis is described in Chapter 5.

2.6.3 Impacts on Aggregate Microscopic Traffic Flow Simulation

Based on the results from the analysis of "disaggregate" models, we investigated which are the impacts of different uncertainty models and model simplifications on the simulation performances of the "aggregate" model.

Therefore, we performed a trace-driven simulation to compare aggregate *measured* data with *simulated* ones. As to comprehend the impact on simulation performances, in Chapter 6, we conducted different experiments to investigate the following topics:

- Impact of measurement errors on "aggregate" simulation results;

- Impact of parameter correlation on "aggregate" simulation results;

- Impact of model simplifications on "aggregate" simulation results.

2.7   Summary

Despite of the underlined importance of uncertainty management in scientific modeling, in the field traffic flow simulation modeling it is a very under investigated issue. Further, its implications on model applicability, credibility, and robustness, are frequently neglected by both practitioners and researchers.

A literature review of field applications revealed that the roles of uncertainty modeling, quantification and sensitivity analysis are largely misinterpreted, and these analyses usually cover only a marginal role in traffic flow simulation studies.

However, neglecting uncertainty in traffic flow simulation models contribute to the issues of model indeterminacy, equifinality and over-parameterization.

Therefore, in this Chapter, we presented the problem of uncertainty management in the analysis of microscopic traffic flow simulation models, specifically focusing on driver behavioral models. To this aim, we adopted the general framework proposed in de Rocquigny et al. (2008), and specialized it to handle the steps of Uncertainty Modeling, Propagation, Quantification and Sensitivity Analysis in traffic flow simulation modeling.

We pointed out at the (possible) computational and modeling drawbacks of applying such framework for the analysis of a micro-simulation software. Therefore, we provided

a distinction between the micro-simulation software – referred as "aggregate" simulation model – and its model components – referred as "disaggregate" models or sub-models.

On this basis, we substantiate the lack of appropriate literature on the relationship between the results of analyses on "disaggregate" models and the performances of the "aggregate" simulation model.

The Chapter ends with an outline of the applications carried out in the remaining of the thesis.

In Chapter 3 and 4 we focused on the problem of indirect estimation of driver behavioral model parameters (Uncertainty Modeling). In particular, we evaluated the impacts of the calibration problem specification (Chapter 3) and of the measurement errors in vehicle trajectory data (Chapter 4) on the quality and robustness of parameter estimates.

In Chapter 5, we focused on Uncertainty Propagation and Sensitivity Analysis, presenting the results of a global sensitivity analysis of driver behavioral models in a *factor fixing* setting.

Finally, in Chapter 6 we investigated the relationship between "disaggregate" modeling/analysis and "aggregate" performances of the micro-simulation software.

Chapter 3

Uncertainty in the Procedure for Calibration of Microscopic Traffic Flow Simulation Models[1]

3.1 Introduction

In the field of traffic simulation – provided the structural inadequacy of models – it is common practice to include model uncertainty alongside the uncertainty in parametric inputs. Considering model parameters as uncertain inputs to calibrate against real data, indeed, is usually taken also to cover the epistemic uncertainty regarding the un-modeled details of the phenomena and the aleatory not predicted by the models (de Rocquigny et al., 2008; Punzo et al., 2014b). In microscopic traffic flow models, for instance, indirect estimation of the probability density function of model parameters, is needed not only to account for the heterogeneity within the driver population – the so called inter-driver variability (Ossen et al., 2006; Kim and Mahmassani, 2011) – but also for compensating the model errors and the system aleatory, like the time variability of driving behaviors (i.e. intra-driver variability; Kesting and Treiber, 2008).

As clarified in Chapter 2, the compensation of the modeling errors and of the system irreducible uncertainty is the basic theoretical motivation for the indirect estimation of model parameters in traffic simulation. Conversely, it generally advises against the direct estimation of the observable parameters, namely, of those parameters which have

[1] Regarding the contents of this Chapter, the reader can refer also to Punzo et al. (2012).

a physical equivalent in the reality and can be directly measured, like for example the reaction time or the maximum acceleration in car-following models (this point has been controversial for long time, with many scientific works supporting to some extent the direct estimation of parameters, e.g. in Gipps (1981): "the parameters in the model should correspond to obvious characteristics of drivers and vehicles so that most can be assigned values without resorting to elaborate calibration procedure").

In this Chapter, we focused on the calibration of "disaggregate" driver behavioral models, and specifically on car-following models.

The calibration of car-following models based on real trajectory data has been widely applied as the basis for different type of studies ranging from the investigation and benchmarking of models (e.g. Wilson, 2001; Newell, 2002; Treiber et al., 2008) to the study of driving characteristics and model features like multi-anticipation (e.g. Hoogendoorn et al., 2006), driver heterogeneity (e.g. Ossen and Hoogendoorn, 2007, 2011) and the correlation structure of model parameters (e.g. Kim and Mahmassani, 2011) (for a review of previous calibration studies: Brackstone and McDonald, 1999).

In spite of the large number of studies attempting to deepen the properties of the models/phenomena through the results of calibrations, very few of them attempted also to analyze and quantify the uncertainty entailed in the calibration process itself, and its impacts on the accuracy and reliability of results. For example, Brockfeld et al. (2004) recognized that many optimization algorithms get stuck in local minima and suggested to start the algorithms (at least five times) from different starting points, as also indicated in Ossen et al. (2006). Punzo and Simonelli (2005) pointed at the effect on calibration results of using different measure of performances in the objective function, namely speed, inter-vehicle spacing and headway, providing numerical comparisons and a conceptual justification of the advantage of using spacing. Kesting and Treiber (2008) confirmed the justification by Punzo and Simonelli (2005) for preferring inter-vehicle spacing and compared the effect on results of using different goodness of fit functions in the objective, like relative error, absolute error and mixed error. Finally, Ossen and Hoogendoorn (2008a, 2009a) asserted the preeminent role of experiments with synthetic data to investigate calibration issues, and, in such a framework, showed that measurement errors can yield a considerable bias in the estimation results. They also raised the crucial issue that parameters minimizing the objective function do not

necessarily best capture car-following dynamics and, as a general conclusion, they stated that "*calibration based on real trajectory data turns out to be far from trivial*". Therefore, despite these specific investigations, so far there is not a thorough understanding of the mutual effect of the specific issues raised as well as of the whole problem of car-following model calibration (against vehicle trajectory data), in light of the traffic flow theory.

If, on the one hand, the complexity of the calibration problem stems mainly from: *i)* the scarceness, incompleteness or inconsistency of data as to the model complexity, *ii)* the measurement errors in the data, *iii)* the computational complexity of the analysis, and *iv)* the asymmetry in the importance of uncertain parametric inputs, on the other hand, the true possibility to success in a calibration effort depends on the setup of the calibration problem as partially highlighted in the previous literature. We mostly refer to the methodological choices concerning:

- the optimization algorithm,

- the measure of performance (MoP),

- the goodness of fit function (GOF).

In the following, we refer to the choices of the optimization algorithm, the MoP and the GOF function as the *optimization problem specification*.

In this view, the Chapter focuses on the main findings of a vast exploratory study aimed at investigating and quantifying the uncertainty entailed in the calibration process (Ciuffo et al., 2012a). According to a full-factorial design, all the combinations of algorithms, measure of performances and goodness of fit functions applied so far in the field literature were tested. Each test was performed several times from different starting points in order to unveil the impact of the initial setting on the calibration results. The methodological approach followed in this Chapter is based on experiments with synthetic data, i.e. data generated by the model itself, as this is the only way to ascertain if and how "good" parameters are identified by the calibration procedure (as also pointed out in Ossen and Hoogendoorn, 2009a). Given the objectives, the Gipps' car-following model was chosen because of the relatively high number of parameters (necessary for the study purpose), the presence of a delay in the formulation and for the acknowledged understanding of its properties (Gipps, 1981; Wilson, 2001; Punzo and Tripodi, 2007).

The Chapter is organized as follows. Firstly, Section 3.2 presented a state-of-the-art in microscopic traffic flow model parameters calibrations, jointly with a review of the principles and methodologies commonly adopted in the literature for car-following model calibration. Then, the basic idea beyond the verification procedure is introduced in Section 3.3. Section 3.4 presents the formulation of the Gipps' car-following model, and of the specialized calibration problem. Section 3.5 is dedicated to the data description. Section 3.6 describes the experimental design, while the analysis of calibration results is presented in Section 3.7. The work ends with conclusions and recommendations for future research.

3.2   State-Of-The-Art in Car-Following Model Calibration

In this section, a literature review of the studies on the calibration of car-following models is presented in Section 3.2.1, followed by a background on the principles and methodologies adopted for car-following model parameters estimation (Section 3.2.2).

3.2.1 Review of the studies on car-following model calibration

As soon as microscopic traffic flow models started to be used to simulate complex contexts, to support the design of new traffic facilities or to evaluate the effect of alternative traffic operations, weaknesses of models in representing the real world became clear. One of the major reasons for these was identified in the inadequacy of their sub-models, including car-following ones.

First car-following models were developed after the pioneering study performed by a group of researchers of the General Motors (GHR model by Gazis et al., 1961; and successive model derivations), between the mid-1950s and the 1970s, and after the contributions of some other investigators in the beginning of the 1980s (Gipps' model by Gipps, 1981; LWR model by Leutzbach and Wiedemann, 1986). These models were developed through a straightforward deductive approach, based on simple assumptions. Later, in the 1990s, researchers started to focus on the study of the theoretical properties of such models, and this lead to the development of new theories (for example, the Optimal Velocity Model by Bando et al., 1995; the Asymmetric Full Velocity Different Model by Helbing and Tilch, 1998; the Intelligent Driver Model by Treiber et al., 2000;

the Newell's model by Newell, 2002; the Kerner's Stochastic Three Phases Traffic model by Kerner et al., 2007).

In contrast with the increasing efforts in model development, empirical verification of the assumptions and model calibration have encountered serious difficulties across the years, due to the accuracy and quality of collected data. Different data sources were used to run calibration experiments: data from loop detectors, section measurements (such as individual travel times), and trajectory data. However, major technological issues arouse in the collection of microscopic data related to the vehicle motion and, also for this reason, previous findings from calibration studies have often been contradictory.

First experiments were based on small-scale observations of flow data which provided information only at an aggregate level (1-min or 5-min averages of flows and speeds) and at a limited number of cross sections, where dual loop detectors were located. Little was known about differences in car-following behavior between individual driver–vehicle combinations, also because of the lack of detailed microscopic traffic data (for a comprehensive review of previous studies, see the work by Brackstone and McDonald, 1999).

More recently, Brockfeld et al. (2003) performed car-following model calibrations by using individual travel times between several observers along a one-lane rural road, given as boundary conditions the flow into this road and the flow out of it. The task was to predict individual travel times and to estimate the best matching set of parameters for each of the tested models, by using nonlinear optimization techniques. The models with better performances were the ones with the smallest number of parameters. However, the average error rate of the estimates was not reliable at all, with fluctuation between 2.5% and 25% among different parts of the dataset. Therefore, from the very first, it was clear how complex was the calibration procedure on this kind of models, when attempting to obtain reliable estimates of model parameters.

Soon after, the same authors compared the models by calibrating and validating them with data from dual loop detectors on a multilane freeway (Brockfeld at al., 2005). To simplify this task, the models were tested by a single-lane simulation in the place of a multilane freeway simulation. The results show that, although lane-changing maneuvers were completely left out of the simulation, low calibration errors of 14% to 16% could have been obtained. However, the reliability of such estimates was unknown.

Starting from the early 2000s, advancements in digital technology have opened up wide new horizons for the research in this field. Technological innovation in microscopic data-collection methods (such as Differential GPS, and so on) have caused a considerable increase in the number of studies using trajectory data for calibrating car-following models. These studies were important, first, because they provided the opportunity to assess the performances of existing car-following models at the microscopic level, that is, the accuracy of the models in predicting the behavior of an individual driver. Second, the outcomes of these studies could be used to obtain new empirical insights into car-following behaviors, such as the heterogeneity of traffic flows, the degree of multi-anticipation behavior, or the effect of model errors and system aleatory, like the inter-driver variability of driving behaviors.

In this view, several projects were initiated with a primary focus on microscopic modeling, and thus, high-quality traffic and trajectory data were collected in order to support the research. Hereafter, trajectory data has become the most extensively used data source in car-following calibration studies.

First studies with trajectory data aimed at the benchmarking of car-following models, in order to evaluate the variation in the performances of different models and their ability to catch traffic dynamics. Ranjitkar et al. (2004) performed the calibration of several car-following models against trajectory data collected via real-time kinematic Global Positioning Systems equipped on cars moving on a test track in Japan. Then, they set up a methodology to evaluate and compare model performances, testing various driving conditions by means of different levels of disturbance of the leader vehicle's speed. Brockfeld et al. (2004), instead, attempted to calibrate ten different models using the same trajectory dataset, and showed that the error measurements on time-headways were as low as in a previous study (Brockfeld et al., 2003), ranging from 17% to 22%. However, they found out that no models perform better than the others, but those with high-number of parameters were prone to model overfitting, that is the adaptation of the model to a particular situation which limited the capability to extend results to other situations. Evidence on model overfitting was provided also in Punzo and Simonelli (2005). They analyzed the behavior of four car-following models that differed greatly in both approach and complexity. Calibration was performed against a set of trajectory data acquired through kinematic differential Global Positioning System instruments installed on four vehicles driving under real traffic conditions on both urban and extra-urban

roads. Model calibrations showed results similar to those obtained in other studies that used test track data (Ranjitkar et al., 2004; Brockfeld et al., 2004). Instead, cross validation using different trajectories by the same drivers resulted in higher deviations compared with those from previous studies (with peaks in cross validations between urban and extra-urban experiments), confirming the scarce robustness of models and the need for empirical investigation of the variability in time of driving behaviors (intra-driver variability).

These studies indirectly questioned also the capability of the models in reproducing inter-driver variability of driving behaviors, and this topic became soon after the objective of several other research works. Ossen and Hoogendoorn (2005) estimated the parameters of different specifications of the well-known GHR car-following rule for individual drivers, using vehicle trajectory data extracted from high-resolution digital images collected at a high frequency from a helicopter. They found out that considerable differences between the car-following behaviors of individual drivers could be identified, which lead to stress the idea of inadequacy of models in reproducing heterogeneity. On this basis, the same authors performed, in simulation, a cross-comparison analysis of car-following models regarding their average performances and their specific performances for each individual driver (Ossen et al., 2006). The prime objective of this cross-comparison was to study the inter-driver differences. Average model performances revealed that the simplest models are generally not able to capture the dynamics of car-following behavior correctly, whereas individual estimates showed that the performances of more elaborate models differ between drivers. As a conclusion, they showed that inter-driver differences cannot be caught by different parameter settings by themselves, and more complex models are needed. Same evidence was provided by Hoogendoorn et al. (2006) where the multi-anticipative car-following behavior (i.e., driver behavior that includes responses to multiple vehicles ahead) was studied. Two well-known models incorporating multivehicle stimuli (Bexelius, 1968; Lenz et al., 1999) were calibrated against the trajectory data discussed in Ossen and Hoogendoorn (2005). The study investigated the nature of multi-leader stimuli, giving insights into the number of vehicles ahead to which drivers react and the kind of stimuli to which drivers respond. Large inter-driver variability in multi-leader driving behavior was found, and, thus, different models were needed to describe driver heterogeneity correctly. For interested readers, an extended study that gives insights into the level of

heterogeneity in car-following can be found in Ossen and Hoogendoorn (2011). Following this research line, Kim and Mahmassani (2011) studied the expected correlation among car-following model parameters. They focused on the investigation of the impact of neglecting parameter correlations on the resulting movement and properties of a simulated heterogeneous vehicle traffic stream. Results suggested that the use of parametric distribution with known correlation structures could reduce the errors due to ignoring correlation; however, the effect also varied depending on model specification.

Ultimately, indirect estimation of model parameters was considered not only to account for the heterogeneity within the driver population, but also for compensating the model errors and the system aleatory, like the time variability of driving behaviors. On this topic, Kesting and Treiber (2008) showed that intra-driver variability rather than inter-driver variability accounts for a large part of the calibration errors.

3.2.2 Review of principles and methodologies for parameters calibration

In Table 3.1, a collection of studies dealing with calibration efforts is presented. Various approaches have been used to solve the calibration problem, by combining different *estimation methods*, *measures of performance*, *goodness of fit* functions and *optimization algorithms*. A review of the possible combination of the previous elements is provided in the following.

Table 3.1: Review of the most used settings for car-following model parameters calibration.

ESTIMATION METHOD	MEASURE OF PERFORMANCE	GOODNESS OF FIT function / ESTIMATOR			
		Error measures	*GEH statistics*	*Theil's Inequality Coefficient*	*Likelihood function*
Least Squares (LS)	*Time-headway*	Brockfeld et al. (2004) Punzo and Simonelli (2005)		Punzo and Simonelli (2005)	
	Inter-vehicle spacing	Ranjitkar et al. (2004) Punzo and Simonelli (2005) Kesting and Treiber (2008) Punzo et al. (2012)	Punzo et al. (2012)	Punzo and Simonelli (2005) Ossen and Hoogendoorn (2008a) Punzo et al. (2012)	
	Speed	Ranjitkar et al. (2004) Punzo and Simonelli (2005) Punzo et al. (2012)	Punzo et al. (2012)	Punzo and Simonelli (2005) Ossen and Hoogendoorn (2008a) Ciuffo et al. (2012a) Punzo et al. (2012)	
	Speed and inter-vehicle spacing	Punzo et al. (2012)		Ossen et al. (2006) Ossen and Hoogendoorn (2008a, 2009) Kim and Mahmassani (2011) Punzo et al. (2012)	
	Acceleration	Ossen and Hoogendoorn (2005)			
Maximum Likelihood Estimation (MLE)	*Speed*	-	-	-	Hoogendoorn et al. (2006) Hoogendoorn and Hoogendoorn (2010a) Hoogendoorn and Hoogendoorn (2010b)
	Acceleration				Ahmed (1999) Toledo et al. (2009)
Bayesian	*Speed*	-	-	-	van Hinsbergen et al. (2009)

Estimation methods

Above all, three main estimation techniques have been used so far in the literature:

- Least Squares (LS) method;

- Maximum Likelihood Estimation (MLE) method;

- Bayesian method.

The Least Squares method is definitely the most widely applied technique in car-following calibration studies. A review of this approach can be found in Punzo and Simonelli (2005), as well as in Ossen and Hoogendoorn (2005, 2008b). Basically, the problem formulation is presented in Eq. 3.1:

$$P^* = \arg \min_{P \in D} f(Y^{obs}, Y(P)^{sim}) \qquad (3.1)$$

where:

- P is the vector of the model parameters p_i, with $i = 1, ..., m$;

- D is the domain of feasibility of the model parameters, eventually constrained by the upper and lower bounds and by the linear and non-linear constraints;

- $f(\cdot)$ is a scalar valued non-linear function which measures the distance between observed and simulated following driver's behaviour;

- Y^{obs} and $Y(P)^{sim}$ are, respectively, the observed and simulated outputs;

The domain of feasibility of model parameters is defined by the parameters' bounds and potentially by other (linear and non-linear) constraints:

$$LB_i \le p_i \le UB_i \qquad i = 1, ..., m \qquad (3.2)$$

$$g_j(P) ? b_j \qquad j = 1, ... n \qquad (3.3)$$

where:

- LB_i and UB_i are, respectively, the lower and upper bounds of the parameter p_i;

- $g_j(\cdot)$ is a scalar valued linear or non-linear function of the vector of model parameters P, that evaluates the left hand side of the *j*-th constraint;

- b_j is a constant value equal to the right hand side of the j-th constraint;

- $?$ is one the following relational operators: " $\leq$ ", " $\geq$ " or " $=$ ";

According to this framework, the most applied estimators in the literature were error measures. Among the others, absolute errors, square errors, percentage errors and mixed errors were the most used. On the other hand, also the Theil's Inequality coefficients and the GEH statistics were recently applied in some other calibration studies (for a comprehensive review, see Table 3.1).

The Maximum Likelihood Estimation has been widely applied in car-following model calibration. Ahmed (1999) presented the formulation of the unconditional distribution of the accelerations that constituted the likelihood function formulation for the follower driver. Hoogendoorn and Ossen (2005), Hoogendoorn et al. (2006) and Hoogendoorn and Hoogendoorn (2010a, 2010b) reformulate the likelihood to estimate the parameters of a generalized form of car-following models, while Toledo et al. (2009) applied the method for the estimation of the parameters of the extended non linear GM model, shown in Ahmed (1999).

In a discretized form, car-following models can be expressed as follows:

$$v_i^{sim}(t_{k+1}) = \mathbb{f}[T, y_i(t_k), y_i(t_k - \tau)|\mathbf{\theta}]$$

here v_i^{sim} is the simulated speed of driver i, $\mathbf{\theta}$ denotes the set of parameters describing the car-following behavior, while T denotes the time step used for discretization. The vector $y_i(t_k)$ denotes the state that is relevant for driver i at time instant t_k.

The following relation between the speed data and the predicted speed is assumed:

$$v_i^{obs}(t_{k+1}) = \mathbb{f}[T, y_i(t_k), y_i(t_k - \tau)|\mathbf{\theta}] + \epsilon(t_k) = v_i^{sim}(t_{k+1}) + \epsilon(t_k)$$

The error term $\epsilon(t_k)$ is introduced to reflect errors in the modeling, similar to the error term used in multivariate linear regression. The error terms are generally serially correlated, which is described later in this section. For now, assume that the error term is a zero mean normally distributed variable with standard deviation σ.

Assuming that the difference between the prediction and the observation follows a normal distribution with mean 0 and standard deviation σ, the likelihood of a single prediction can be thus determined as follows:

$$\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{\epsilon(t_k)^2}{2\sigma^2}\right)$$

Because it has been assumed that the errors are uncorrelated, the probability of a set of observations k = $1, \ldots, n$ can be determined, with the likelihood of the sample as a result:

$$L(\Theta, \sigma) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{\epsilon(t_k)^2}{2\sigma^2}\right)$$

Applying a log-transformation, it results:

$$\tilde{L}(\Theta, \sigma) = -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^{n} \epsilon(t_k)^2$$

Maximum-likelihood estimation involves finding the parameters that maximize the log likelihood. A necessary condition for the optimum allows the determination of the standard deviation:

$$\frac{\partial \tilde{L}}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma} = \frac{1}{n} \sum_{k=1}^{n} \epsilon(t_k)^2$$

That is, the maximum-likelihood estimate for the variance of the error is given by the mean standard error of the predictions and the observations. For the remaining parameters, the maximum-likelihood estimates can be determined by numerical optimization:

$$\boldsymbol{\theta} = \arg\max \tilde{L}(\Theta, \sigma)$$

with:

$$\tilde{L}(\Theta, \hat{\sigma}) = -\frac{n}{2}\left[ln\left(\frac{2\pi}{n} \sum_{k=1}^{n} \epsilon(t_k)^2\right) + 1\right]$$

This expression shows that maximization of the log-likelihood is equivalent to minimization of the prediction error (mean squared error).

However, subsequent error terms in trajectory data are not independent, showing the existence of serial correlation or autocorrelation. A review of the approach to deal with serial correlation can be found in Hoogendoorn and Hoogendoorn (2010a, 2010b).

The Bayesian approach is a generalization of the Likelihood Ratio Test (LRT) introduced in Hoogendoorn et al. (2006). To test whether one model performs better than another model, the likelihood ratio test is performed. To this end, the zero-acceleration model is used as a reference model:

$$v_i^{obs}(t_{k+1}) = v_i^{sim}(t_k) + \epsilon(t_k)$$

For this model, one can determine the (null) log-likelihood:

$$\tilde{L}_0 = -\frac{n}{2}\left[ln\left(\frac{2\pi}{n}\sum_{k=1}^{n}\epsilon(t_k)^2\right) + 1\right]$$

The LRT involves testing the statistic:

$$2[\tilde{L}(\Theta, \hat{\sigma}) - \tilde{L}_0]$$

which follows a χ^2 distribution with degrees of freedom equal to the number of model parameters to calibrate. The LRT is passed with (1-α) confidence if:

$$2[\tilde{L}(\Theta, \hat{\sigma}) - \tilde{L}_0] > \chi^2(1 - \alpha, d)$$

The likelihood-ratio test can also be used to cross-compare the performance of two different car-following models. In this case, *d* denotes the difference in the number of parameters of the complex model and the simple model. The test accounts for the number of parameters (via the degrees of freedom *d*) and thereby makes it possible to fairly compare simple and complex models.

In the Bayesian method, prior probabilities are transformed into posterior probabilities for each parameter in the car-following model, for which Bayes' rule is used. The exact formulation of this method for calibration and model selection is presented in van Hinsbergen et al. (2009).

Measures Of Performance

In the case of calibration of car-following models the measures of performances should capture the dynamics of the phenomenon as it develops (Punzo et al., 2005). To this aim, the MoPs most used in the literature so far were the following:

- Time series of the follower's speeds (V);
- Time series of the inter-vehicle spacing between leader and follower (S).

However, in some other cases (see Table 3.1), also the time-headway or the acceleration have been adopted.

Goodness Of Fit functions

Most widely used error measures were the Root Mean Square Errors and the Mean Absolute Error, defined in the following:

- Root Mean Square Error RMSE (Punzo et al., 2005; Ciuffo et al., 2008; Ciuffo and Punzo, 2010; Punzo et al., 2011a):

$$RMSE(Y) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(Y_i^{obs} - Y_i^{sim}\right)^2} \tag{3.4}$$

- Mean Absolute Error MAE (Ma and Abdlulhai, 2002; Kim et al., 2005; Ciuffo and Punzo, 2010; Punzo et al., 2011a):

$$MAE(Y) = \frac{1}{N} \sum_{i=1}^{N} \left|Y_i^{obs} - Y_i^{sim}\right| \tag{3.5}$$

In other studies, instead, different statistics have been adopted, such as the GEH statistics, and the Theils' Inequality Coefficient:

- GEH Statistics with a threshold value equal to 1 (Ma et al., 2007; Ciuffo and Punzo, 2010; Punzo et al., 2011a):

$$GEH_{Threshold\ Value}(Y) = \frac{N - \sum\limits_{i=1}^{N}\delta_i}{N}$$

$$\delta_i = \begin{cases} 1 & if\ GEH_i(Y) \leq Threshold\ value \\ 0 & elsewhere \end{cases}$$

$$GEH_i(Y) = \sqrt{\frac{2 \cdot \left(Y_i^{obs} - Y_i^{sim}\right)^2}{\left(Y_i^{obs} + Y_i^{sim}\right)}}$$

(3.6)

- Theil's Inequality Coefficient U (Punzo et al., 2005; Brockfeld et al., 2004; Brockfeld et al., 2005; Ossen et al., 2008a; Ossen et al., 2009; Kim and Mahmassani, 2011, Ma and Abdlulhai, 2002; Ciuffo and Punzo, 2010; Punzo et al., 2011a):

$$U(Y) = \frac{\sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_i^{obs} - Y_i^{sim}\right)^2}}{\sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_i^{obs}\right)^2} + \sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_i^{sim}\right)^2}}$$

(3.7)

$$U(Y_1,Y_2) = \frac{\sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_{1,i}^{obs} - Y_{1,i}^{sim}\right)^2}}{\sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_{1,i}^{obs}\right)^2} + \sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_{1,i}^{sim}\right)^2}} + \frac{\sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_{2,i}^{obs} - Y_{2,i}^{sim}\right)^2}}{\sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_{2,i}^{obs}\right)^2} + \sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}\left(Y_{2,i}^{sim}\right)^2}}$$

(3.8)

Optimization Algorithms

Given the problem specifications reported in Table 3.1, three main optimization algorithms have been used in the field of car-following model calibration to find the model parameter estimates, and are reviewed in the following.

- Downhill Simplex;

- Genetic Algorithm;

- OptQuest Multistart.

The Nelder–Mead method or Downhill Simplex method was proposed by John Nelder and Roger Mead (Nelder et al., 1965). The Nelder–Mead technique is a gradient-free optimization method, widely used in many car-following model calibration studies since 2004 (Brockfeld et al., 2004; Brockfeld et al., 2005; Ossen et al., 2006; Ossen et al., 2008a; Ossen et al., 2009; Kim and Mahmassani, 2011; Punzo et al., 2012).

It is a common unconstrained nonlinear optimization technique, and relies on a well-defined numerical method for twice differentiable problems. However, the Nelder–Mead technique is only a heuristic, since it can converge to non-stationary points (Powell, 1973; Lagarias et al., 1998; McKinnon, 1999). As the algorithm does not allowed the setting neither of parameters bounds nor of constraints, most car-following calibration studies have added a penalty value to the objective function value to account for the possibility that parameter values are not defined in the domain of feasibility, or violate any constraints (see the problem specification in the Least Squares method).

Genetic algorithms are widely used algorithms for the calibration of microscopic traffic simulation models. The reason is quite straightforward since no information on the objective function is required for their application, and thus they are suitable for "black-box" optimization. For the calibration of microscopic traffic flow simulation models, they have been applied several times (see, for example, Ma and Abdlulhai, 2002; Schultz and Rilett, 2004; Kim et al., 2005; Ma et al., 2007). With regards to car-following models, see Ranjitkar et al., 2004; Kesting and Treiber, 2008; Punzo et al., 2012.

Even though the genetic algorithm is suitable for solving constrained non-linear optimization problem, only the parameter bounds were set to design the problem, since it was recognized that non-linear constraints heavily slowed down the optimization. Indeed, the genetic algorithm uses the Augmented Lagrangian Genetic Algorithm (ALGA) to solve nonlinear problems. With this approach, bounds and linear constraints are handled separately from nonlinear constraints. Thus, a sub-problem is formulated by combining the fitness function and nonlinear constraint function using the Lagrangian and the penalty parameters. A sequence of such optimization problems are approximately minimized using the genetic algorithm such that the linear constraints and bounds are satisfied. As a result, the algorithm minimizes a sequence of the sub-problem, which is an approximation of the original problem, resulting in an increase of the number of function evaluations needed to solve it (Goldberg, 1989; Conn et al.,

1991, 1997). Thus, to limit the computing time, it is often applied a penalty function, to simulate violations of any non-linear constraints.

Since it is not a global optimizer, the genetic algorithm could face difficulties in finding a stationary global solution. However, the genetic algorithm can sometimes overcome this deficiency with the right settings. Indeed, with a large population size, the genetic algorithm searches the solution space more thoroughly, thereby reducing the chance that the algorithm will return a local minimum (Powell, 1973). Concurrently, a large population size also causes the algorithm to run more slowly.

The OptQuest/Multistart heuristic (Ugray et al., 2005) is an optimization algorithm for solving both constrained and unconstrained global optimization problems. It has been recently used for the calibration of car-following models in Punzo and Simonelli (2005) and in Punzo et al. (2012).

Basically, the algorithm employs a Scatter Search meta-heuristic (Glover, 1998) to provide starting points for a Generalized Reduced Gradient NLP solver (Smith et al., 1992; Drud, 1994). In this way it tends to combine the seeking behavior of a gradient-based local NLP solvers with the global optimization abilities of a Scatter Search. In practice, the Scatter Search performs a preliminary exploration in the parameters' domain in order to locate different starting points for a local gradient-based descent (which converges to the "nearest" local solution). Adopting a high number of maximum local search allowed, the probability to find the global solution of the optimization problem could increase. The major shortcoming with this approach is in the high number of objective functions evaluations (which increases with the numbers of parameters to be calibrated) required to converge towards a (possible) global solution.

3.3   Proposed Verification Framework

As emerges from the analysis of the literature, the inner source of uncertainty of any optimization procedure derives from the problem specification itself and, thus, relies on several factors, among which, at least, the following: *i*) the choice of the model parameters to calibrate, *ii*) the choice of a Measure of Performance (MoP) to describe the status of the system, *iii*) the choice of the Goodness Of Fit (GOF) function used to

evaluate the overall performance of the simulation model in the objective function, and
iv) the choice of the optimization algorithm to solve the problem.

Since each choice produces different results for the same optimization problem, a
methodology to verify the goodness of the adopted specification is crucial for evaluating
the reliability of the estimation results.

The proposed verification framework is presented in Figure 3.1.
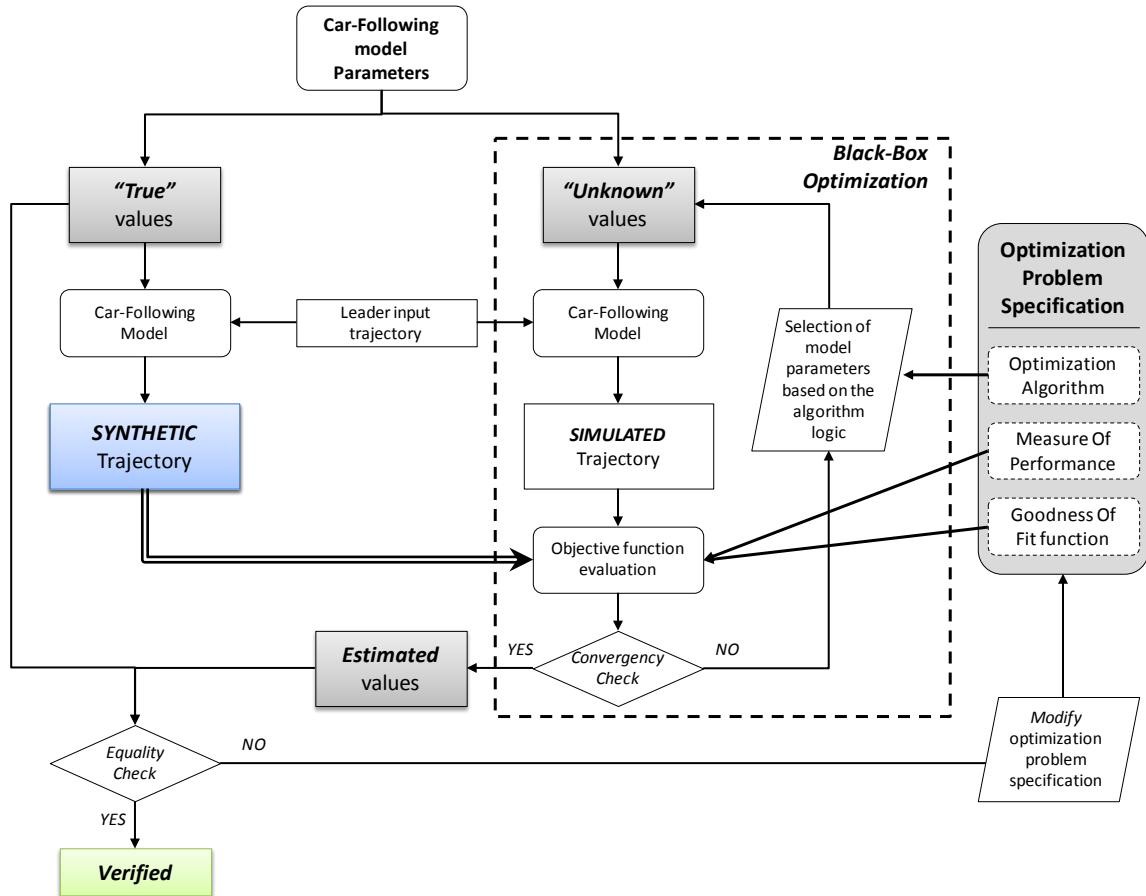
Figure 3.1: General Framework for the *verification* of the optimization problem specification.

The basic idea is that whenever one knew the values of the parameters which turned into
the global minimum, the overall problem specification should first guarantee that such
global optimal solution can be found and, then, that the algorithm is actually able to find
it. The only way to asses this, is using synthetic data, that are generated from the model
itself by fixing the parameters to a set of "known" (or "true") values. Then the
calibration aims at rediscovering the "known" values. If not succeeding in that, the
problem specification needs to be changed.

It is worth noting that the proposed verification framework is absolute general and can be apply also to different estimation problem, such as in the case of lane-changing model calibration or OD matrix estimation/correction.

3.4   Application to the Gipps' Car-Following Model

The verification approach presented in the previous section is here applied to the estimation of the parameters of the Gipps' car-following model (Gipps, 1981). The section is organized as follows. Section 3.4.1. reviewed the original formulation of the model from the literature, while Section 3.4.2 is dedicated to the description of the simulation setup. Finally, Section 3.4.3 illustrates the specification of the optimization problem for the model at hand.

3.4.1 The model

The Gipps' model (1981) is a safety-based model. It provides different transfer functions according to the two following driving regimes: the free-flow regime (Eq. 3.9) and the proper car-following regime (Eq. 3.10). A simple switching rule between the two (Eq. 3.11) drives the simulation through the motion equation (Eq. 3.12). For further details, please refer to Gipps (1981), Wilson (2001), Punzo and Tripodi (2007), Ciuffo et al. (2012b). For a review refer to Appendix B.

$$v_{a,n}(t+\tau) = v_n(t) + 2.5 \cdot a_n^{Max} \cdot \tau \cdot \left(1 - \frac{v_n(t)}{V_n^{Max}}\right) \cdot \sqrt{0.025 + \frac{v_n(t)}{V_n^{Max}}} \tag{3.9}$$

$$v_{b,n}(t+\tau) = -b_n \cdot \left(\frac{\tau}{2} + \theta\right) + \sqrt{b_n^2 \cdot \left(\frac{\tau}{2} + \theta\right)^2 + b_n \cdot \left[2 \cdot (x_{n-1}(t) - x_n(t) - S_{n-1}) - \tau \cdot v_n(t) + \frac{v_{n-1}(t)^2}{\hat{b}_{n-1}}\right]} \tag{3.10}$$

$$v_n(t+\tau) = Max\{0, min\{v_{a,n}(t+\tau), v_{b,n}(t+\tau)\}\} \tag{3.11}$$

$$x_n(t+\tau) = x_n(t) + \tau \cdot \left[\frac{v_n(t) + v_n(t+\tau)}{2}\right] \tag{3.12}$$

where:

- $v_n(t)$ and $v_{n-1}(t)$ are, respectively, the follower's and leader's speed at time t [m/s];

- a_n^{Max} is the follower's maximum acceleration rate [m/s²];

- τ is "the apparent reaction time, a constant for all vehicles" (Gipps, 1981) [s];

- V_n^{Max} is the follower's maximum desired speed, that is "the speed at which the driver of vehicle n wishes to travel" (Gipps, 1981) [m/s];

- b_n is "the most severe braking that the driver of vehicle n (i.e. the follower) wishes to undertake" (Gipps, 1981) [m/s²];

- $\theta = \tau/2$ is an additional "comfort" time lag that allows the follower not to brake always at his or her maximum desired rate [s];

- $x_n(t)$ and $x_{n-1}(t)$ are, respectively, the follower's and leader's position at time t, measured at the front bumper [m];

- $S_{n-1} = L_{n-1} + Safety$ is the effective size of the leader's vehicle, that is "the physical length plus a margin into which the following vehicle is not willing to intrude, even when at rest" (Gipps, 1981) [m];

- L_{n-1} is the physical length of the leader's vehicle of the leader [m];

- *Safety* is the safety margin "into which the following vehicle is not willing to intrude, even at rest" (Gipps, 1981) [m];

- $\hat{b}_{n-1}$ is the follower's estimate of the leader's maximum deceleration rate [m/s²];

Please note that the deceleration rates, b_n and $\hat{b}_{n-1}$, are intended as absolute values.

The Gipps' car-following model has been largely applied in the literature since its first appearance, and it is also at the basis of some commercial simulation packages (e.g. AIMSUN, 2012). A review of alternative model versions proposed in the literature so far can be found in Appendix B.

3.4.2 Integration scheme

The Gipps' car-following model is a delayed differential equation (being τ the delay). In his original paper (Gipps, 1981), Gipps found the solution of Eq. 3.4 by adopting an integration step just equal to the delay.

The original integration scheme is presented in Figure 3.2. At the instant t, the model calculates the follower's speed for the instant $t+\tau$. The reaction time τ is assumed to be a multiple of the data resolution (i.e. 0.1 s) and, thus, it is treated as a discrete variable. The follower's speed function between the instants t and $t+\tau$ is assumed linear. Finally, a forward Euler method on acceleration (i.e. a trapezoidal integration scheme on speed, see Figure 3.2) is adopted for calculations.
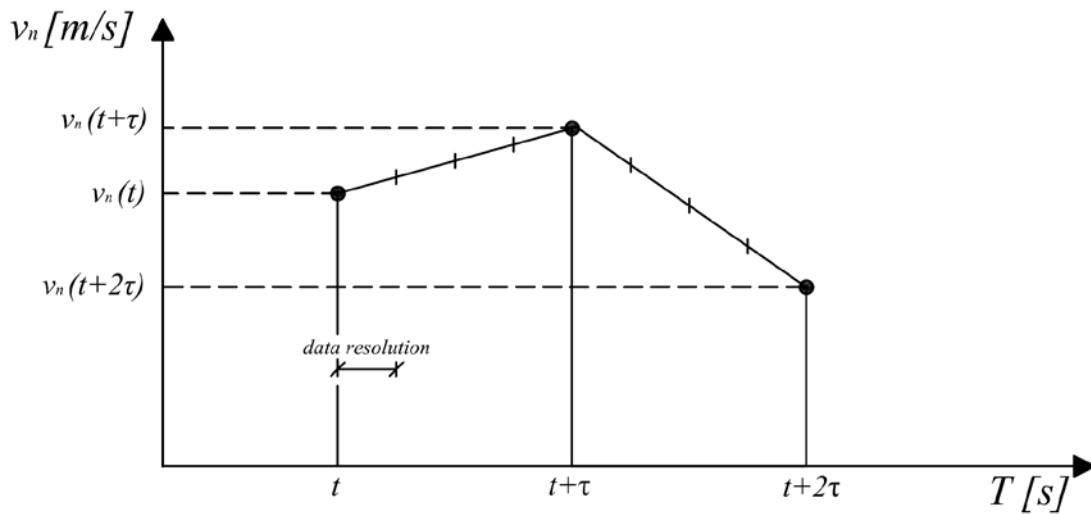
Figure 3.2: The original integration scheme for the Gipps' car-following model.

Please note that in the literature several other approaches for numerical integration of the original Gipps' car-following model have been proposed. For a review, please refer to Appendix B.

3.4.3 Optimization problem setup

As reviewed in the literature, the calibration of car-following model parameters based on vehicle trajectory data, consists of a "black-box" constrained non-linear optimization problem, where one looks for the best values of the model parameters that minimize a measure of the distance between the observed and the (model) simulated follower driver's behaviour.

Independently from the estimation method adopted, a general problem formulation can be found in Eqs. 3.1 – 3.3 (see Section 3.2.2), and here resumed for clarity:

$$P^* = \arg \min_{P \in D} f(Y^{obs}, Y(P)^{sim}) \tag{3.1}$$

$$LB_i \leq p_i \leq UB_i \qquad i = 1,...,m \tag{3.2}$$

$$g_j(P) \, ? \, b_j \qquad j = 1,...n \tag{3.3}$$

where:

- P is the vector of the model parameters p_i, with $i = 1, ..., m$;

- D is the domain of feasibility of the model parameters, eventually constrained by the upper and lower bounds and by the linear and non-linear constraints;

- $f(\cdot)$ is a scalar valued non-linear function which measures the distance between observed and simulated following driver's behaviour;

- Y^{obs} and $Y(P)^{sim}$ are, respectively, the observed and simulated outputs;

- LB_i and UB_i are, respectively, the lower and upper bounds of the parameter p_i;

- $g_j(\cdot)$ is a scalar valued linear or non-linear function of the vector of model parameters P, that evaluates the left hand side of the j-th constraint;

- b_j is a constant value equal to the right hand side of the j-th constraint;

- $?$ is one the following relational operators: " $\leq$ ", " $\geq$ " or " $=$ ";

It is worth noting that in the literature there can be also found approaches based on a Maximum Likelihood Estimation (MLE; for details, see Section 3.2.1). However, in the case of models based on time-chained equations, an analytical close-form Likelihood formulation does not exist (Law and Kelton, 2000). As a consequence, the MLE problem becomes a maximization of the (model) simulated Likelihood function that is a maximization problem equivalent to that of Eq. 3.1.

With regards to the setup of the calibration problem, specification of Eq. 3.2 (i.e. parameters' bounds) and Eq. 3.3 (non-linear constraints) are reported in the following.

Parameter upper and lower bounds

The following Gipps' car-following model parameters were calibrated: τ, V_n, a_n, *Safety*, b_n and $\hat{b}_{n-1}$. The upper and lower bounds of the parameters were fixed at the values reported in Table 3.2.

Table 3.2: Parameters' upper and lower bounds.

Parameters	Lower bound	Upper bound
τ [s]	0.1	3.0
V_n^{Max} [m/s]	10	40
a_n^{Max} [m/s^2]	0.1	8
Safety [m]	0.1	10
b_n [m/s^2]	0.1	8
$\hat{b}_{n-1}$ [m/s^2]	0.1	8

As described in Section 3.3, the verification framework here adopted is based on the use of synthetic data to estimate model parameters. Indeed, in this case, we expected that the response surface of the model would have been very steep in the neighborhood of the well-defined global optimum point (i.e. the "known" values of the parameters). Therefore, in such a case, if the combination of optimization algorithm, Measure of Performance (MoP) and Goodness of Fit function (GOF) was effective in finding the unique global minimum, the width of the range of variability of the parameters values should not influence the finding procedure.

Non-linear constraints

In order to preserve the simulation from crashing (i.e. not to obtain, at a certain time step, imaginary follower's speed values, given by negative values under the square root in Eq. 3.10), the feasible domain of the parameters was further constrained. In particular, the following two non-linear constraints were applied:

$$b_n{}^2 \cdot \left(\frac{\tau}{2} + \theta\right)^2 + b_n \cdot \left[2 \cdot \left(x_{n-1}(0) - x_n(0) - S_{n-1}\right) - \tau \cdot v_n(0) + \frac{v_{n-1}(0)^2}{\hat{b}_{n-1}}\right] \geq 0 \qquad (3.13)$$

$$V_n^{Max} \leq \frac{\tau + \theta}{\dfrac{1}{\hat{b}_{n-1}} - \dfrac{1}{b_n}} \qquad (3.14)$$

The first one (Eq. 3.13) relates to the initial state of the simulation (i.e. at t=0) and preserves from the generation of a set of parameters that make the model loosing the existence of a solution. The second constraint (Eq. 3.14), instead, assures that the speed-distance function at equilibrium is single valued, as demonstrated by Wilson (2001). In facts, Wilson conjectured that such constraint preserves the model from the global loss of existence of the solution and, in this work, such conjecture has been extensively verified in simulation. For more details, please refer to Appendix B.

3.5   Data Description

According to the objective of the work, the data used for this study were synthetic. It means that the follower trajectory has been generated through the simulation, by fixing the model parameters to a set of "known" values. The values of the parameters that were used to generate the synthetic follower trajectory were the following: $\tau = 1.0$ s, $V_n^{Max} = 30$ m/s, $a_n^{Max} = 2$ m/s^2, $Safety = 2$ m, $b_n = 2$ m/s^2, $\hat{b}_{n-1} = 2$ m/s^2. These values were chosen accordingly to common values used in the literature to simulate the drivers' behavior in urban environment. However, resuming the aim of the study, we point out that the proposed verification approach should be independent from the values of the model parameters used to generate the synthetic data (being them in the feasibility domain of parameter values).

The leader's trajectory, used to feed the Gipps' car-following model, was taken, instead, from one of the experiments carried out on a two-lane rural highway, in the area surrounding Naples (Italy). Data were acquired through instrumented vehicles, equipped with kinematic differential GPS receivers that recorded the position of the vehicle at 0.1

second interval. Differential GPS data were further processed by means of the procedure described in Punzo et al. (2005) based on a non-stationary Kalman filter. More details on the data can be found instead in Punzo and Simonelli (2005). The complete set of trajectory data is available on the MULTITUDE website (2014) for the forum members.

For the current study, the leader's trajectory is taken from the experiment 30B (Punzo and Simonelli, 2005) carried out on a two-lane extra-urban highway.

The leader's and the (synthetic) follower's speed profiles are shown in Figure 3.3(a), while the spacing profile is presented in Figure 3.3(b).
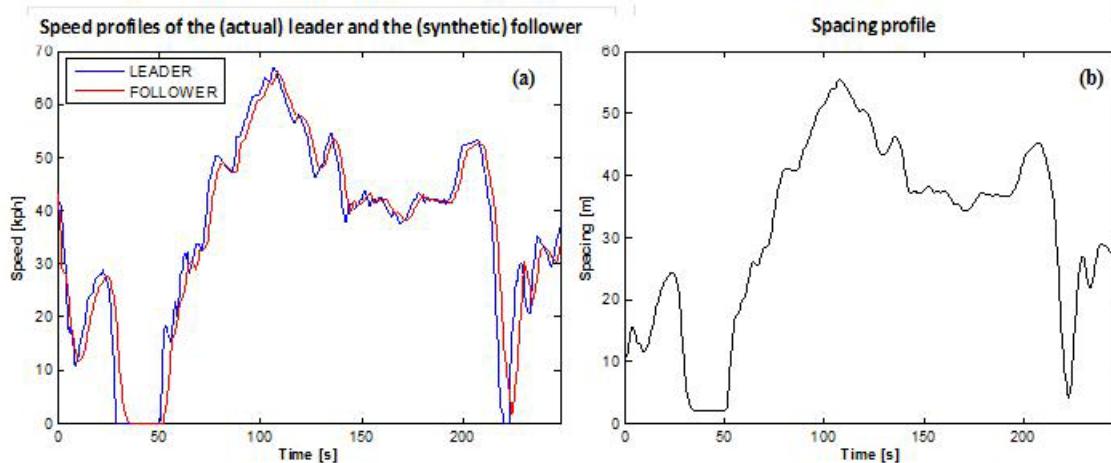
Figure 3.3: Leader's and (synthetic) follower's speed (a) and spacing (b) profiles.

3.6   Design Of Experiment

In this section, the design of the calibration experiments is presented.

Each calibration experiment was defined as an optimization problem according to Eq. 3.1, given the functional form of the objective function and the optimization algorithm to be used. The former defines the mathematical properties of the objective function and, thus, the shape of the response surface. Both affect the possibility to solve the optimization problem.

To the best of our knowledge, in the literature there is not a consolidate approach for the specification of the optimization problem, that is to define the combination of Algorithm/MoP/GOF function. Further, the lack of evidence about the capability of each

problem specification in finding the global minimum, even when synthetic data are used, led us to test all the combinations of Algorithm/MoP/GOF function most used so far in the literature (as reviewed in Section 3.2.2), in order to verify the performances of the calibration procedure and evaluate the uncertainty there entailed.

3.6.1 Tested algorithms

As reviewed in Section 3.2.2, the most used algorithms in the calibration of microscopic traffic flow models are the following:

- Downhill Simplex;

- Genetic Algorithm;

- OptQuest Multistart.

Since none of the cited algorithms is considered to be a global optimization tool, each calibration experiment (i.e. a problem with a defined combination of Algorithm/MOP/GOF) was repeated *64 times*, by using different starting points (in the case of gradient-based algorithms) or different random seeds (in the case of search-based algorithms). This approach allowed us to perform an analysis of local minima, in order to evaluate the power of the heuristic towards the capability of finding the (existing) global solution.

The different starting conditions were sampled using the Sobol' LPτ low-discrepancy quasi-random sequence, coded in MATLAB (Sobol et al., 1992), which is often used to explore the parameters' domain when conducting global sensitivity analyses.

In the following, a description of the adopted algorithm configurations is reported.

Downhill Simplex

With regards to the Downhill Simplex, we used the algorithm code embedded in MatlabR2009b (MATLAB, 2009). Since the algorithm does not allowed the setting neither of parameters bounds nor of constrained, we applied the following penalty function:

$$ObjFuncValue = \begin{cases} 100,000 & if \ x \notin D \\ ObjFuncValue & elsewhere \end{cases} \qquad (3.15)$$

where:

- *ObjFuncValue* is the value of the objective function related to the calibration experiment;

- *x* is the set of parameters value chosen by the algorithm at each functional evaluations;

- *D* is the domain of feasibility of the parameters, constrained by the upper and lower bounds and by the non-linear constraints (see Section 3.4.3);

Stringent termination criteria were set in order to try avoiding to get stuck in local minima. Here are the defined stopping rules:

- Maximum number of function evaluations allowed is 100,000;

- Maximum number of iterations allowed (i.e. the maximum number of non-stationary points that can be found) is 100,000;

- Termination tolerance on the function value is 1e-30;

- Termination tolerance on the parameters values is 1e-30.

Genetic Algorithm

Though the genetic algorithm allows for a non-linear constrained optimization, we applied the penalty function described in Eq. 3.15 (with the domain of feasibility *D* defined by the non-linear constraints only) to limit the computing time (for details, see Section 3.2.2).

As for the downhill simplex, stringent termination criteria were set in order to try avoiding to get stuck in local minima. Here are the defined stopping rules:

- Maximum number of generations (i.e. the maximum number of iterations allowed) is 100,000;

- Maximum number of stalling generations (i.e. with no improvements in the objective function) is 1,000;

- Cumulative change in the fitness function value over the maximum number of stalling generations is less than 1e-6.

Since it is not a global optimizer, the genetic algorithm could face difficulties in finding a stationary global solution. However, the genetic algorithm can sometimes overcome this deficiency with the right settings, such as increase in the population size (for details

see Section 3.2.2), but at the expensive of higher computational burden. Thus, a compromise was found, and the number of individuals in each generation was set equal to 20.

OptQuest Multistart

In the present work, we have used the OptQuest algorithm implemented in Lindo API 2.0 (LINDO, 2003).

As for the genetic algorithm, to improve the capability of the algorithm in finding the "known" global solution, a compromise was needed in the choice of the maximum allowed number of local searches. Indeed, a high number of maximum local searches increase the probability to find the global solution, but at the cost of increasing the number of model evaluations needed to reach convergence. To this aim, the maximum number of local searches was set to 20.

3.6.2 Tested measures of performance

Accordingly to the review of the literature, in this study we adopted both the speed and the inter-vehicle spacing as possible MoPs in the optimization problem specification.

3.6.3 Tested goodness of fit functions

It has been previously recognized (Ciuffo and Punzo, 2010) that the joint choice of the MoP and the functional structure of the objective function strongly influences the results. Indeed, the shape of the response surface associated to the specific optimization problem can vary considerably once we adopt different configuration of MoPs and GOFs. Thus, according to the basic idea of the experimental design, we tested a large number of objective functions in a setting with synthetic data, in order to understand their influence on the possibility to find the global solution.

In this view, the selection was made on the basis of the possible options reviewed in Section 3.2.2:

-        Root Mean Square Error (RMSE);

-        Mean Absolute Error (MAE);

- GEH Statistics with a threshold value equal to 1;

- Theil's Inequality Coefficient (U);

For details on the analytical formulations of the cited GOFs, please refer to Section 3.2.2.

3.6.4 Summary of the experiments

According to the presented setting, each *calibration experiment* was defined as an optimization problem given the solving heuristics and the response surface, which is univocally defined by the choice of the measure of performance and the functional form of the objective function.

Combining the 3 tested optimization algorithms (Downhill Simplex, Genetic Algorithm, OptQuest Multistart) and the 9 different response surfaces (RMSE(V), RMSE(S), MAE(V), MAE(S), GEH1(V), GEH1(S), U(V), U(S) and U(V)+U(S)), it resulted into 27 experiments. Moreover, each calibration experiment was solved *64 times* (i.e. 64 replications – here indicated as *calibration attempts*), in order to investigate the stability of the solution, thus resulting in a total number of 1728 calibration attempts.

3.7   Analysis of Calibration Results

In this section, the analysis of the results of the calibration experiments on synthetic data is presented.

Firstly, we were interested in assessing the ability of each problem setting (Algorithm/GOF function/MoP) in finding the "known" global solution. For a single calibration attempt, this can be measured in terms of the distance between the optimal solution found by the heuristic and the "known" global solution.

However, results from a single calibration attempt are not really informative on the uncertainty in the specific calibration process. In facts, calibration attempts differing in the starting point of the optimum search often provide different results. For this reason, multiple calibration attempts starting from different initial points were needed. This is even more the case of real trajectory data which often give flat and waved response

surfaces, with no "well-defined" global minimum but multiple local minima, each one potentially very far from the others.

The section is organized as follows. In Section 3.7.1, we first introduced four indicators that allowed us to compare the performances of the different calibration setting. Then, we evaluated them on the results of the calibration experiments. In Section 3.7.2, a novel representation of the map of the solutions found by the optimizer is presented. Finally, in Section 3.7.3 this graphic method is applied to explore the existence and the nature of local minima.

3.7.1 Proposed performance indicators

To evaluate performances of a specific problem setting we proposed and applied the following four indicators:

- The "Frequency of the original parameters", which measures the number of times, out of the 64 attempts of a calibration experiment, in which the optimization algorithm was able to rediscover the original parameters (i.e. the values which generated the synthetic global optimum) with an error on the single parameter of ±5%. This indicator reveals the ability of the specific calibration setting to find a solution in the close neighbourhood of the known global solution that is to rediscover the original parameters.

- The "Frequency of the best score" which measures the number of attempts in which the optimization algorithm attained its best score i.e. the lowest score of the objective function over the 64 attempts of a calibration experiment. Such solution, of course, is the best one provided by the specific calibration setting but does not necessarily coincide with the known global minimum. Therefore, the indicator measures the robustness of the specific calibration setting as to the variation of the starting point of the search (but not the ability to rediscover the global minimum).

- The "Optimization Performance Indicator" (OPI) given by Eq. 3.15b, evaluated at the best minimizer over the 64 calibration attempts, and labelled as OPI* (see Eq. 3.15a). Such indicator provides a measure of the accuracy of the best solution of a calibration experiment in terms of both the parameters values, and the score of the

objective function. It is a normalized indicator which therefore allows also different GOF functions and MoPs to be compared (e.g. RMSE(V) vs. GEH(S)).

$$OPI^* = \min_{i=1,\ldots,64}\left\{OPI\left(GOF_{alg,i}\right)\right\} \tag{3.15a}$$

where:

$$OPI\left(GOF_{alg,i}\right) = \sqrt{\sum_{j=1}^{m}\left(\frac{p_{j,i} - p_{j,Global}}{UB_j - LB_j}\right)^2} \cdot \exp\left(\frac{GOF_{alg,i}}{\max\limits_{alg=1,\ldots,3}\max\limits_{i=1,\ldots,64} GOF_{alg,i}}\right) \tag{3.15b}$$

with:

- $p_{j,i}$, equal to the value of the j-th parameter resulting from the i-th calibration attempt, with $j=1,\ldots, m$ and $i=1,\ldots, 64$.

- $p_{j,Global}$, the value of the j-th parameter at the known global optimum;

- LB_j and UB_j, respectively, the upper and lower bounds for parameter j;

- $GOF_{alg,i}$, the score of the specific objective function at the i-th attempt, for the alg-th algorithm, with $alg=1,\ldots 3$ (the nine GOFs are listed in Section 3.6.4).

The OPI measures the Euclidean distance between the values of the parameters resulting from a calibration attempt and those corresponding to the known global optimum (each term under the square root is normalized over the corresponding interval). In order to penalize a parameter set "near" to the optimal one which gives, however, a high score of the objective function, such distance is exponentially weighted with the score of the objective function found in that calibration attempt (normalized against the best value found among all the experiments which share the same GOF function and the same MoP).

In facts, the philosophy behind the OPI is that, in such type of investigation, one is mainly interested in understanding whether the problem setting allows the true parameters values to be rediscovered, rather than to see if the algorithm is able to achieve low scores of the objective function. Low values of the objective function

indeed can be sometimes obtained also with parameter values really distant from the optimal ones (when dealing with actual data this can be the symptom of model overfitting).

- The Total OPI, which gives a synthetic measure of the performance of a whole calibration setting, combining the information on the dispersion in the space of the cloud of the solutions found with the corresponding scores of the objective function. It is therefore a global indicator which measures the robustness of the calibration setting as to the variation of the initial search point, in terms of the accuracy of the solution found. It is given by Eq. 3.16:

$$Total\ OPI = \sum_{i=1}^{64} \left\{ OPI \left(GOF_{alg,i} \right) \right\} \tag{3.16}$$

3.7.2 Results from the performance indicators

The results are presented in Table 3.3.

The following considerations can be made:

1. The analytical formulation of the GEH Statistics, even with a strict threshold value set at 1 (in the place of 5, which is considered to be a good match between the observed and the model simulated outputs; Ma et al., 2007), does not allow any algorithm to ever find the global solution, that is to rediscover the original value of the parameters used to generate the synthetic data. All the algorithms converge (more or less frequently) to different points which share a zero value of the objective function, but which differ from the known global minimizer. This is told by the high values of the "frequency of the best score" and by the null percentages in the "frequency of the original parameters". This is also confirmed by the OPI scores which are multi valued (as each one of the 64 solutions returns a zero value of the objective function but different parameters values) and by the Total OPI that is higher than in other settings, especially in the case of the most performing algorithm.

Table 3.3: Analysis of the performances of each calibration procedure.

ALGORITHM	GOF/MoP	Frequency of the original parameters ± 5% error (%)	Frequency of the best score (%)	OPI*	Total OPI
Downhill Simplex with penalty function	RMSE(V)	0	2	9.01E-02	42.21
	RMSE(S)	0	2	5.54E-01	51.16
	MAE(V)	0	2	3.18E-01	49.62
	MAE(S)	0	2	3.57E-01	52.00
	GEH1(V)	0	14	More than one value	57.29
	GEH1(S)	0	6	More than one value	55.38
	U(V)	0	2	1.73E-01	41.55
	U(S)	0	2	4.99E-01	52.23
	U(V)+U(S)	0	2	3.63E-01	49.07
Genetic Algorithm with penalty function	RMSE(V)	94	2	1.41E-04	1.56
	RMSE(S)	25	2	1.95E-03	14.57
	MAE(V)	95	9	1.41E-05	0.96
	MAE(S)	17	2	3.55E-05	16.47
	GEH1(V)	0	100	More than one value	35.87
	GEH1(S)	0	73	More than one value	40.60
	U(V)	98	25	1.38E-04	0.91
	U(S)	30	3	2.59E-04	13.13
	U(V)+U(S)	41	2	1.05E-03	10.23
OptQuest Multistart	RMSE(V)	75	75	2.56E-05	5.62
	RMSE(S)	34	34	9.92E-04	17.36
	MAE(V)	58	58	5.66E-05	11.80
	MAE(S)	44	2	1.38E-03	16.44
	GEH1(V)	0	100	More than one value	33.87
	GEH1(S)	0	61	More than one value	31.37
	U(V)	58	58	2.56E-05	9.07
	U(S)	25	25	9.92E-04	19.59
	U(V)+U(S)	23	23	5.11E-04	18.82

2. The Downhill Simplex algorithm is never able to find the global solution in any of the problem settings. Further, the algorithm is not even robust as to the starting conditions, since it converges in almost all the replications to different optimal solutions while the best score of the algorithm is obtained only twice (except in the case of the GEH for the reasons above). This is because the algorithm gets always stuck in local minima. It also gave the highest values of the Total OPI among all the algorithms.

3. Both the Genetic algorithm and the OptQuest Multistart are able to rediscover the "true" values of the parameters with high frequency. The Genetic Algorithm resulted in the lowest values for the OPI and the Total OPI indicators, among all the tested settings (with the exception of the GEH case), though the stochastic nature of the heuristic mostly influenced the repeatability of the best score. On the other hand, the OptQuest Multistart showed a strong independence from the initial condition, converging repeatedly to the same global solution with the highest frequency.

3.7.3 *Proposed graphical inspection method*

We adopted a graphical representation of the map of the solutions of each different calibration experiment. The so-called *Cobweb* plots were used for this purpose (e.g. see Figure 3.4). Basically, they are line charts that display information as a series of data points (vertexes) connected by straight line segments. Unlike the time series, the horizontal axis is made of different categories and the vertexes of the plotted line are the values associated to each category. Since the range of values associated to different categories (for example, to the model parameters) can be wide and different, a normalization of those values is required, limiting the range of variability between 0 and 1 for each category.

The *Cobweb* plots were constructed as follows. The categories were:

-   The number of evaluations of the objective function when the stopping conditions were reached (*Nr_of_Iter*);

-   The value of the specific function adopted to compare the solutions of the different calibration experiments (*Validation_Score*). The chosen validation function was

the sum of the Theil's Inequality coefficients related to speed and spacing (U(V)+U(S)), and its value was computed for each calibration attempt;

- The optimal value of the objective function resulting from the current calibration attempt (*Obj_Funct*);

- The values of the calibrated model parameters (Tau=τ, Max_Vel=V_n, Max_Acc= a_n, Safe_Dist= *Safety*, Est_Dec=$\hat{b}_{n-1}$, Max_Dec=b_n);

In order to compare the results of all the calibration experiments with the same optimization algorithm the *Validation_Score* was used. Therefore, the values of this category were normalized between the minimum and the maximum *Validation_Score* among all the calibrations attempts made with a specific optimizer. Further, to give visual information of the best overall solutions (i.e. those associated to the minimum *Validation_Score*, as defined before) among all these calibrations, a colour bar was added.

Regarding the number of evaluations of the objective function, they were normalized between 1 and the maximum number among all the calibrations with a specific optimizer (i.e. 64x9 calibrations).

Conversely, the optimal value of the objective function, resulting from a single replication of the same calibration experiment, was normalized between 0 and the maximum value among the results of all the 64 replications with a specific GOF and optimization algorithm. Moreover, in the Cobweb plot, the bold line is associated to the results of the replication where the objective function was the minimum.

Concerning the optimal values of the parameters resulting from each replication (i.e. independently from the optimization algorithm and the objective function), they were normalized between the lower and upper bounds of the parameters.

It is worth noting that, according to the normalization methodology used, when the algorithm finds the "known" global solution of the optimization problem, both the validation score and the objective function value are equal to 0.

3.7.4 Results from the graphical inspection method

In the following, the *Cobweb* plots concerning some of the calibration experiments performed are presented. For a complete report, please refer to Ciuffo et al. (2012a).

First, the results of the calibration experiments using the GEH statistics with a threshold value of 1 are presented (Figure 3.4). Then, we showed the results of the calibrations with the different algorithms to compare their attitude towards globality (Figures 3.5 – 3.7). Finally, we reported on some minor findings on the use of the different objective functions in the calibration procedure (Figures 3.8 and 3.9).

Insights into the GEH statistics: the threshold value

Without loss of generality, in Figure 3.4 we showed the *Cobweb* plots concerning the 64 calibration attempts using the OptQuest Multistart algorithm. In facts, similar results were obtained with the other optimization algorithms (Ciuffo et al., 2012a). Figure 3.4(a) relates to the calibration on speeds, while (b) on spacing. Moreover, we drew the speed profiles of the leader (blue line), of the (synthetic) follower (red line) and, for each of the resulting set of calibrated parameters (i.e. 64 sets of parameters), of the (model) simulated follower (green lines). In addition, we drew the real (black line) and the simulated (green lines) spacing profiles.

From the figure, we can see that in none of the 64 attempts the algorithm was able to find the global minimum. Indeed, even if the algorithm is able to find the zero of the objective function, the validation scores are not equal to zero (which would mean that the global minimum has been found), and they are also very variable. The explanation of these outcomes relies on the functional form of the objective function, which requires the setting of a correct threshold value.

Specifically, we may recall here the analytical formulation reviewed in Section 3.2.2:

$$GEH_{Threshold\ Value}(Y) = \frac{N - \sum_{i=1}^{N} \delta_i}{N}$$

$$\delta_i = \begin{cases} 1 & \text{if } GEH_i(Y) \leq Threshold\ value \\ 0 & \text{elsewhere} \end{cases} \tag{3.6}$$

$$GEH_i(Y) = \sqrt{\frac{2 \cdot \left(Y_i^{obs} - Y_i^{sim}\right)^2}{\left(Y_i^{obs} + Y_i^{sim}\right)}}$$

(a)

(b)
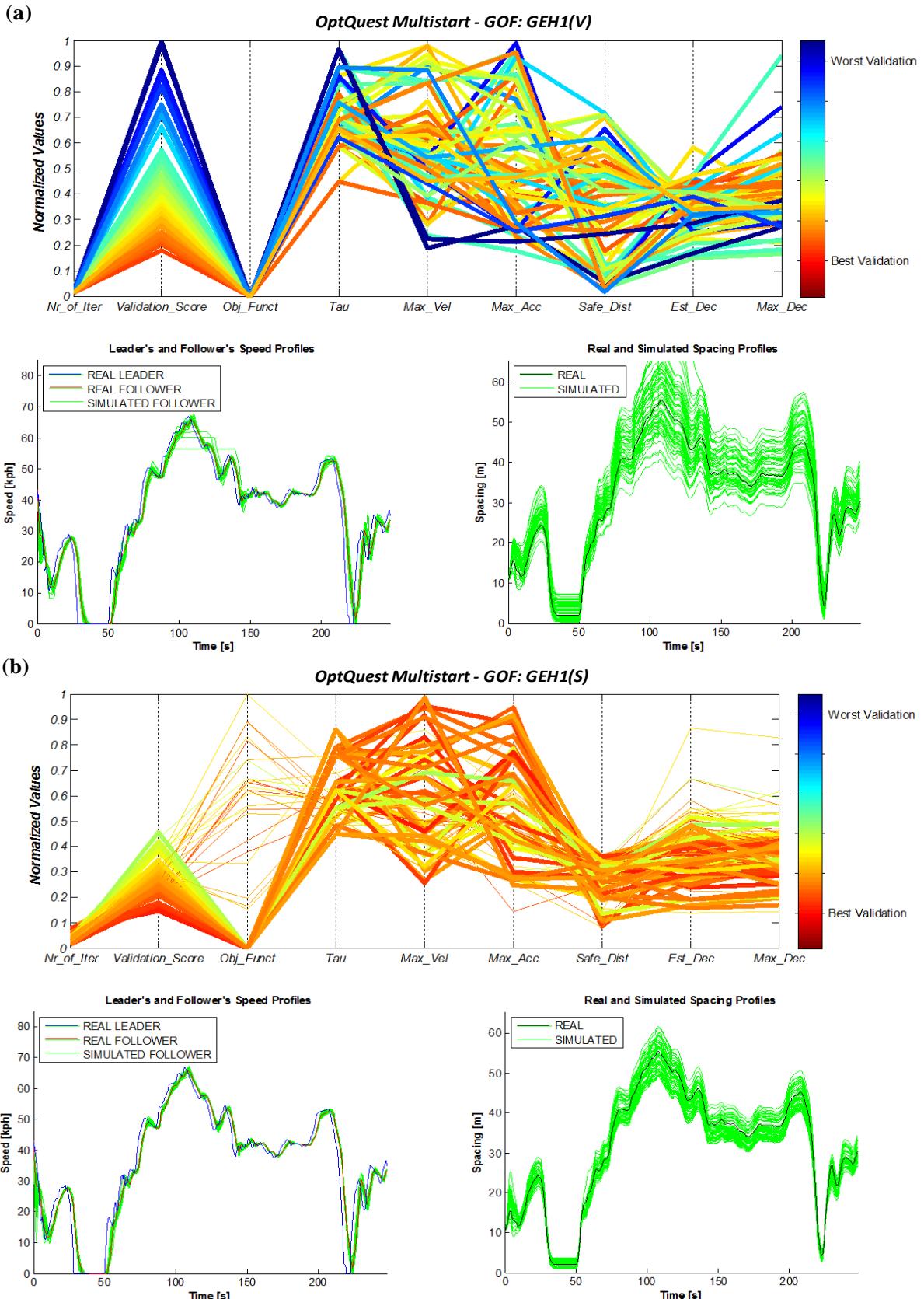
Figure 3.4: *Cobweb plots*, together with speed and spacing profiles, related to the calibration experiment using the OptQuest Multistart optimization algorithm and the GEH Statistics with a threshold value of 1.

It is worth noting that it slightly differs from the common formulation that can be found in the literature, when it is used for calibration purposes. In Ma et al. (2007), for example, the objective function is the sum of the GEH statistics ($GEH_i(Y)$) computed for each pair of observed and simulated outputs. However, such a formulation does not preserve the initial idea beyond the GEH statistics. Indeed, it was adopted to compare sets of traffic volumes (one observed, while the other simulated) and a good match between them was considered acceptable when the GEH value was less than 5 in the 85% of the observations. Thus, the formulation presented in Eq. 3.6 reflects this concept, that is to minimize the number of observations where the GEH statistics was above the threshold.

According to this, when we compare the real and the (model) simulated outputs, respectively Y_i^{obs} and Y_i^{sim}, at each simulation step (i.e. step i, with $i = 1,\dots, N$, given the number of observations N), the GEH statistics does not take into account the actual measure of distance between the observed and the simulated outputs ($GEH_i(Y)$), but only its being above or beneath a fixed threshold value (δ_i).Thus, it implies that the uniqueness of the global solution of the optimization problem cannot be preserved.

Thus, the main challenge is to set the optimal value for the threshold.

In facts, if one knew the actual level of approximation of the model to the reality, the threshold value could be set appropriately. Unfortunately, it is data dependent. In the context of synthetic data, for example, the model is able to reproduce exactly the "known" reality and, thus, a threshold value of 0 (i.e. δ_i equals to 1 only if $Y_i^{sim} = Y_i^{obs}$) would guarantee that, whenever the objective function is zero, the unique global minimizer has been found. On the contrary, moving to real data, the effective capability of the model in reproducing the world is unknown and, thus, setting a threshold value of 0 would not allow the algorithm to find the zero of the objective function.
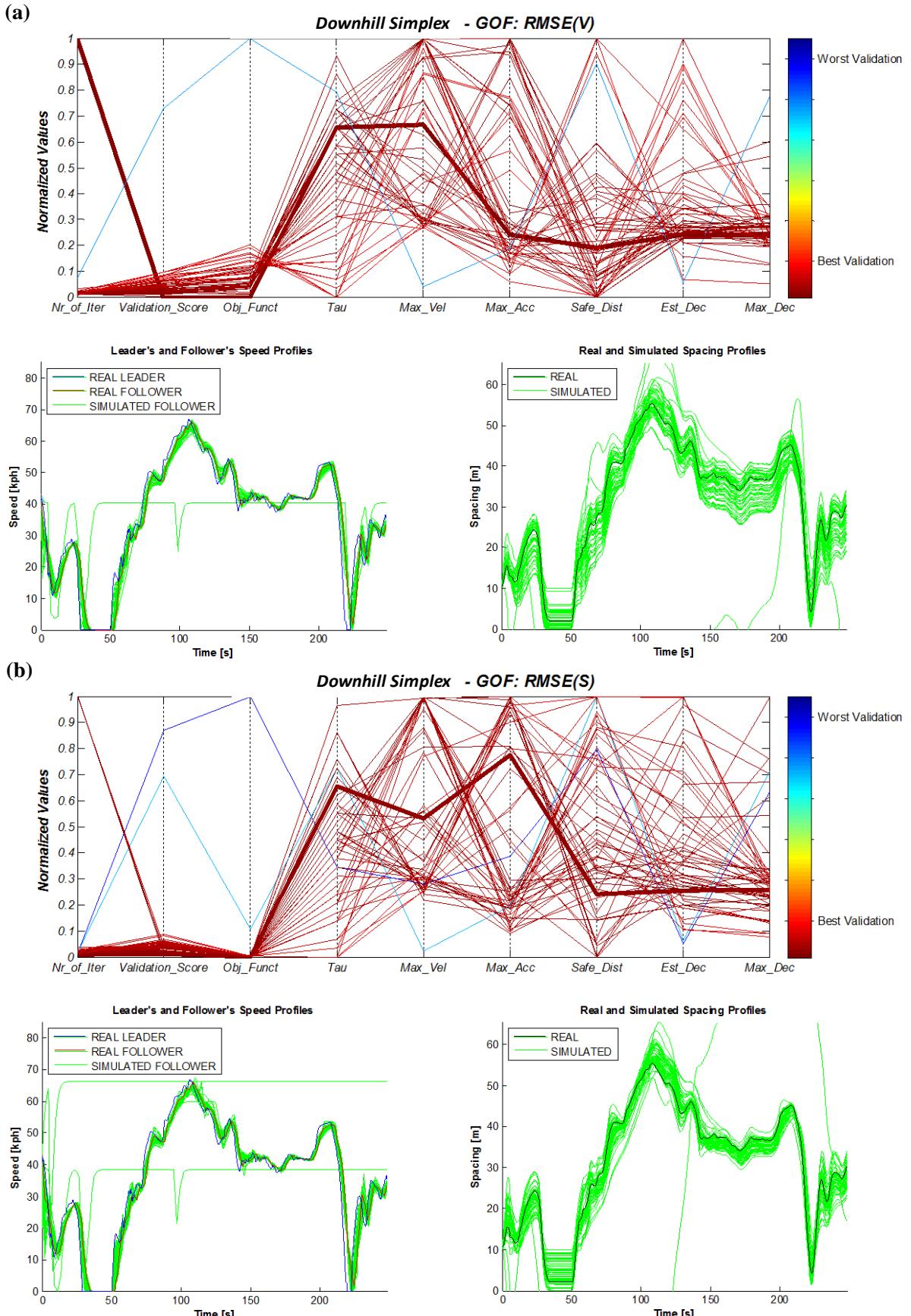
Comparison among the algorithms

Figures 3.5 – 3.7 show the *Cobweb* plots, together with speed and spacing profiles, related to the calibration experiments with the Downhill Simplex, the Genetic Algorithm and the OptQuest Multistart, when the RMSE was used as GOF function. We adopted this GOF function to compare the algorithms since, in our findings, it was the one that performed the best with all the heuristics (see Table 3.3), but the following considerations can be extended also to other GOF functions (Ciuffo et al., 2012a).

Comparing the three algorithms, it emerges that the Downhill Simplex is highly affected by the initial starting points, as the algorithm stopped in different points of local minima (see Figure 3.5).

Further, the setting of a penalty function, in the place of the parameter bounds and of the non-linear constraints, influences the algorithm even in finding the local solution. Indeed, in 5 out of 64 replications (both when calibrating on speed or on spacing), it stopped at the penalty value itself without actually performing optimization at all (see the blue lines in the *Cobweb* plots of Figure 3.5). In terms of simulation, it means that the resulting optimal set of parameters does not preserve the consistency of the *speed-headway* function and produces a model crash.

Regarding the Genetic Algorithm and the OptQuest Multistart, they are both able to rediscover the "true" value of the parameters, at least once. Moreover, even when they get stuck in local minima, they always find the "known" values of the most sensitive parameters (see e.g. Est_Dec and Max_Dec in Figures 3 and 4), where "sensitive" is intended in the framework of a global sensitivity analysis (for details, see Chapter 5).

Further, the two algorithms confirmed the conjecture, proposed in Punzo and Simonelli (2005), that calibrating the model against the spacing between the leader and the follower gives acceptable results also in terms of the vehicle speed, while the opposite is not equally true (see Figures 3.6 and 3.7).

(a)

(b)

Figure 3.5: *Cobweb plots*, together with speed and spacing profiles, related to the calibration experiment using Downhill Simplex algorithm and the RMSE as the GOF function.
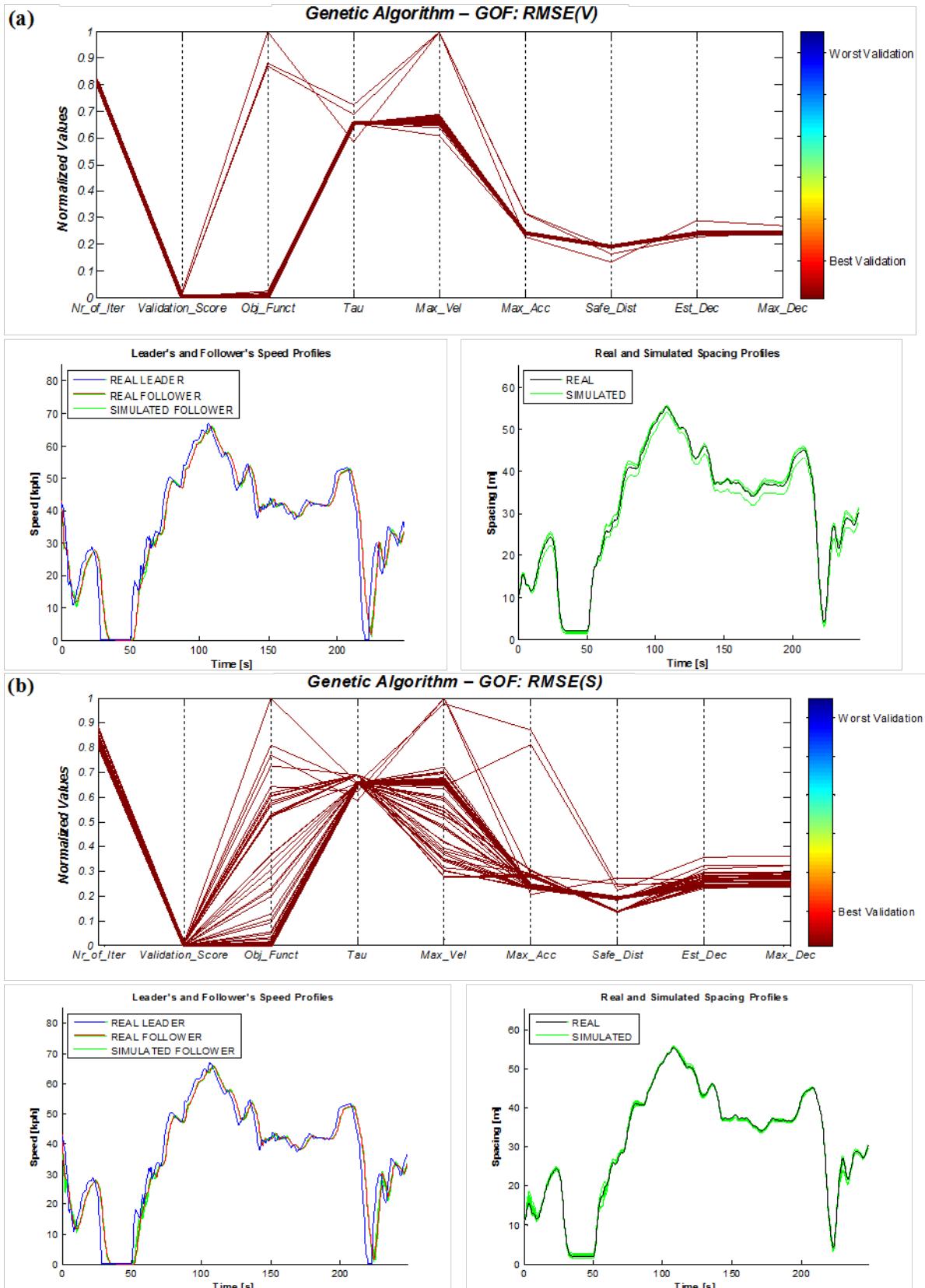
Figure 3.6: *Cobweb plots*, together with speed and spacing profiles, related to the calibration experiment using the Genetic Algorithm for the optimization and the RMSE as the GOF function.
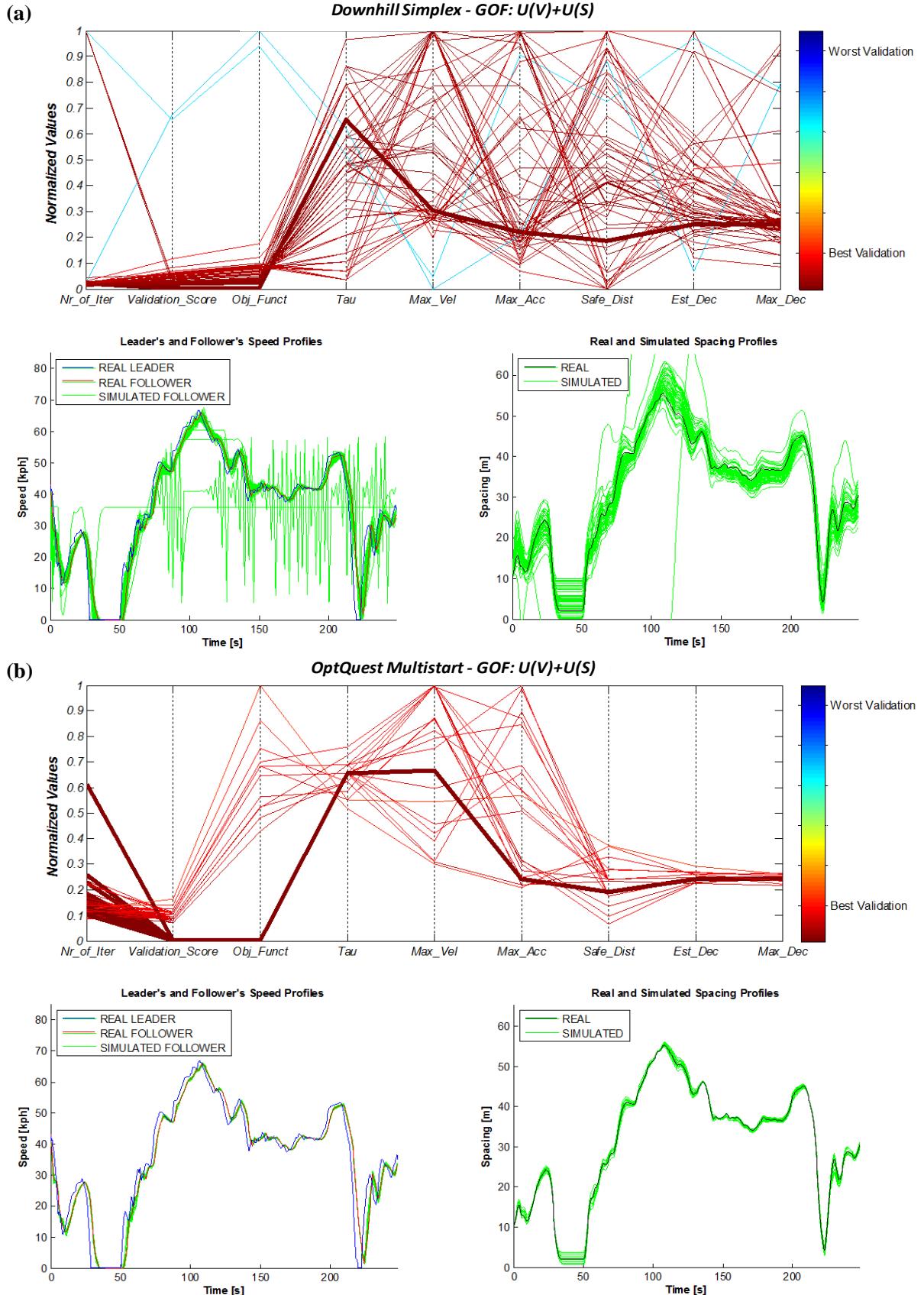
(a)

(b)
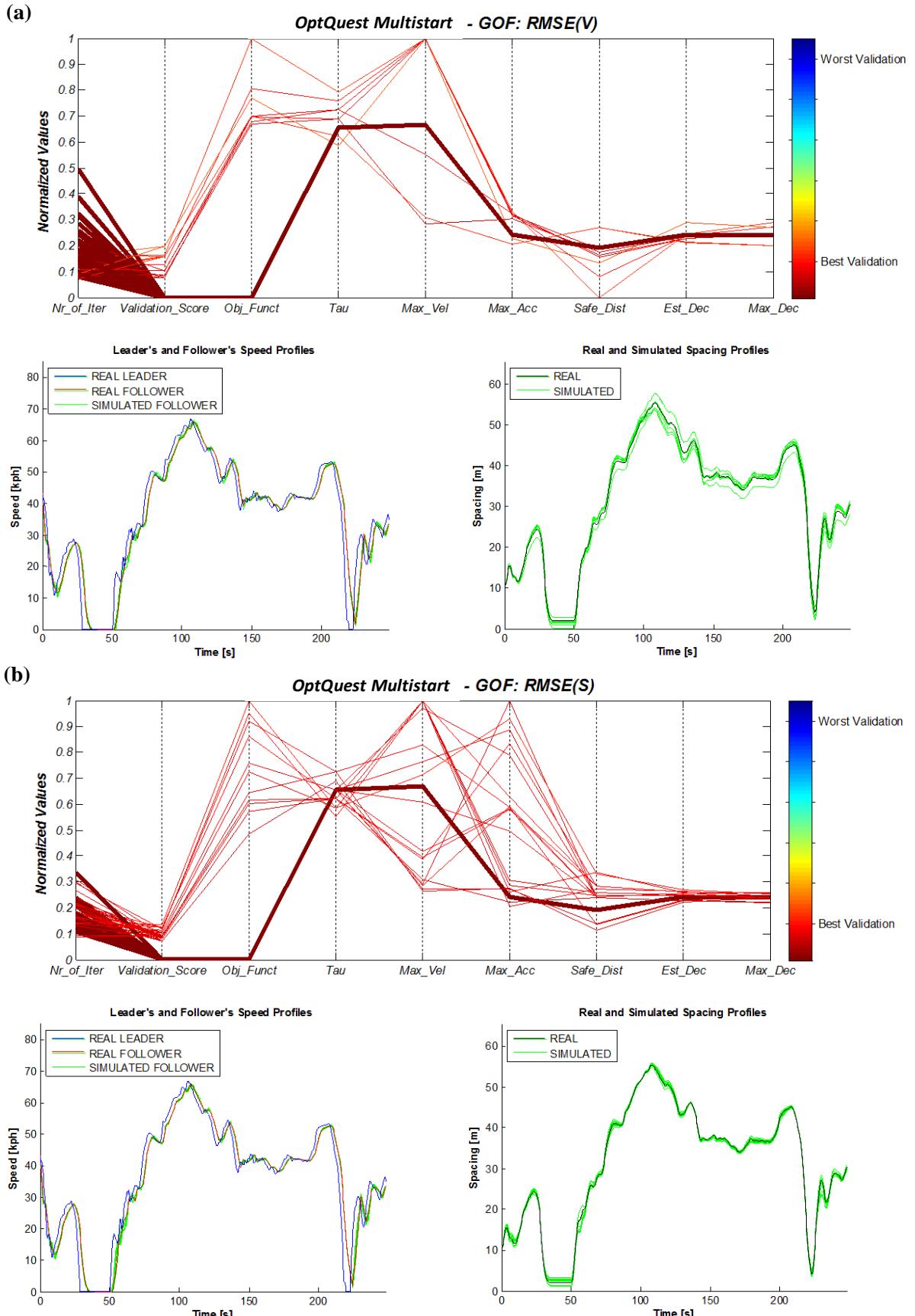
Figure 3.7: *Cobweb plots*, together with speed and spacing profiles, related to the calibration experiment using the OptQuest Multistart algorithm and the RMSE as the GOF function.

Minor findings on the GOF functions

In Figure 3.8, the *Cobweb* plots related to the calibration experiments where the sum of Theil's Inequality coefficients was used as GOF function. The plots refers to the calibrations with the Downhill Simplex and the OptQuest Multistart. In both the cases, it appears that the use of a combined function that takes into account both speed and spacing is not effective as those that perform optimization separately on speed or on spacing (Figures 3.6 and 3.7; see also Table 3.3).

Finally, Figure 3.9 shows the results of the calibration experiments when the MAE was used as the GOF function (for further details, please refer to Ciuffo et al., 2012a). It was found that this objective function (both when calibrating on speed or on spacing) is by far the least efficient in the optimization process, since it requires a very high number of objective function evaluations, while the improvement in finding the global minimizer is negligible.

(a)

(b)

Figure 3.8: *Cobweb plots*, together with speed and spacing profiles, related to the calibration experiments using the Downhill Simplex (a) and the OptQuest Multistart (b). The GOF function is the sum of the Theil's Inequality coefficients.

Figure 3.9: *Cobweb plots*, together with speed and spacing profiles, related to the calibration experiment using the OptQuest Multistart algorithm and the MAE as the GOF function.
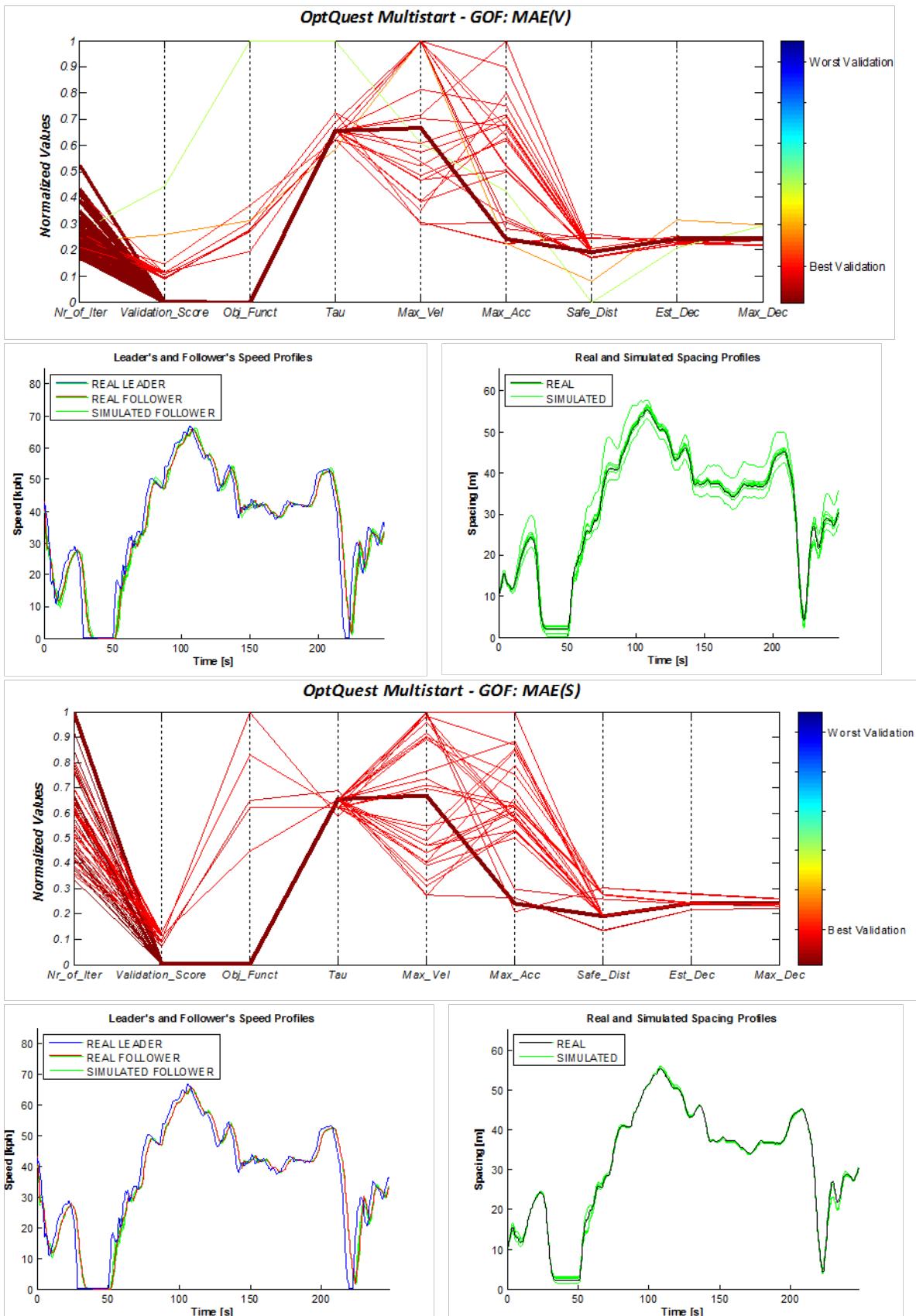
3.8   Summary

In the field of car-following models and, more in general, of traffic simulation, there is a lack of general and established methods to verify a calibration procedure and quantify the entailed uncertainty. As a result, the suitability of a particular calibration setting – here intended as the combination of the optimisation algorithm, the goodness of fit function and the measure of performance – as well as the reliability of the corresponding results, are unknown. As the calibration is deemed necessary to fruitfully apply these models, because acknowledged as the only way to deal with the inaccuracy of models and the uncertainty in the system, quantitative methods to assess any calibration setting are claimed to be necessary before one can apply any calibration method.

In this Chapter, a general methodology was proposed and applied to the calibration of the Gipps' car-following model. The methodology was based on the use of synthetic trajectory data, as this is the only way to ascertain the ability of a calibration setting to discover the global optimum. Compact indicators were proposed to evaluate the capability of a calibration setting to find the "known" global solution, in terms of both the accuracy and the robustness as to the variation of the starting conditions of the optimization algorithm. Then, a novel graphic inspection method, based on the so-called Cobweb plots, was used to explore the existence and the nature of the local minima found by the algorithms, as well as to give insights into the measures of performance and goodness of fit functions used in the calibration experiments.

Such methodology has been applied to all the calibration settings used so far in the literature to calibrate car-following models. Though any comparison involving optimization algorithms can never be fair or definitive as it always depends on the particular algorithm setting adopted, the present analysis allowed us to emphasize some specific relevant behavior. In particular, the main outcomes of the study were the following:

- GOF functions based on the GEH statistics are highly affected by the setting of the threshold value. When used in calibration, a wrong setting of this value lead to the loss of uniqueness of the global solution, even in the case of optimization problems on synthetic data, where the global minimizer is unique and well-defined.

- The Downhill Simplex has not been able to rediscover the true set of the parameters' values in none of the experiments performed. Further, the heuristics was very sensible to the initial starting condition, providing very different sets of optimal parameters depending on the starting point.

- Both the Genetic Algorithm and the OptQuest Multistart were able to find the "known" global minimizer, at least once on 64 replications of the same calibration experiment. Moreover, they were able to rediscover the true value of the most sensible parameters in almost all the replications.

- These two algorithms confirmed the conjecture, proposed in Punzo and Simonelli (2005), that calibrating the model against the spacing between the leader and the follower gives acceptable results also in terms of the vehicle speed, while the opposite is not equally true.

- The use of mixed GOF functions that combine both the MoPs (speed and spacing), such as the sum of Theil's Inequality coefficients, performed worse than when calibrating separately on speed or on spacing. Further, the use of absolute measures of the distance between observed and (model) simulated outputs, such as the MAE, entails very low efficiency in the optimization, as they require a high number of evaluations of the objective functions to satisfy the same stopping rules adopted with the other GOF functions. Moreover, the improvements in finding the global minimizer are negligible.

As a general conclusion the present study confirmed the complexity of the problem of calibrating car-following models against real trajectory data. As a matter of facts, none of the tested settings gave completely satisfactory results, and future research shall necessary aim at finding more robust settings.

Therefore, in the light of the previous findings, the following research lines can be drawn for future investigations:

1. to limit the calibration process to the most sensitive parameters, via e.g. sensitivity analysis of model outputs, in order to reduce both the number of dimensions of the input space and the flatness of the response surface. This would drastically decrease the computational complexity of the optimization problem.

2.   to seek for "global" GOFs which were able to capture the inner structure/driving behaviour/driving style contained in the trajectory data, as expressed/interpreted by the specific model in use. This is also in the course of the recent studies performed by Chiabaut et al. (2010). Local GOF indeed are sensitive to errors in the data, and especially least square ones tend to compensate errors over the whole length of the trajectory.

3.   to appropriately bound the space of the admissible inputs in order to preserve the well established macroscopic characteristics of the traffic flow.

The first two points are discussed in Chapter 5 and in Appendix C, respectively.

Previous points would contribute at the end to address the problem of model overfitting – mostly relevant for car-following models given their manifest inadequacy – and to increase the transferability of calibration results. A contribution on this topic is given in Chapter 6.

Chapter 4

Uncertainty in Vehicle Trajectory Data and Impacts on Model Estimation[1]

4.1   Introduction

Availability of all the vehicle trajectories in a traffic stream is the Holy Grail of traffic flow theory. Since the publication on the internet of the first and unique publicly available database of vehicle trajectories, i.e. the datasets from the FHWA's Next Generation SIMulation Program (NGSIM, 2014), plenty of researchers have made use of such data to interpret traffic phenomena, support theories, benchmark, calibrate and validate traffic flow models (see e.g. Chiabaut et al., 2010; Kim and Mahmassani, 2011; Koutsopoulos and Farah, 2012; Laval and Leclercq, 2008, 2010; and so on).

Within the field research community, however, an increasing concern is taking off about the accuracy of such data and its potential impact on the results of studies applying them. Recently, Punzo et al. (2011b) proposed a method to inspect the accuracy of trajectory data and applied it to all the datasets of the NGSIM Program. They focused, in particular, on quantifying the *internal* consistency of data – that is the consistency among space travelled, speeds and accelerations – and the *platoon* consistency which refers to the physical consistency of the inter-vehicle spacing as resulting from the

---

[1] Regarding the contents of this Chapter, the reader can refer also to Montanino and Punzo (2013).

individual trajectories of a pair of leader-follower vehicles. Study results supported previous concerns of the scientific community, showing low degree of accuracy for both the criteria. In reason of the big errors and inconsistencies in the NGSIM speeds and accelerations[2], therefore, Punzo et al. suggested not to use such quantities but to estimate them directly from the space travelled[3] after an appropriate filtering of such data in light of the two consistency criteria proposed.

Despite of the problem relevance, however, relatively few studies exist on the subject of vehicle trajectory data correction (for a review and classification of techniques see Punzo et al., 2011b). In addition, it is argued herein that none of the techniques proposed and applied so far in the literature is suitable to reconstruct effectively vehicle trajectories from the NGSIM data. This is because none of such techniques is able to treat effectively those extremely biased values, often present in such data, to which we referred to as "outliers". They limited to smoothing out the noise, indeed, by removing the high-frequency and, in part, the medium-frequency disturbances from the data. This is explained hereafter where the mechanism at the basis of the errors in the NGSIM data is clarified together with the limits of currently available techniques.

In addition, very few studies attempted to quantify the impact of measurement errors on model estimation. To the best of our knowledge, the only contribution on this topic was given by Ossen and Hoogendoorn (2008a, 2009). In their studies, the authors evaluated the reliability of estimated car-following model parameters in presence of measurement errors in trajectory data, concluding that measurement errors can have a large influence on estimation results in terms of both median differences among estimates and robustness of estimate.

However, in these studies, position errors were synthetic, i.e. not obtained from data collection but added ex-post, and with white noise structures (independent and identically normal distributed random variables with zero mean). Conversely, real trajectory data present much more complex error structures, often with time-correlation properties, and locally distributed only in certain time-windows with peaks where observations are massively biased.

---

[2] Values in the NGSIM dataset fields "Velocity" and "Acceleration".

[3] Values in the NGSIM dataset field "LocalY".

Therefore, in this Chapter, the evaluation of the impact of *real* measurement errors (in vehicle trajectory data) on model parameter estimation results is provided.

To this aim, the availability of trajectory data with real measurement errors is crucial for the study purpose, as the reproducibility of reliable synthetic error structures is controversial. Therefore, in this study we relied on the NGSIM vehicle trajectory data whose low degree of accuracy in terms of several criteria is widely recognized in the transportation community.

Therefore, in the first part of the Chapter, a multi-step procedure for reconstructing vehicles' trajectories is presented. The proposed methodology aimed at eliminating the main inconsistencies and noise from *raw* measurements while preserving *i)* the actual driving dynamics (vehicle stoppages, shifting gears, etc.), *ii)* the *internal* consistency of trajectories (i.e. the consistency among space travelled, speed and acceleration) and *iii)* the platoon consistency (i.e. the actual inter-vehicle spacing).

In the second part of the Chapter, provided both *raw* and *reconstructed* trajectory data, we evaluated the impact of real measurement errors on estimation of car-following and lane-changing model parameters.

In this work, the reconstruction procedure has been applied to the NGSIM dataset from the northbound traffic on I80 in Emeryville, California (NGSIM, 2005), recorded from 4:00 p.m. to 4:15 p.m. on April 13, 2005 – in the following referred as *I80-1*. It is worthwhile mentioning, however, that the proposed technique is absolute general and could be used to filter trajectories from any other dataset.

The Chapter is organized as follows. Section 4.2 describes the mechanism at the basis of the errors in the NGSIM data and the limits of currently applied smoothing/averaging techniques. The proposed methodology is presented in Section 4.3, together with the discussion of the results from each step of the sequential reconstruction and a comparison with low-pass filtering techniques. Then, Section 4.4 discusses the peculiarities of the NGSIM trajectory data and the requirements for robust filtering. Results from the application to the NGSIM I80-1 dataset are illustrated in Section 4.5, in terms of both individual vehicle trajectories and acceleration distributions. Successively, Section 4.6 presents the comparison of the results of car-following model parameter calibration against *raw* and *reconstructed* data. Finally, the work ends with conclusions and recommendations for future research.

4.2   Error Generation in Trajectory Data

In this section we investigate the mechanism at the basis of the errors in the NGSIM data which is necessary to understand *i)* which are the most appropriate data to start the reconstruction with (i.e. spaces, speeds, accelerations) *ii)* why the usual techniques fail or, conversely, *iii)* which are the desirable features of a reconstruction method.

4.2.1 Errors in video-processed data

Figure 4.1 depicts a general situation which may arise after measuring positions of a vehicle at discrete times: measured positions (black full points) apparently follow an irregular path (zigzag), which is actually due to the measurement errors.
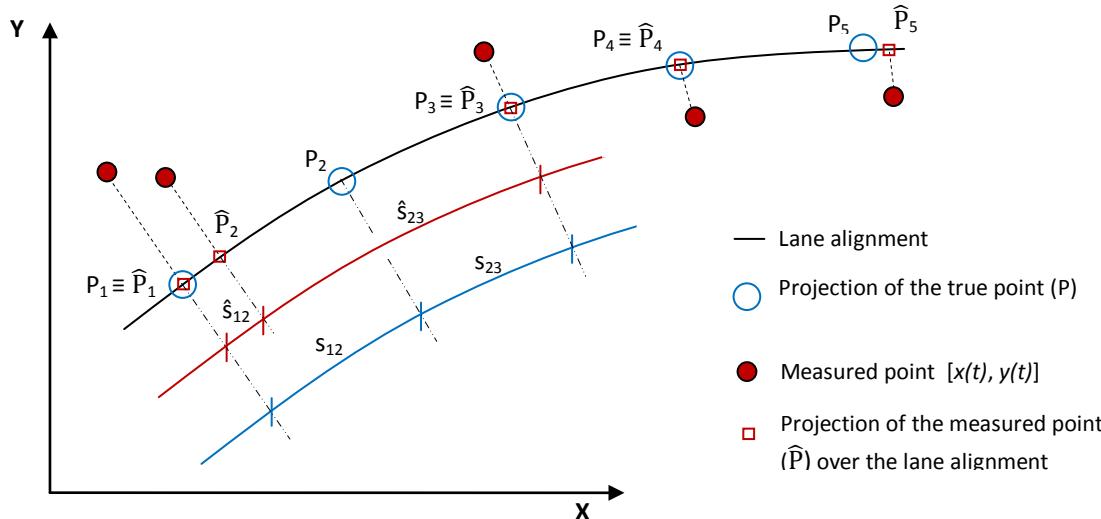

Figure 4.1: Insight in the nature of errors in the NGSIM data. Though the total space travelled measured between instants 1 and 3, $(\hat{s}_{12} + \hat{s}_{23})$, coincides with the true one $(s_{12} + s_{23})$, i.e. there is no bias in the cumulative space travelled, the error in the position of point 2 implies an error in the space travelled in the two intervals 1-2 and 2-3, which is amplified in the corresponding speeds and accelerations.

Punzo et al. (2011b), show that such scattering of points around the actual unknown path implies a bias between the actual space travelled and the one calculated from such measurements, howsoever one reconstructs a path among these points. To eliminate such

bias, however, it is possible to project the points over the road/lane alignment. In all the cases where it is acceptable to confuse the actual vehicle path with its projection on the lane alignment, as is the case of "pure" car-following studies indeed, the projection of the coordinates on the lane alignment is therefore the basic way to eliminate such bias in the space travelled (Punzo et al., 2011b). This is actually the way in which the values of NGSIM's "LocalY" (i.e. the longitudinal coordinate of the front centre of the vehicle with respect to the entry edge of the section in the direction of travel; NGSIM, 2005) have been derived: they are the projections over the lane alignment of the coordinates of the measured points (the ones recorded in the datasets as "GlobalX" and "GlobalY"). The truthfulness of such process has been verified numerically in this work.

Unfortunately, the projection on the lane alignment eliminates the bias in the space travelled, but not the noise on the same measure. In facts, it's easy to verify that the projected points $\hat{P}_i$ (the gray squares in Figure 4.1) are often positioned differently from the projections P_i of the "true" points (the black empty circles in the same figure). For example, the projection $\hat{P}_2$ happens to be nearer to $\hat{P}_1$ and farther from $\hat{P}_3$ than the actual point P_2 is, respectively, from P_1 and P_3. Though the total space travelled measured between the instants 1 and 3, $(\hat{s}_{12} + \hat{s}_{23})$ coincides with the true one $(s_{12} + s_{23})$ (i.e. there is no bias in the cumulative space travelled), the error in the position of point 2 implies an error in the spaces travelled in the two intervals. Given the high frequency of the measurements, i.e. 10 Hz., and the amplification occurring in the differentiation process, even errors of few centimeters in space produce significant errors in the speed and even more in the acceleration.

This strong effect is shown in Figure 4.2 with regards to a trajectory sample from the NGSIM I80-1 dataset. In Figure 4.2(a), the cumulative space travelled by the vehicles, as stored in the NGSIM "LocalY" field, is reported, while speeds in Figure 4.2(b) are calculated as the ratio between the distance travelled in 0.1 seconds (from the "LocalY") and the same interval. Alike, accelerations in Figure 4.2(c) are calculated as the variation of such mean speeds between two consecutive time intervals.

a)

b)

c)
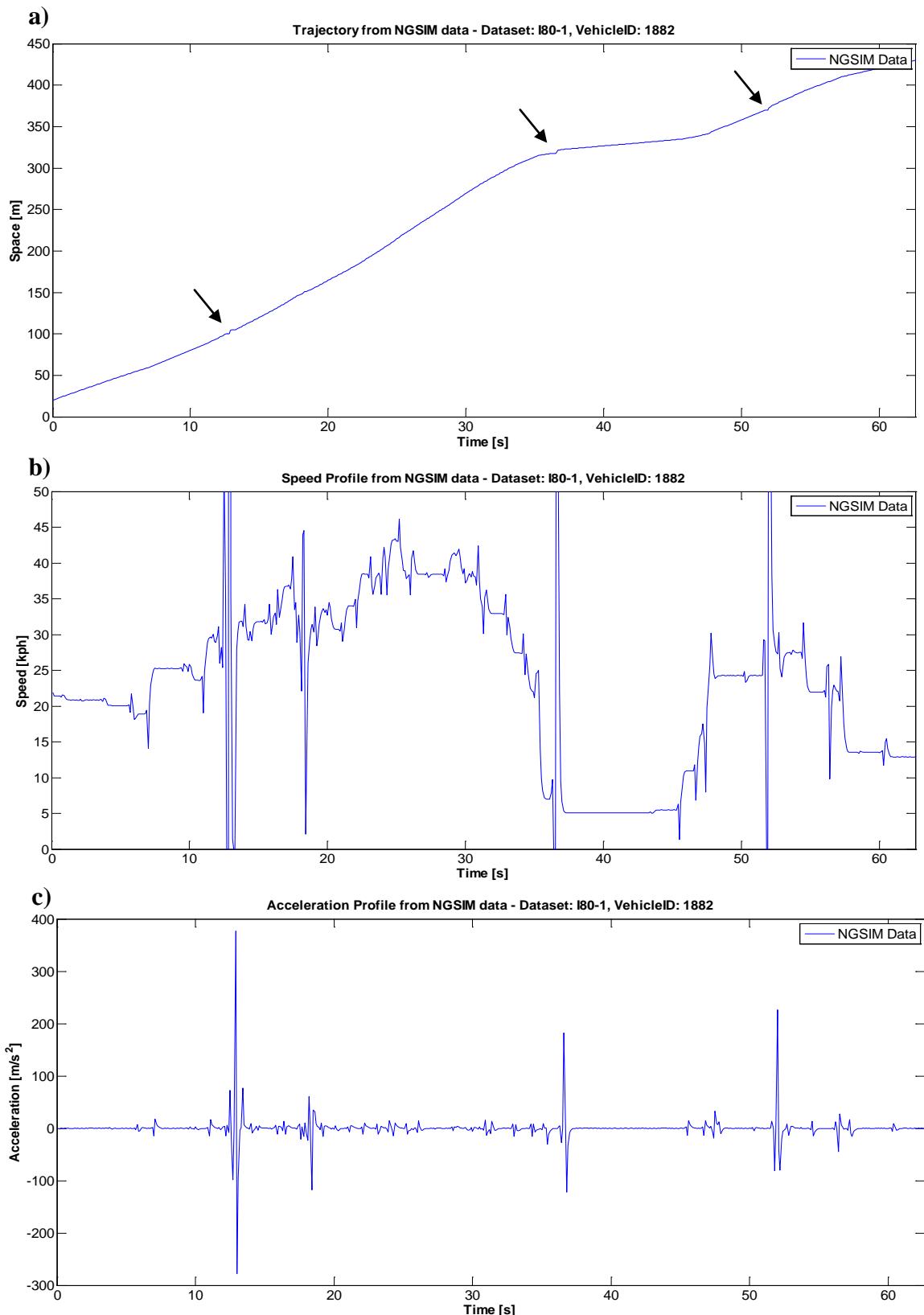
Figure 4.2: Cumulative space (a), Speed (b) and Acceleration (c) profiles of vehicle 1882 from dataset I80-1. Values from "Local Y" field are reported in (a). Speeds (b) are calculated from "LocalYs", as travelled distance in 0.1 seconds interval divided by the same interval. Alike, accelerations (c) are calculated from Speeds, as their variation in a 0.1 seconds interval.

The errors in the space travelled that appear locally in the trajectory (see the arrows in Figure 4.2(a)), produce the totally unreliable acceleration/deceleration rates shown in the bottom figure which reach peaks of almost 400 m/s^2. Apart from such unrealistic spikes, Figures 4.2(b-c) show a significant random disturbance affecting the measurements especially during speed transitions (e.g. shifting gears, sudden brakes/accelerations, etc.) which give rise to accelerations up to 40 m/s^2. Both these errors are extremely frequent in the NGSIM datasets and make such data unusable, without an appropriate treatment, for any study on traffic flow theory.

For the sake of simplicity in the following we define as *outliers* the measurement errors in the "LocalYs" that produce the greatest bias in the accelerations, and as *noise* the residual errors.

### 4.2.2 Desirable features for robust trajectory data filtering

Apart from the data in the LocalY field there are two other measurements available in the NGSIM datasets which point to the question of which is the most appropriate measure to start the trajectory reconstruction with. These other measures are the "Vehicle velocity" and the "Vehicle acceleration" which represent the vehicle instantaneous speed and acceleration, respectively. Such quantities, which have been estimated by the "LocalYs" with local regression techniques (see Punzo et al., 2011b, for a discussion) and successively further corrected, apparently resolve (only) the highest errors just highlighted for the "LocalY" (i.e. the outliers). However, in Punzo et al. (2011b) it is shown how this was obtained with non optimal practices like, e.g. simply cutting accelerations above/below a specific threshold or reversing the sign of negative speeds, which actually left unvaried the unrealistic and noisy patterns of such variables. Moreover, this also yielded speed and acceleration profiles macroscopically inconsistent with the space travelled (i.e. with the LocalY data) (see, again, Punzo et al., 2011b).

Therefore, especially considering that the original (highly noisy) pattern of speeds and accelerations has been so sharply corrupted, it does not really make sense to filter such data, but it is necessary to concentrate only on the reconstruction of the "LocalY" data. This means trying to reconstruct physically consistent mean speeds and accelerations from the LocalYs and, whenever needed, to calculate instantaneous values from these.

The examples of Figure 4.1 and 4.2 on the error nature in the "LocalY" data allow us to put forward also a second major consideration: the so-called *outliers* cannot be easily treated at the same time of the residual errors (i.e. the *noise*) and with the same techniques.

Indeed, the filtering techniques currently applied in the field literature are based on smoothing or averaging (like low-pass filters and moving average filters, based on Gaussian or symmetric exponential kernels, as in Hamdar and Mahmassani, 2008, and Thiemann et al., 2008, respectively). Independently from the variable to which they are applied (coordinates, spaces, speeds, accelerations), they essentially remove the noise from the data. Inevitably, the design of the response of such filters is the result of a compromise between the need of eliminating the noise, even at low frequencies (low cut-off frequency) and that of preserving the actual driving kinematics (high cut-off frequency). This means that the elimination of the highest peaks, if possible, is paid at the cost of smoothing too much the real speed and acceleration profiles. The following example clarifies the issue.

A low-pass filter (i.e. a Butterworth filter; Buttherworth, 1930), has been applied to the speed profile in Figure 4.2(b), and the resulting filtered profile is shown in Figure 4.3.Two cut-off frequencies (0.75 Hz and 0.25 Hz, respectively) were used to show the different results obtainable. When allowing a frequency response of 0.75 Hz in the pass-band that reduces the error noise at most still preserving the driving dynamics in the original data (i.e. vehicle stoppages, shifting gears during accelerations/decelerations, etc.), the so-called outliers are barely smoothed out (see the green line at around seconds 10, 40 and 50 in Figure 4.3(a)). In turn, if the cut-off frequency is set in order to obtain only accelerations in the range of physical values (i.e. to 0.25 Hz) very smoothed speed profile is obtained (see the red line in Figure 4.3(a)), but still the outliers are not completely removed (see, for example, the unrealistic behavior around second 40 in Figure 4.3(a)).

a)

b)
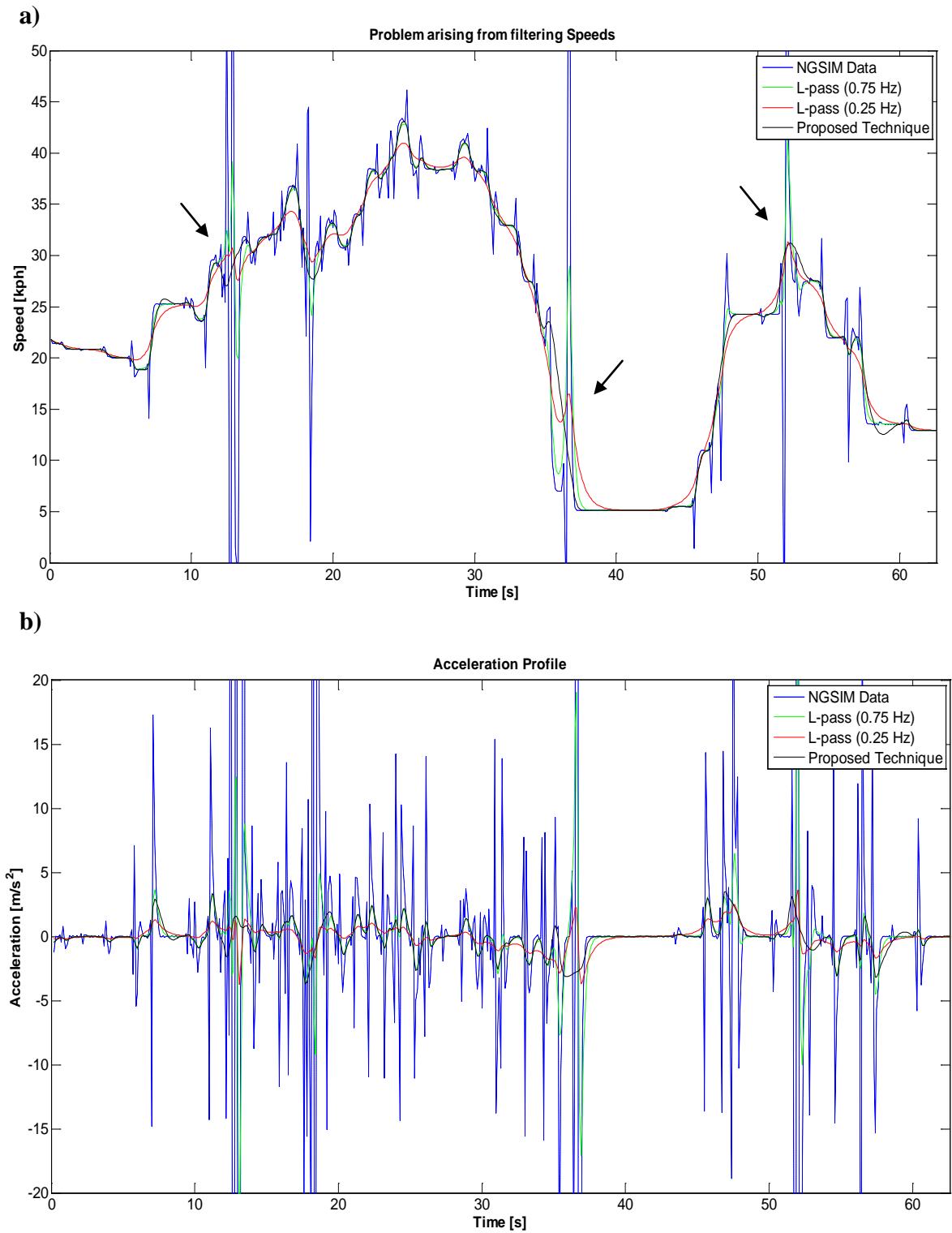
Figure 4.3: Speed (a) and Acceleration (b) profiles when Speeds are filtered with a Butterworth filter (L-pass) of order 1 and cut-off frequency of 0.75 Hz (green line) and 0.25 Hz (red line), and with the procedure here proposed (black line).

In order to solve this problem it is claimed that the highest errors (the so-called *outliers*) need to be fixed before filtering out the residual *noise*.

Moreover, traditional filtering techniques may alter the total space travelled by a vehicle – the so called *internal* consistency of the trajectory (for details, see Punzo et al., 2011b). Indeed, whatever filter is applied to speeds, it would inevitably modify the space travelled by the vehicle in each time interval, thus corrupting the total space travelled.

In addition, when individually filtering vehicle trajectories, also with internal consistency requirement properly taken into account, problems related to inter-vehicle spacing with leader and follower vehicles still may arise.

Therefore, in order to solve both these problems, in this work we proposed a multi-step filtering procedure to reconstruct vehicle trajectories by fixing the outliers, reducing the residual noise in the data, and preserving the *internal* and the *platoon* consistency requirements.

4.3   Multistep Vehicle Trajectory Reconstruction

In this section, the sequential multi-step procedure for vehicle trajectory reconstruction is presented. According to the requirements for robust trajectory data filtering, introduced in Section 4.2.2, the filtering procedure has the main goal of removing the unphysical values of accelerations, while preserving *i)* the *driving dynamics*, especially in acceleration and deceleration phases (e.g. . vehicle stoppages, shifting gears, etc.), *ii)* the total space travelled, i.e. the *internal consistency* of the trajectory, and *iii)* the inter-vehicle spacing between successive vehicles, i.e. the *platoon consistency* of the entire dataset.

With this aim, the filter operates on individual vehicle's positions ("LocalY", as suggested in Punzo et al., 2011b), modifying locally the vehicle position in time.

As the quality of each trajectory (e.g. the type of errors or the frequency component) may vary sensibly in a large dataset, different filter's parameters should be appropriately set. However, as we needed to apply the same procedure to sequentially filter all trajectories in the dataset, we needed to design a common filter whose parameters do not vary across individual trajectories.

Therefore, the procedure is organized in the following 4 steps, which require to *sequentially*:

1. remove the *outliers* (Section 4.3.1);

2. cut-off the high and medium frequency responses in the speed profile (Section 4.3.2);

3. remove the residual unphysical acceleration values, preserving the internal and platoon consistency requirements (Section 4.3.3);

4. cut-off the high and medium frequency responses eventually generated from the previous step (Section 4.3.4).

It is worth noting that the procedure is absolute general, and can be applied to whatever vehicle's trajectory dataset.

In the following, the description of each step is accompanied by figures showing the sequential gain reached during reconstruction, for a sample trajectory.

4.3.1 Step 1: Removing the outliers

This stage aims at removing the errors in the "LocalYs" measurements that give rise to the unreliable values of acceleration shown in Figure 4.2(c). This can be done by filtering locally the trajectory when the observed (absolute) accelerations exceed a certain threshold. Such threshold should be set appropriately high in order to capture only the big measurement errors (i.e. the outliers) and not the random disturbances that affect the observations. In this light, a threshold value of 30 m/s^2 was appropriate for the scope.

To achieve the objective, a filtering technique is needed to interpolate between the sequences of non-outliers that can be found respectively before and after consecutive big measurement errors. In this light, we applied a natural cubic spline interpolation using ten reference points (i.e. one second of observations) both before and after the outliers.

Figure 4.4 shows the speed (a) and acceleration (b) profiles before (the blue line – *raw* data) and after (the green line) this step. The arrows in Figure 4.4(a) indicate the outliers detected with a threshold value of 30 m/s^2. As a result, the local cubic spline interpolation on the "LocalYs" completely removed the four outliers.

a)

b)
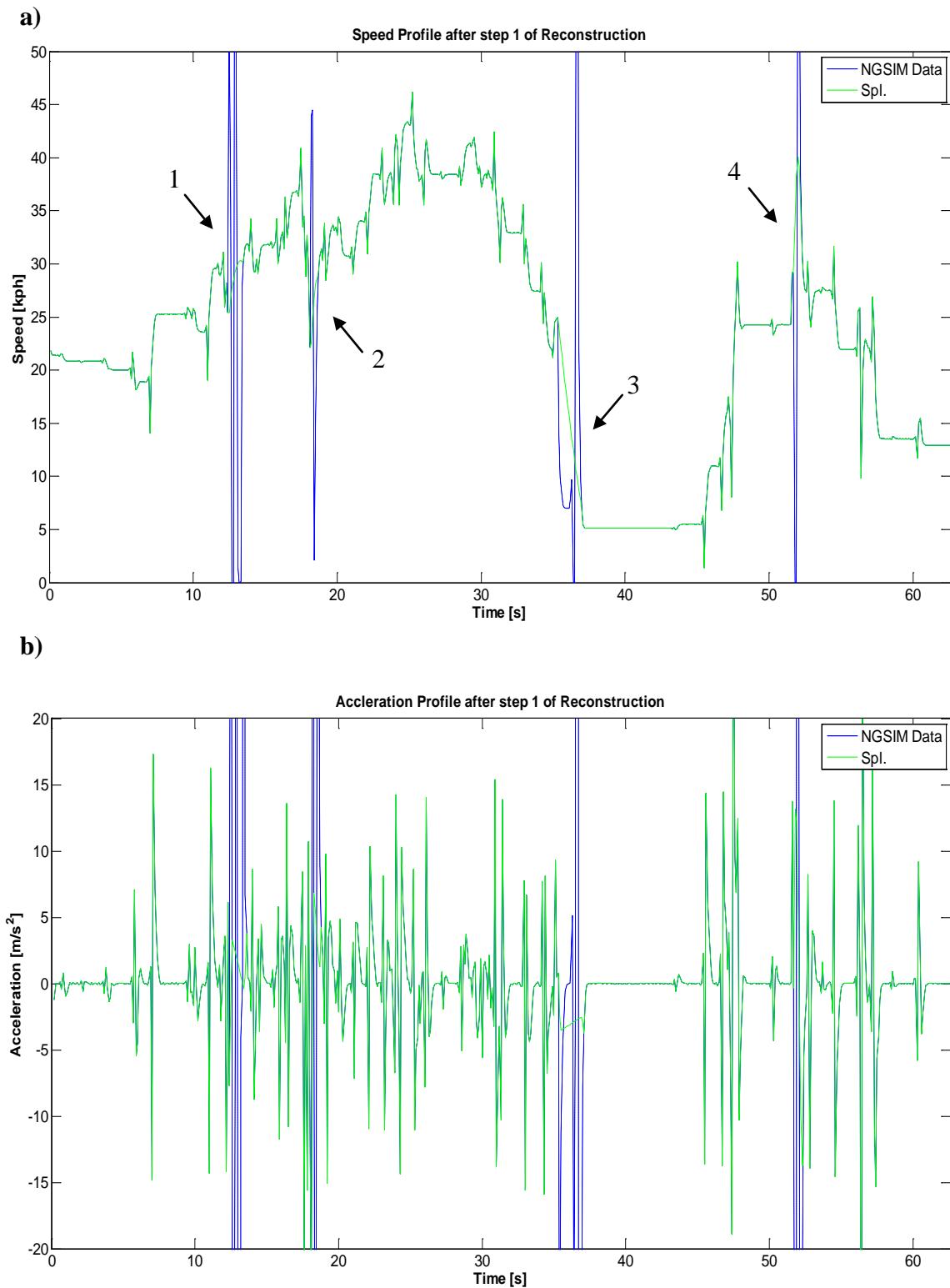
Figure 4.4: Speed (a) and Acceleration (b) profiles after step 1 (Spl.) of reconstruction. The resulting profiles (green line) differ from the original NGSIM data only in the time windows in which the spline interpolation was applied to remove the outliers (see the arrows).

To better appreciate the impact of removing the outliers, we invite the reader to skip to Figure 4.12 in Section 4.6.2, which shows the representation of the acceleration profile in the frequency domain. The analysis was limited to frequencies up to 5 Hz, according to the Nyquist–Shannon sampling theorem (Shannon, 1949). Comparing the original data (the blue line) with the results from the current step (the green line), we may see how removing outliers is equivalent to strongly attenuate (i.e. reduce the amplitude of) the signal, that is decreasing the peak amplitude of the oscillations in the acceleration profile.

However, this step does not guarantee that resulting accelerations fall in a range of physical values, since any constraint is applied to the interpolating curve, except for the passing condition at the reference points. Evidence on it can be found in Figure 4.4(b) focusing on the resulting values of the accelerations after the elimination of outliers 2 and 4.

4.3.2 Step 2: Cutting-off high and medium frequency responses in the speed profile

The objective of the current step is to remove the noise (i.e. the random error component) from the signal. This can be done by linear smoothing the signal with low-pass or moving average filters.

In contrast with the applications of the low-pass filter in Figure 4.3(a), at this stage the input signal has less frequency content at medium and high frequency (see the green line in Figure 4.12) and, thus, a higher cut-off frequency can be set to preserve the driving dynamics more accurately. Figure 4.5 presents the results in terms of speed (a) and acceleration (b) profiles.

The contribution of smoothing out the noise can be appreciated in Figure 4.5(b), where the maximum accelerations/decelerations are further lowered. In turn, the speed profile (Figure 4.5(a)) shows clear speed transitions (e.g. shifting gears) without disturbances, and does not suffer from loosing information at frequencies higher than the pass-band.

Despite of that, resulting accelerations might still be on the borderline of acceptable values (see the peaks indicated with the arrows in Figure 4.5(b)). However, if on one hand the identification of the maximal physical acceleration values could be rather

subjective, it ultimately depends on the type of vehicle and on the speed regimes at which the vehicle is moving. On the other hand, it is widely recognized that accelerations and decelerations exceeding $5m/s^2$ and 6 m/s^2, respectively, are above common values for ordinary conditions (Marczak and Buisson, 2012; Thiemann et al., 2008).

a)

b)
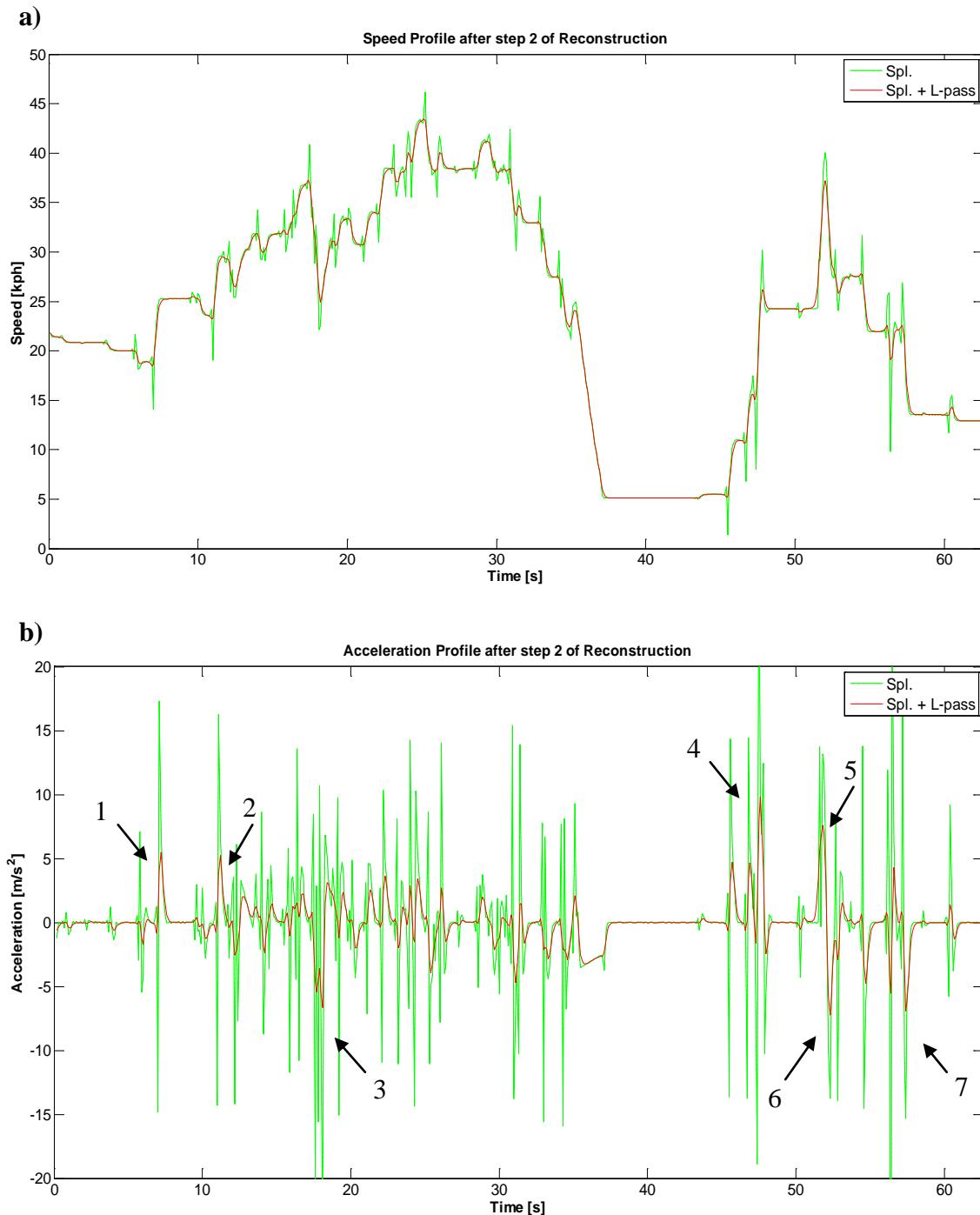
Figure 4.5: Speed (a) and Acceleration (b) profiles after step 2 (L-pass) of reconstruction. The low-pass filter here used is a first-order Butterworth filter with cut-off frequency of 1 Hz.

4.3.3 Step 3: Removing residual unphysical acceleration values, preserving trajectory consistency requirements

The objective of the current step is to remove the residual peaks – in the following referred as *outsiders* – in the acceleration profile that exceed defined thresholds (eventually variable with speed levels), preserving the *internal* consistency of the trajectory (i.e. without modifying the total space travelled) and the *platoon* consistency (i.e. not generating negative inter-vehicle spacing values).

A possible way of doing it is by modifying the position of the outsider (i.e. the vehicle's position at time t) so that the resulting accelerations fulfil the thresholds. Recovering the example in Figure 4.1, it means re-positioning the measured point 2, so that the new mean speeds V_{12}^{*} and V_{23}^{*} produce compatible values for the acceleration. In this study, $3m/s^2$ and 5 m/s^2 were adopted as maximum valid values for accelerations and decelerations, respectively.

To reach this goal aim, the basic idea is to locally modify the position of the outsider assuring *i)* that the resulting accelerations/decelerations fall in the bounds of acceptable values, and *ii)* that no negative inter-vehicle spacing with the actual vehicle in front is produced.

In the following subparagraphs, the necessary pre-processing to take into account the *platoon* consistency requirement is first described, followed by the description of the filtering method.

Leader-Follower Dependency Tree

As the platoon consistency aims at preserving positive inter-vehicle spacing between successive vehicles in a whole platoon, it is clear that the filtering procedure have to be applied *sequentially* to all individual trajectories in the platoon. Indeed, when locally fixing vehicle positions in a certain time window, the position of the vehicle in front (needed to calculate the inter-vehicle spacing) can be used only if the trajectory of that vehicle have been previously filtered. Things are further complicated by the presence of lane-changes in real traffic, as a couple of leader/follower vehicles may switch their positions (e.g. by overtaking each other), thus creating a circular dependency. To clarify, a simple example from the I80-1 dataset is presented in Figure 4.6, where only two vehicles are involved.
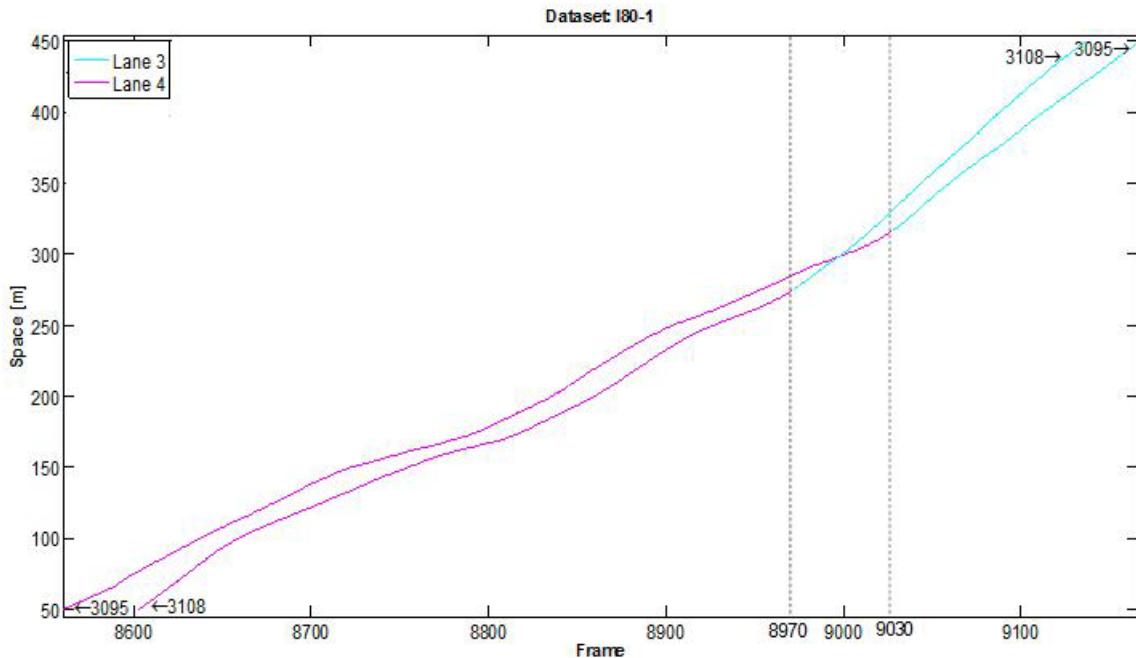
Figure 4.6: Example of a sequence of vehicles that invert their positions (leader/follower) due to lane-changing. Colors represent the lane IDs, while the vehicle IDs is indicated by arrows

The example depicts how vehicle 3108, who was initially following vehicle 3095 in lane 4, passes its leader (vehicle 3095) through a lane-change to lane 3. In this situation, in order to filter the entire vehicle 3108's trajectory, we needed the filtered trajectory of vehicle 3095 (between frames 8600 and 8970) to apply the inter-vehicle spacing constraint in the procedure. However, to filter vehicle 3095's trajectory, we needed vehicle 3108's one (from frame 9030 to the end). This generates the so-called circular dependency between leaders and followers, due to lane-changing. It is worth noting that the example shown here is a simplification case with only two involved vehicles, while there exist situations with up to 60 involved vehicles. A complex example is shown in Figure 4.7.
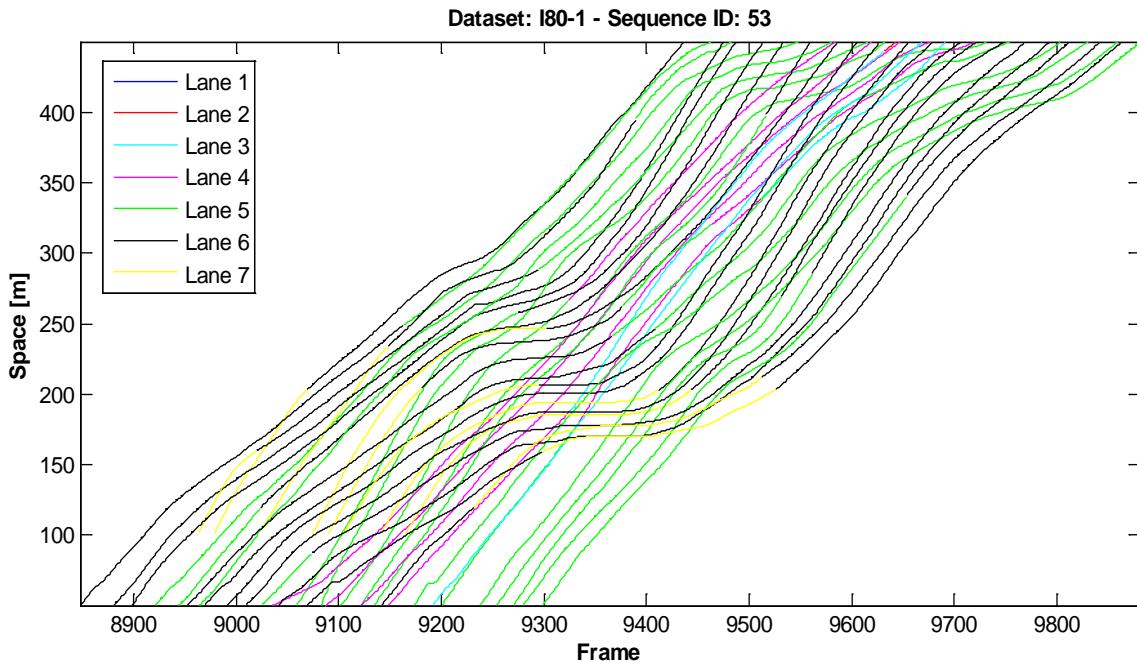
Figure 4.7: Example of a complex circular dependency.

Therefore, in order to apply the platoon consistency constraint in the filtering method, the complete tree of all the dependencies among each vehicle and its leaders was needed[4]. To create it, we designed an algorithm that tags each vehicle in a specific level, if all its leaders belong to upper levels. Therefore, vehicles with no leaders (i.e. the first tracked vehicles) were tagged in level 1. Vehicles with leaders tagged in level 1 were tagged in level 2, while those with leaders tagged in levels 1 or 2, were tagged in level 3, and so on. Defined as k the last completed level, if there are no more vehicles that can be tagged in level $k+1$ (because not all their leaders have already been tagged in levels up to k), a circular leader-follower dependency is encountered. Indeed, among the remaining vehicles, there would exist at least one subset of vehicles where each of them has some leaders already tagged (in levels up to k) while others not tagged yet, but belonging to the subset. In this case, the vehicles belonging to that subset (i.e. the vehicles involved in the circular dependency) can be tagged in level $k+1$, with a different label to distinguish them from regular tags. By tagging the vehicles involved in the circular dependency, other vehicles (i.e. those whose leaders were involved in the

---

[4] In the filtering procedure here proposed, for each vehicle the inter-vehicle spacing is evaluated from its current leader. Indeed, the vehicles with no leaders are first filtered and the procedure continues scanning the platoon upstream. The same procedure could be applied considering the inter-vehicle spacing evaluated in terms of distance from the follower. In this case the scanning direction would be downstream.

circular dependency) can be successively tagged and the algorithm proceed until there are no more vehicles to tag.

Filtering method

As it is clear from the example in Figure 4.1, changing the longitudinal position of the vehicle at time *t* modifies the distance travelled by the vehicle both between *t-1* and *t*, and between *t* and *t+1*. Thus, fixing the vehicle's acceleration between *t-1* and *t* would change also the acceleration between *t* and *t+1*, possibly producing a new outsider. This chaining process ends when the acceleration of the vehicle between *t+k-1* and *t+k* falls in the range of the acceptable values. Therefore, given an outsider at time *t*, the first step consists in finding the instant *t+k* at which the vehicle turns to have an acceleration compatible with the (modified) position at *t+k-1*. The detecting algorithm is based on the hypothesis of linear variation of the mean speeds in the interval [*t*; *t+k*]. It is worth noting that such hypothesis does not imply that the reconstructing vehicle's trajectory in that interval would be linear (which is only a special case).

Taking for granted the vehicle's positions at times *t-1* and *t+k* (i.e. the related accelerations are acceptable), hypothesis on the curve that reproduces the vehicle's trajectory between *t* and *t+k-1* are needed. This is not an easy task, since the trajectory is always a very flat signal and the impact of the interpolation curve in terms of speed dynamics cannot be directly appreciated. Therefore, instead of operating on the "LocalYs", the filter operated on the mean speeds (as defined in Section 4.2), and the reconstruction curve was a 5-th degree polynomial interpolation.

A constrained interpolation is needed to preserve the internal consistency of the trajectory (i.e. changing mean speeds of the outsiders must not modify the original space travelled), plus additional boundary conditions on the derivatives, and on the number of sign inversions of the jerk in 1-sec time window.

In addition, to accomplish to the *platoon* consistency requirement, an additional constraint on the inter-vehicle spacing was needed, i.e. the local reconstruction of the vehicle's trajectory in the time window [*t*; *t+k*] must not give rise to negative inter-vehicle spacings with the actual vehicle in front, which was possible if individually filtering each vehicle's trajectory.

However, to correctly take into account this constraint, the trajectory of the vehicle in front should have been already filtered. Therefore, the procedure needed to be applied sequentially to the entire dataset following the leader-follower dependency tree introduced earlier. Indeed, vehicles' trajectories were sequentially filtered starting from those tagged in level 1 of the dependency tree (i.e. vehicles with no leaders), and proceeding accordingly to the levels in the tree. Once there were two or more vehicles involved in a circular dependency (see, for example, Figure 4.6), they were repeatedly filtered in sequence until a threshold value for the inter-vehicle spacing is no more activated.

One could argue that, given the methodology described in the current step, the preliminary removal of the *outliers* in step 1 was unnecessary. However, since the threshold values for the identification of the *outsiders* are sensibly lower than those for the *outliers*, in step 3 the detection of the *outliers* at time t was still possible, while the identification of the next good point would lead to an excessively wide reconstruction time window [t; $t+k$], changing too much data in each individual trajectory. Therefore, we decided to split the detection of *outliers* and *outsiders* in two phases, in order to have variable threshold values for the identification and reconstruction.

Further, as pointed out in the previous sections, the quality of each trajectory (e.g. the type of errors or the frequency component) may vary sensibly in a large dataset and thus, the algorithm parameters should need fine tuning for each specific vehicle's trajectory. However, this is not feasible in case of large-data application (as this one). On the other hand, we noted that the performances with constant filter parameters sensibly varied with the amount of medium-high frequency components in the mean speed profile. For this reason, we required a linear smoothing of the mean speed profile in step 2 to make its spectral density function more uniform among all the vehicles with regards to the medium-high frequency content.

Results from step 3 are presented in Figure 4.8, in terms of speeds (a) and accelerations (b). From the figure, the effect on the *outsiders* is clear in terms of resulting accelerations (Figure 4.8(b)). The peaks indicated by the arrows in Figure 4.5 disappeared as a consequence of a (better) positioning of the point in space. In turn, the replaced speed profile has a more regular shape.

a)

b)
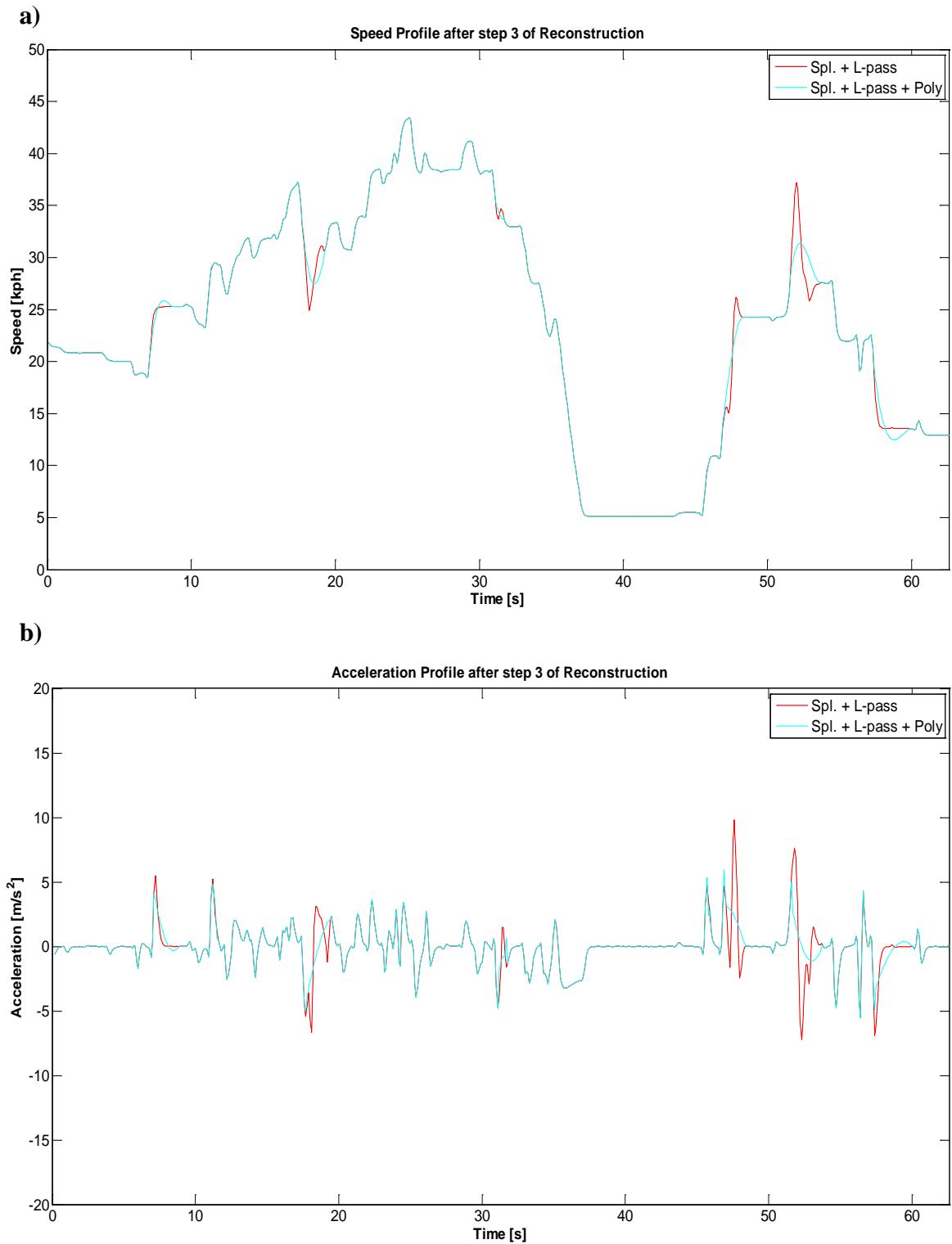
Figure 4.8: Speed (a) and Acceleration (b) profiles after step 3 (Poly) of reconstruction. At this stage, it was used a 5-th degree polynomial interpolation constrained on the space travelled, plus additional boundary conditions and constraints.

4.3.4 Step 4: Cut-off the high and medium frequency responses eventually generated from the previous step

The objective of the current step is to remove the high and medium frequency responses (eventually) generated in the previous step. Operatively, we noticed that the boundary conditions on derivatives applied in the constrained interpolation at step 3 needed a slight relaxation in order to let the algorithm perform faster. As a consequence, the repositioning of the outsiders could (eventually) generate some discontinuities in the speed profile (i.e. angles). Therefore, to smooth-out this sudden irregularities, we re-applied the low-pass filter used at step 2.

Results from the conclusive step, are shown in Figure 4.9.

a)

b)
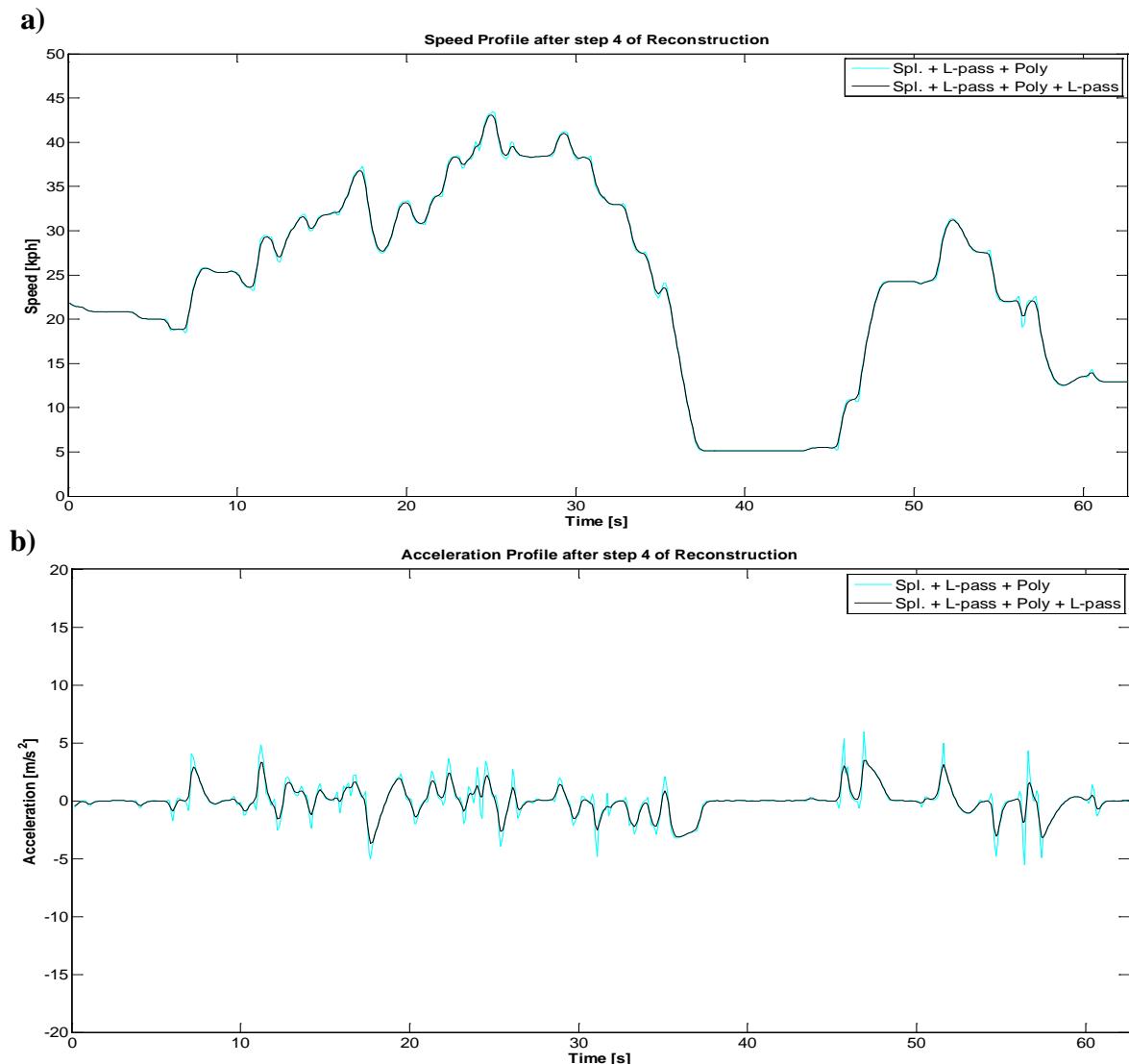
Figure 4.9: Speed (a) and Acceleration (b) profiles after step 4 (L-pass) of reconstruction. The low-pass filter used here was designed as in Section 4.3.2.

4.4   Peculiarities of NGSIM data

As shown in the recent literature, the quality of the NGSIM video-processed data is undermined by a large amount of measurement errors. Punzo et al. (2011b) discussed on the necessity to inspect the quality of trajectory data with regards to the structure of the error in point measurements and its propagation on the space travelled, and presented a method to assess trajectory data accuracy, based on jerks' analysis, consistency analysis and spectral analysis. The resulting considerations suggested to carefully handle values in "Velocity" and "Acceleration" fields of NGSIM datasets and, whenever possible, to directly estimate speeds and accelerations from an appropriately filtered "LocalY".

The analysis revealed the existence of two main types of errors:

- errors on vehicles' positions giving rise to negative inter-vehicle spacing between successive vehicles;

- errors on vehicles' positions giving rise to unphysical acceleration values (derived from "LocalY").

Based on the results depicted in Punzo et al. (2011b), in this study we conducted a deeper investigation to identify and quantify all the sources of errors included in the data, focusing on the NGSIM I80-1 dataset. As a result, a complete summary of error types was defined, including and further specifying the findings in Punzo et al. (2011b):

1.   Errors due to motorcycle mis-identification;

2.   Errors due to vehicle's lane ID mis-identification;

3.   Errors due to lane-changing;

4.   Errors due to merging;

5.   Errors due to illegal overtaking;

6.   Errors due to large platoon mis-tracking;

7.   Errors due to individual vehicles' mis-tracking;

8.   Errors due to vehicle IDs switching;

It is worth noting that the biggest errors on the individual vehicles' acceleration values, i.e. the so-called *outliers*, are the results of vehicles' mis-tracking, as pointed out in Section 4.2.1.

Based on these observations, the original dataset (*raw* data) presented error types that could not be treated appropriately with the trajectory reconstruction procedure here proposed. Indeed, in the procedure described in Section 4.3, the occurrence of negative inter-vehicle spacing is not a trigger for the reconstruction at step 3, while it is only a constraint on the local reconstruction of the mean speed profile (derived from "LocalY").

Therefore, a pre-processing stage was needed to treat above error types in order to obtain a consistent dataset in terms of total space travelled and inter-vehicle spacing. The methodology here applied aims at identifying (and fixing) errors that give rise to inter-vehicle spacing conflicts, which can be the results of misidentification of motorcycles overtaking other vehicles (error type 1), inaccurate tracking of a lane-changing (error types 3 and 8), the lane-based modeling assumption underneath the tracking algorithm and video-processing (error types 2, 4 and 5), large platoon (error type 6) or individual vehicle (error type 7) mis-tracking.

In order to do that, the identification of vehicles involved in these situations was done by cross-checking the list of all the conflicts occurred in the dataset with the list of all the lane-changes that happened in the monitoring period. Indeed, given a conflict between two vehicles at time t, if at least an influential lane-changing (i.e. a lane-changing performed by one of the vehicles involved in the conflict) happened in a 10 seconds time windows centered on t, that conflict had greater probability to have been caused by error types 3, 4 or 5. In this case, the error cause was detected by the algorithm (with a success rate of 100% of the cases after cross-checking with original video) and the involved vehicles were fixed accordingly (e.g. the instant of lane-changing was anticipated or delayed by less than 1 seconds in order to avoid the conflict, or the presence of a vehicle in the emergency lane was identified with a different lane id). On the other hand, if no influential lane-changing happened, that conflict had greater probability to have been caused by error types 1, 2, 6, 7 and 8. Video analysis allowed to correctly identify the cause of the error, and operate the adequate fixing. For example, for error type 1, the vehicle class was modified accordingly (e.g. Figure 4.10(a)), as well as for error type 2

with regards to the lane id. For error types 6 and 7, the video analysis allowed the identification of the time windows when the mis-tracking happened, and the mis-tracked vehicle's trajectory was reconstructed based on a Newell car-following model behavior (2002) with respect to the correctly tracked leader vehicle (e.g. Figure 4.10(b)). Finally, for error type 8, the correct pieces of vehicle trajectory data were re-assigned to the correct vehicles that were involved in the vehicle ID's switching (e.g. Figure 4.10(c)).

It is worth noting that error types 6 and 7 related to the mis-tracking could also happened without generating a conflict. In these cases, there was no possibility to detect the presence of these errors through a data analysis. However, we noted that long vehicles' mis-tracking were characterized by a constant mean speed profile in data. Cross-checking this information with the visual inspection of the videos, we were able to fix, with the same Newell-based procedure described above, most of the long platoon and individual mis-tracking errors entailed in the data.

From the above analysis, we noted that the amount of errors that happened in vehicle's tracking at the beginning (the road segment monitored by camera 1) and at the end (the road segment monitored by camera 7) of the section was considerably higher than in the road segments monitored by the remaining camera. As a consequence, we decided to avoid pre-processing a large amount of data, and removed these data from the dataset.

a)

b)

c)
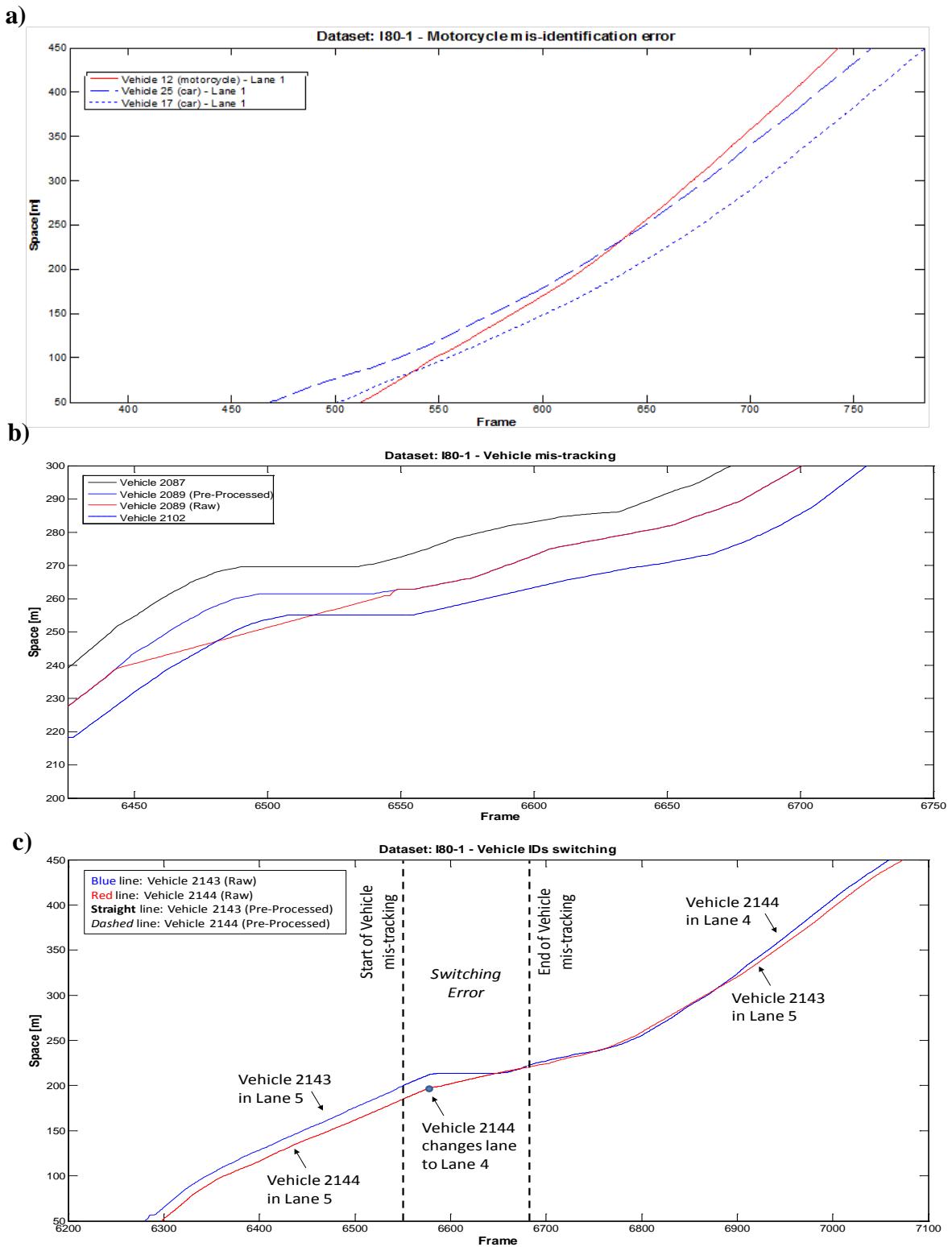
Figure 4.10: Examples of different types of errors entailed in NGSIM I80-1 dataset. (a) showed a spacing conflict arising from mis-identification of the vehicle 12 of type "motorcycle" with two vehicles (17 and 25) of type "car"; (b) showed a Newell-based trajectory reconstruction in a mis-tracking time window; (c) showed a vehicle IDs switching error where two vehicles switch their IDs in a transition window and their tracked positions are mixed up from a give instant in time.

4.5   Application to the NGSIM I80-1 dataset

The procedure described in Section 4.3 was applied to all the vehicles' trajectories of the NGSIM I80-1 dataset. *Reconstructed* NGSIM I80-1 dataset is publicly available on the MULTITUDE website (2014).

The complete picture of the traffic dynamics in the monitored period is depicted in Figure 4.11, by means of the space-time evolution of the space mean speeds (calculated with *raw* data). The figure clearly shows the upstream propagation of three waves, accompanied by intense congestion.

In the following, results from vehicles' trajectories reconstruction are presented in terms of the impacts on *i)* individual vehicle trajectories, and on *ii)* the entire dataset.
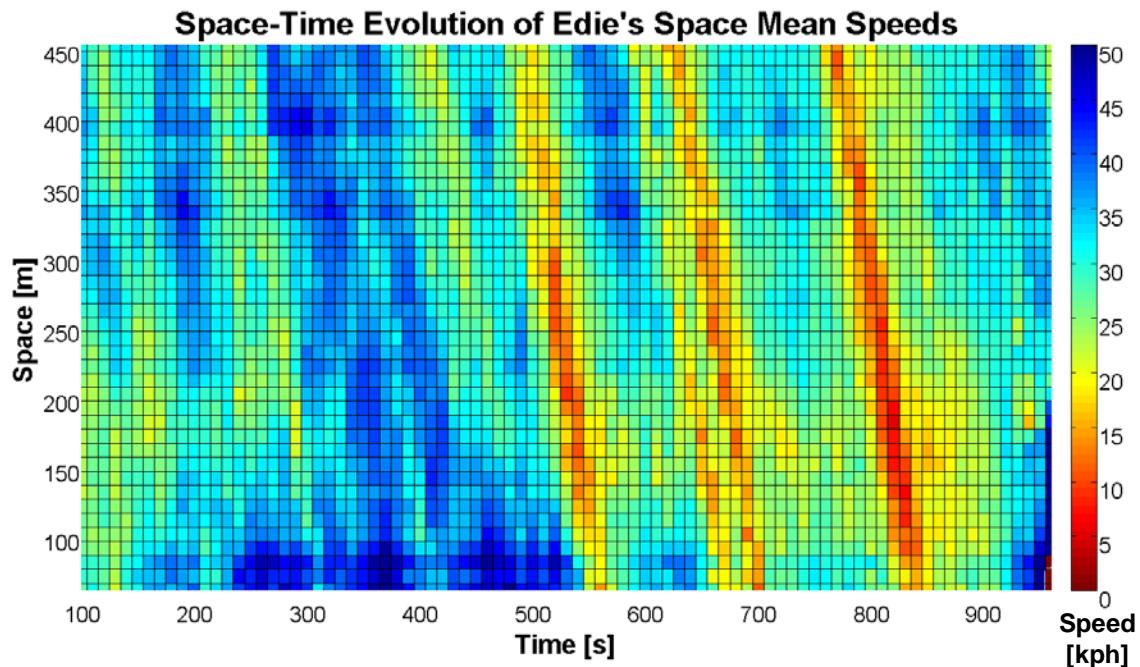
**Figure 4.11:** Time-Space speed contour plot based on Edie's space mean speeds (Edie, 1974).

### 4.5.1 Individual vehicle trajectory analysis

The overall effect from the application of the complete procedure can be appreciated in Figure 4.12, where the resulting speed (a) and acceleration (b) profiles are plotted against the NGSIM data. Figure 4.12(a) shows how the outliers completely disappeared

(step 1), and the driving dynamics are clearer (step 2) and unbiased (step 3) especially at the speed transitions.

In terms of accelerations, Figure 4.12(b), though impressive, does not clearly explain the implication of reducing the bias in the measurements. In turn, if we look at the frequency spectrum of the accelerations presented in Figure 4.13, we may see *i)* the amount of bias enclosed in the original data (blue line), which covers the entire frequency range of the signal, and *ii)* the result from the complete procedure (black line). The characteristics of the frequency spectrum after reconstruction, are comparable with the results from the literature where the whole frequency response of the signal is in the range of frequency up to 2Hz (i.e. human/vehicle responses are unlikely to have a frequency exceeding this value; Punzo et al., 2011b).

a)

b)
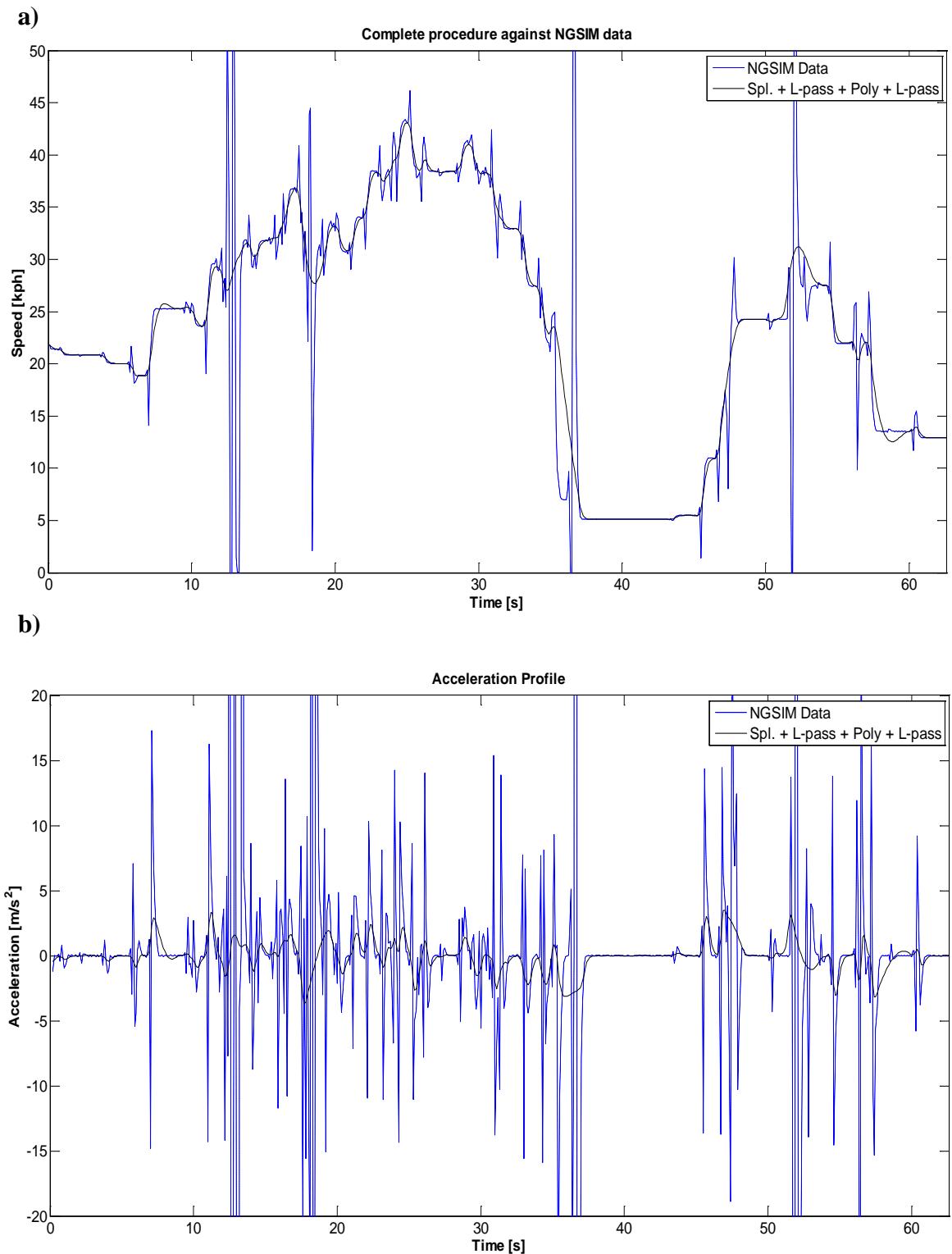
Figure 4.12: Comparison of the Speed (a) and Acceleration (b) profiles after reconstruction, against NGSIM data.
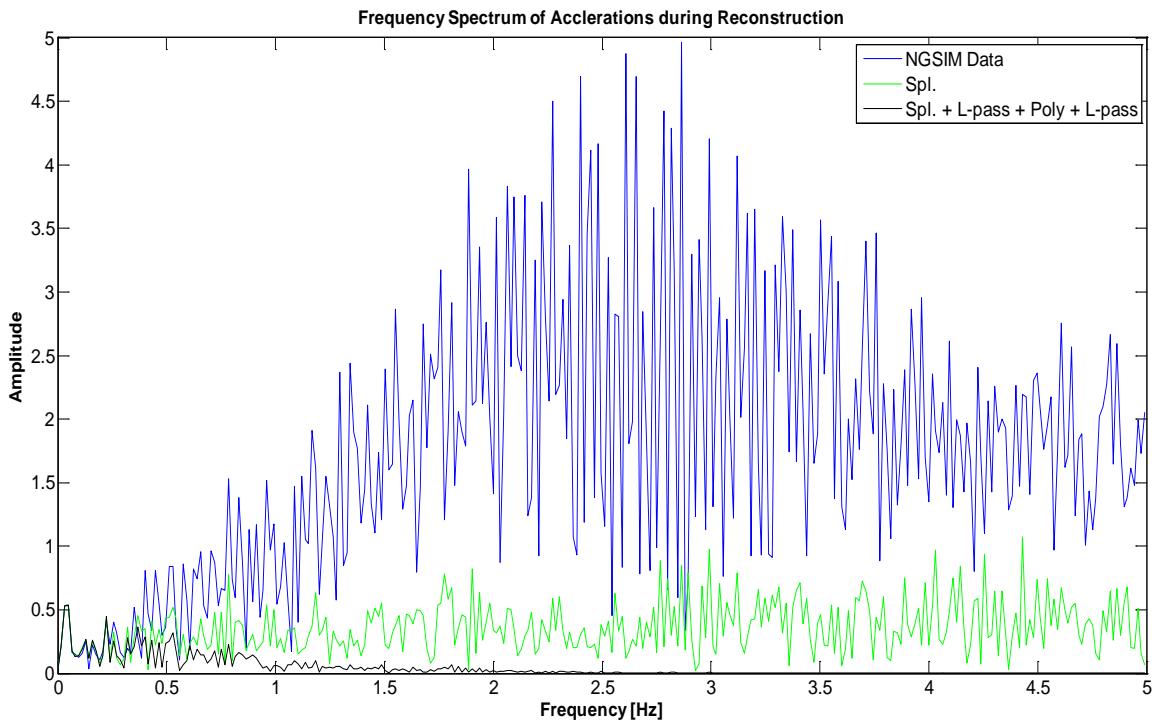
Figure 4.13: Frequency spectrum of accelerations (as defined in Section 4.2) during reconstruction.

4.5.2   *Dataset analysis*

In this section, the impact of the proposed reconstruction method on the quality of the entire NGSIM I80-1 dataset is evaluated. It is worth noting that, as stated in Section 4.3, the filtering procedure could have performed better if parameters of the algorithm were appropriately calibrated for each individual trajectory. However, given the size of the dataset and the limited time available, we needed to equalize the characteristics of the input signal to let the algorithm work faster and with fewer (and lowered) residual peaks in the resulting acceleration profile.

Dataset analysis was performed through the following investigations:

a.   analysis of *acceleration* distributions;

b.   analysis of *maximum speed* distributions;

c.   analysis of *minimum spacing* distributions;

d.   analysis of distribution of *residuals* on "LocalY";

e.   analysis of macroscopic traffic flow characteristics.

Figure 4.14 presents the comparison of the acceleration (as defined in Section 4.2) distributions before and after vehicles' trajectories reconstruction.

The distribution of the *raw* NGSIM data (blue bars) confirmed the findings of Punzo et al. (2011b) regarding the incredibly high percentage of *outliers* and *noise* in the data (i.e. see, for example, the unphysical absolute acceleration values exceeding 10 m/s^2). The result of the proposed technique is that of completely removing this bias (no frequency of extreme acceleration values; see the absence of black bars).
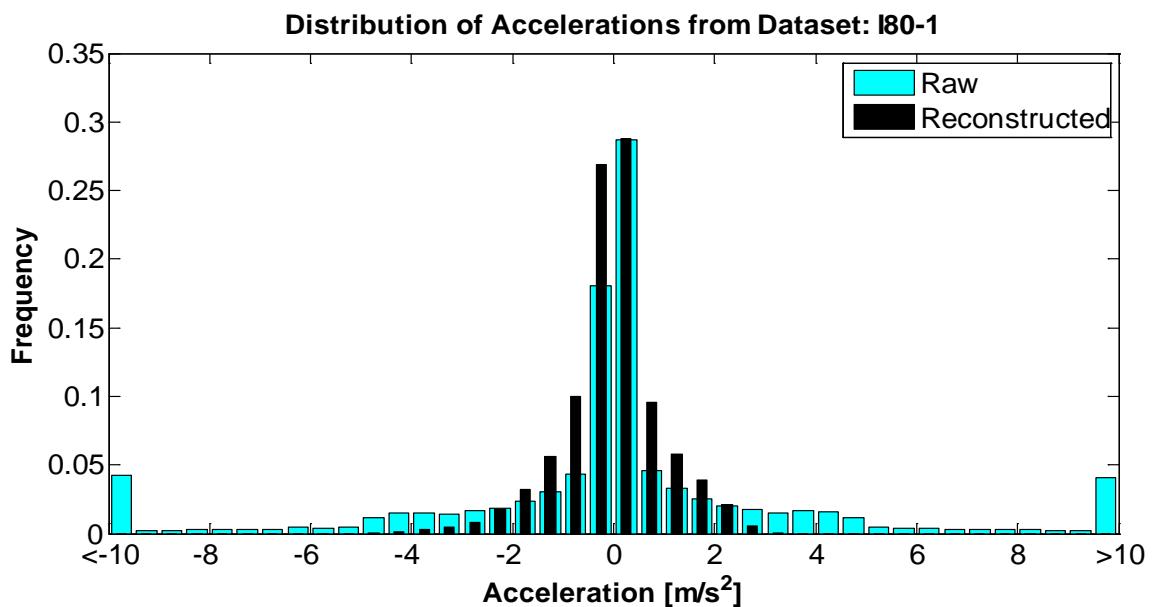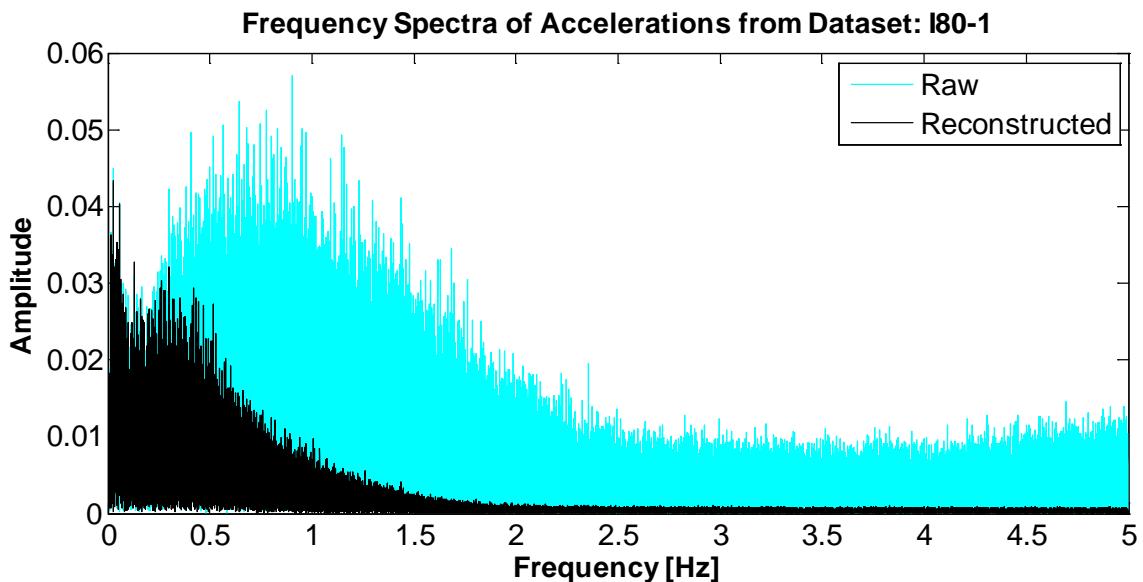

Figure 4.14: Comparison of distributions of accelerations (as defined in Section 4.2) between *raw* (blue bars) and *reconstructed* (black bars) data.

Also with regards to the spectral analysis of the accelerations (see Figure 4.15), the magnification of the high-frequency components of the signal is self-evident in the *raw* data, while, after reconstruction, frequencies are brought back in a range of feasible values (up to 2Hz; for details, see Punzo et al., 2011). It is worth noting that in the spectral analysis conducted here, the accelerations for all individual vehicles were aggregated in a single signal, which explains the differences (in the shape and in the frequency scale) with results shown in Figure 4.13.

Figure 4.16 presents the comparison of the *maximum* speed (as defined in Section 4.2) distributions per each vehicle, before and after vehicles' reconstruction.

The distributions of the *raw* NGSIM data (blue bars) confirmed the presence of a high percentage of vehicles (about 4%) with maximum speed greater than 30 m/s^2 (i.e. the so-called *outliers*, with peaks also greater than 50 m/s, as shown in Figure 4.2). After reconstruction, instead, this percentage lowered down to zero.

Frequency Spectra of Accelerations from Dataset: I80-1

Figure 4.15: Comparison of frequency spectra of accelerations (as defined in Section 4.2) between *raw* (blue bars) and *reconstructed* (black bars) data.
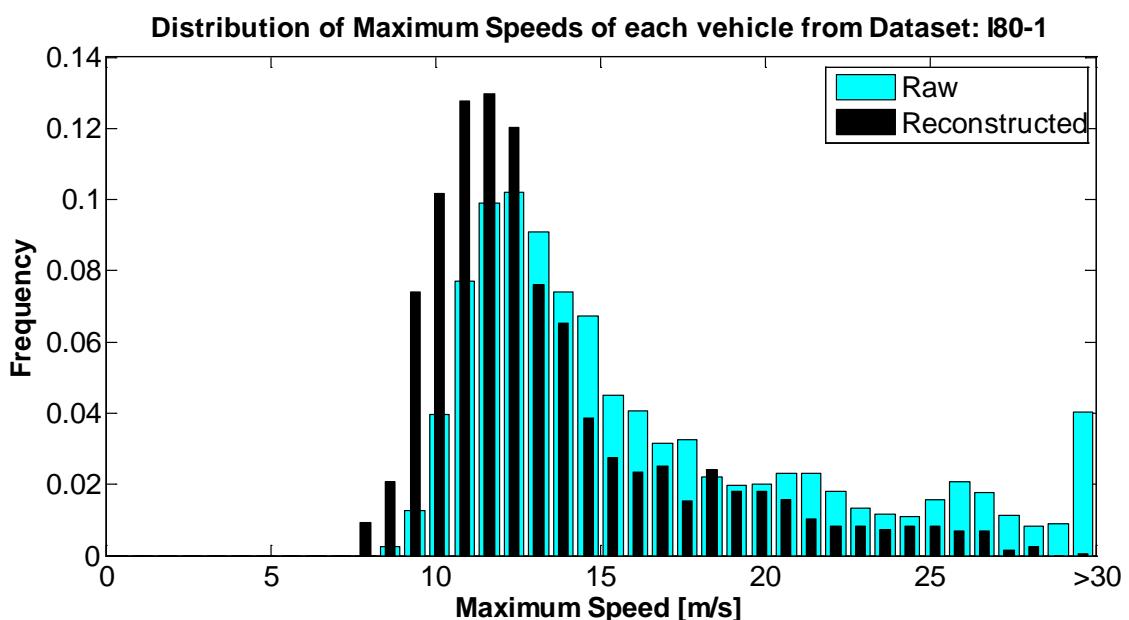
Distribution of Maximum Speeds of each vehicle from Dataset: I80-1

Figure 4.16: Comparison of distributions of maximum speeds (as defined in Section 4.2) of each individual vehicle, between *raw* (blue bars) and *reconstructed* (black bars) data.

In order to appreciate the impact of the proposed reconstruction technique on the *platoon* consistency, Figure 4.17 shows the comparison between *raw* and *reconstructed* distributions of the *minimum* inter-vehicle spacing per each vehicle (with respect to the vehicle in front).

In the *raw* NGSIM data, a high percentage of vehicles (about 7% of the entire I80-1 dataset) present a negative minimum inter-vehicle spacing from the vehicle in front. After reconstruction, instead, platoon consistency was re-established.

Distribution of Minimum Spacings from Dataset: I80-1
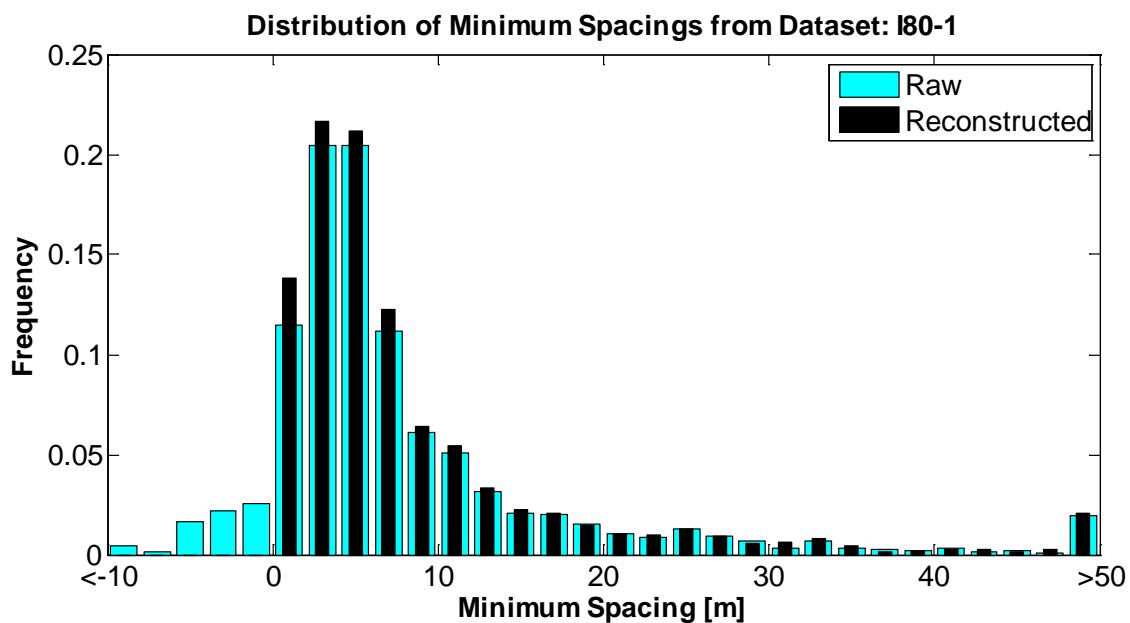
Figure 4.17: Comparison of distributions of minimum inter-vehicle spacings (calculated with respect to the leader) of each individual vehicle, between *raw* (blue bars) and *reconstructed* (black bars) data.

Finally, Figure 4.18 presents the distributions of the residuals on "LocalY" calculated as the difference between *raw* and *reconstructed* vehicles positions each 0.1s., together with normal estimated residuals. From the figure, it is clear that:

- observed residuals are far from being normally distributed (as confirmed by the Normal Probability Plot in Figure 4.19 and by the Kolgomorov-Smirnov test at the level of significance of 5%);
- the observed mean residual is significantly different from zero (as confirmed by the t-Test at the level of significance of 5%).
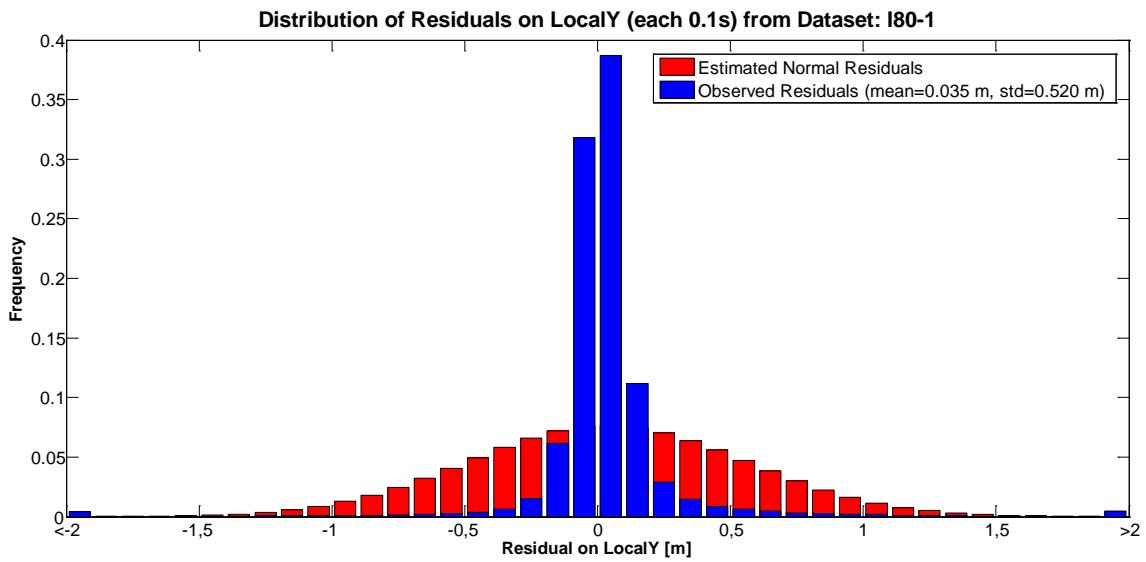
Figure 4.18: Distribution of residuals on "LocalY", calculated as the difference between *raw* and *reconstructed* vehicles' positions each 0.1 seconds (blue bars). Red bars represent the estimated normal residuals based on observed ones.
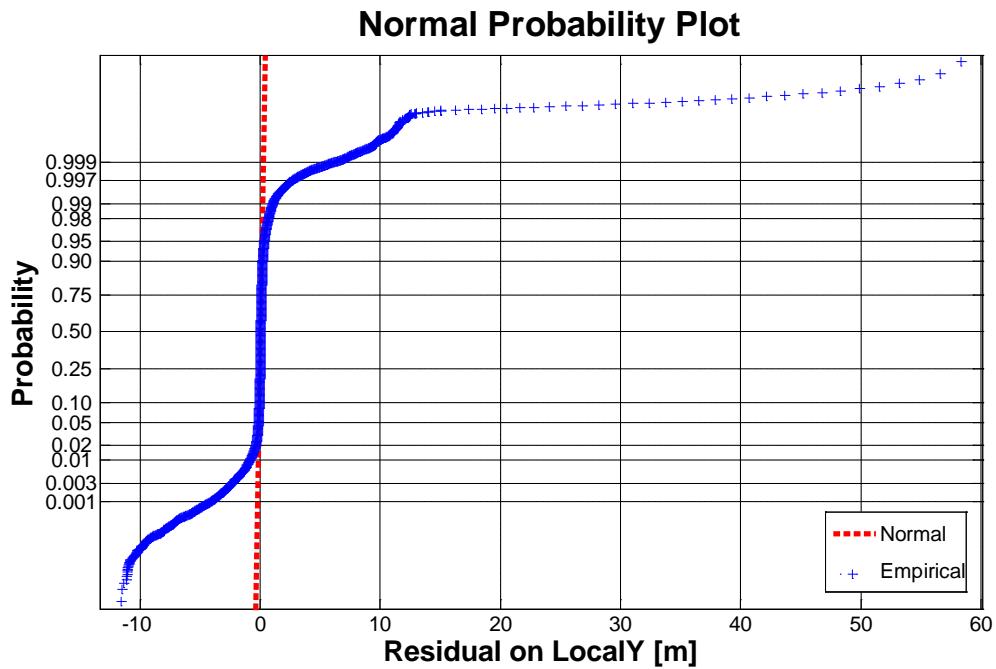
Figure 4.19: Normal Probability Plot of the empirical residuals on "LocalY".

However, despite the large amount of measurement errors in *raw* individual vehicle trajectory data, the impacts on macroscopic traffic flow characteristics is very limited. Indeed, Figure 4.20 shows the comparison between the space-time contour plots of the Edie's space mean speed (Edie, 1974) calculated on *raw* (a) and *reconstructed* (b) data, respectively. It is worth noting that Figure 4.20(a) corresponds exactly to Figure 4.11, and it is here reported again to allow for visual comparison with *reconstructed* data.

a)

b)
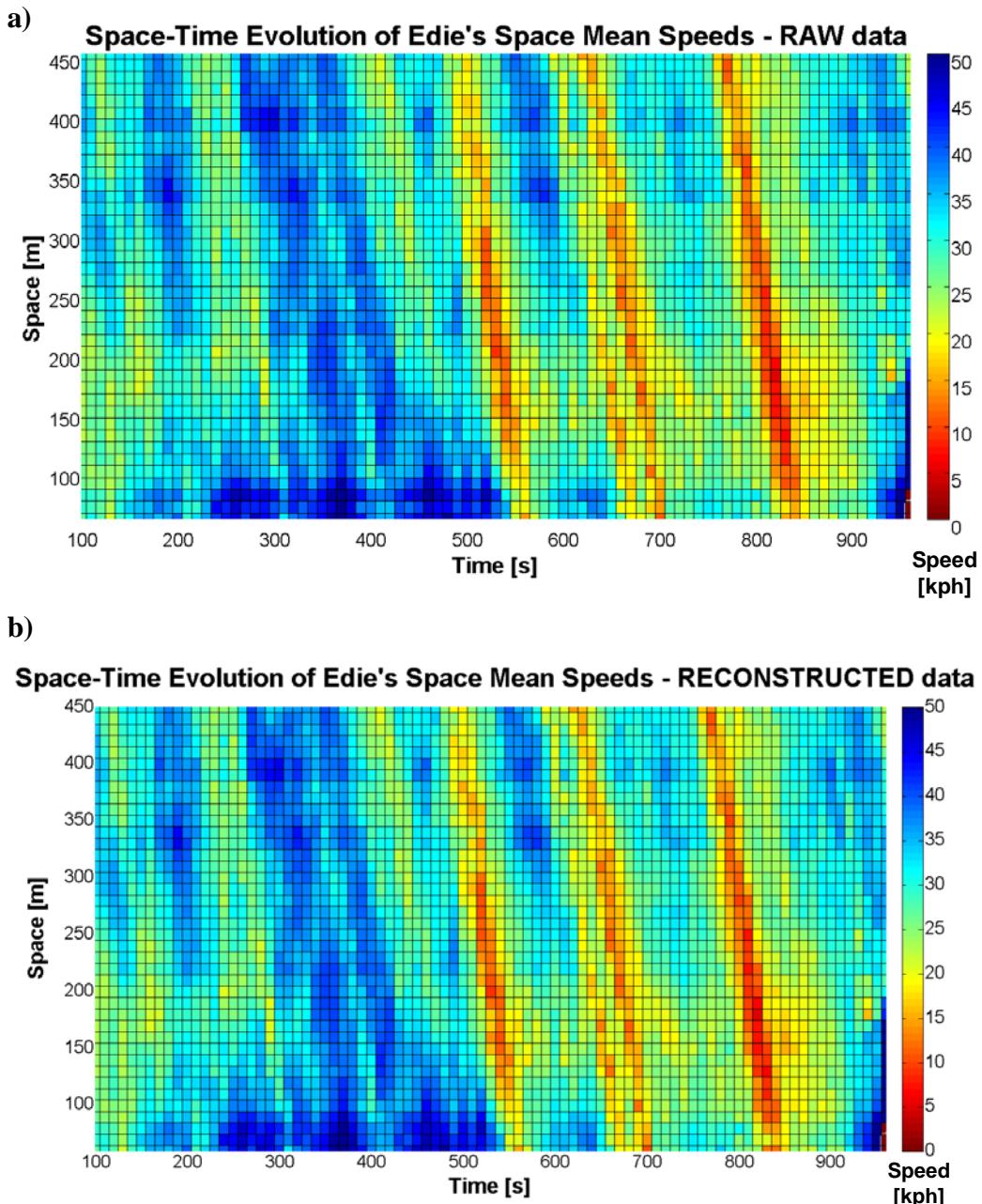
Figure 4.20: Comparison of space-time Edie's space mean speed contour plots calculated with *raw* (a) and *reconstructed* (b) data.

4.6 Impacts on Estimated Distributions of Microscopic Traffic Flow Model Parameters

In this section, the impacts of measurement errors in vehicle trajectory data on model parameters estimated distributions is investigated.

Given these objective, both car-following and lane-changing model parameters were estimated for each individual vehicle in the NGSIM I80-1 dataset against *raw* and *reconstructed* trajectories.

The selected models considered herein were the Intelligent Driver Model (IDM) by Treiber et al. (2000) for car-following, and the MOBIL model by Kesting et al. (2007) for lane-changing. A review of model formulations can be found in Chapter 5, for the IDM model, and in Appendix D, for the MOBIL model.

Car-following model parameter estimation was performed in accordance with the results from Chapter 3 on robust criteria for the specification of the optimization problem. Therefore, in this study, the adopted Measure of Performances (MoPs) were the speed and the inter-vehicle spacing, the Goodness Of Fit (GOF) function was the Root Mean Square Error (RMSE), and the optimization algorithm used to find the parameter values that minimize distance between the simulated and the measured MoP was the OptQuest Multistart (LINDO, 2003).

Conversely, criteria for parameter estimation for the lane-changing model can be found in Appendix D.

It is worth noting that the setup of the estimation problem (e.g. choice of parameter boundaries, convergence threshold of the optimization algorithm, etc.) was the same in both the experiments with *raw* and *reconstructed* data.

In the following sections, the impacts on the IDM (Section 4.6.1) and MOBIL (Section 4.6.2) estimated model parameter distributions are reported. Finally, Section 4.6.3 reports on the evaluation of the impacts of measurement errors on the estimated joint correlation structures among both car-following and lane-changing model parameters.

4.6.1 Car-Following model parameters distributions

The six IDM model parameters (*alpha*, T, $V_f^{Max}, a_f^{Max}, b_f, \Delta S_0$) were estimated for each individual vehicle in the NGSIM I80-1 dataset (excluding those of type "motorcycle"), against both *raw* and *reconstructed* trajectory data.

We adopted the following parameter bounds in the estimation: *alpha* $\in [0.1, 20]$, $T \in [0.1, 5]$, $V_f^{Max} \in [15.6, 40.0]$, $a_f^{Max} \in [0.1, 15]$, $b_f \in [0.1, 15]$, $\Delta S_0 \in [0.1, 10]$.

Figures 4.21 and 4.22 present the comparison of IDM model parameters distributions estimated against *raw* and *reconstructed* data. Figure 4.21 refers to the estimation on speed, while Figure 4.22 on spacing. In the figures, we indicated T with *minTimeHead*, V_f^{Max} with *maxV*, a_f^{Max} with *maxAcc*, b_f with *normDec* and, ΔS_0 with *s0*.

From the figures, it can be seen that the impact of the measurement errors on parameter estimation is very limited. Indeed, two-sample Kolgomorov-Smirnov tests revealed that estimated parameter distributions against *raw* and *reconstructed* data were not significantly different at the level of significance of 5%. Further, also model parameters correlation structures do not change between estimates against *raw* and *reconstructed* data, as shown in Figure 4.23 and 4.24.

These results are not in line with the findings of Ossen and Hoogendoorn (2008a, 2009) where calibration experiments were performed using synthetic data with *normally distributed* error structures added ex-post. A possible explanation of such difference could be the considerably different empirical distribution model of the error, as shown in Figure 4.18, with respect to the synthetic one adopted in Ossen and Hoogendoorn (2008a, 2009).

To confirm this guess, we adopted the *reconstructed* data (leader and follower trajectories) as the "ground truth" to compare the simulation errors, and performed two simulation experiments for each individual vehicle, in one using parameters estimated against *raw* data, while in the other those estimated against *reconstructed* data. Figure 4.25 shows the distribution of simulation errors on both speed (a) and inter-vehicle spacing (b). It is worth noting that, in the legend, the labels "raw" and "reconstructed" refer to the parameters dataset adopted in the simulation.

Simulation results confirmed that measurement errors slightly affected the results of parameter estimation, producing an average increase in the error between the simulated and the observed (i.e. *reconstructed*) follower trajectory of about 8%.

According to these findings, we may conclude that the car-following model operates like a filter. This behavior can be observed by plotting *raw* (a) and *reconstructed* (b) follower trajectories together with the simulated ones. Figure 4.26 refers to the simulation using model parameters estimated on speed, while Figure 4.27 to those estimated on spacing.
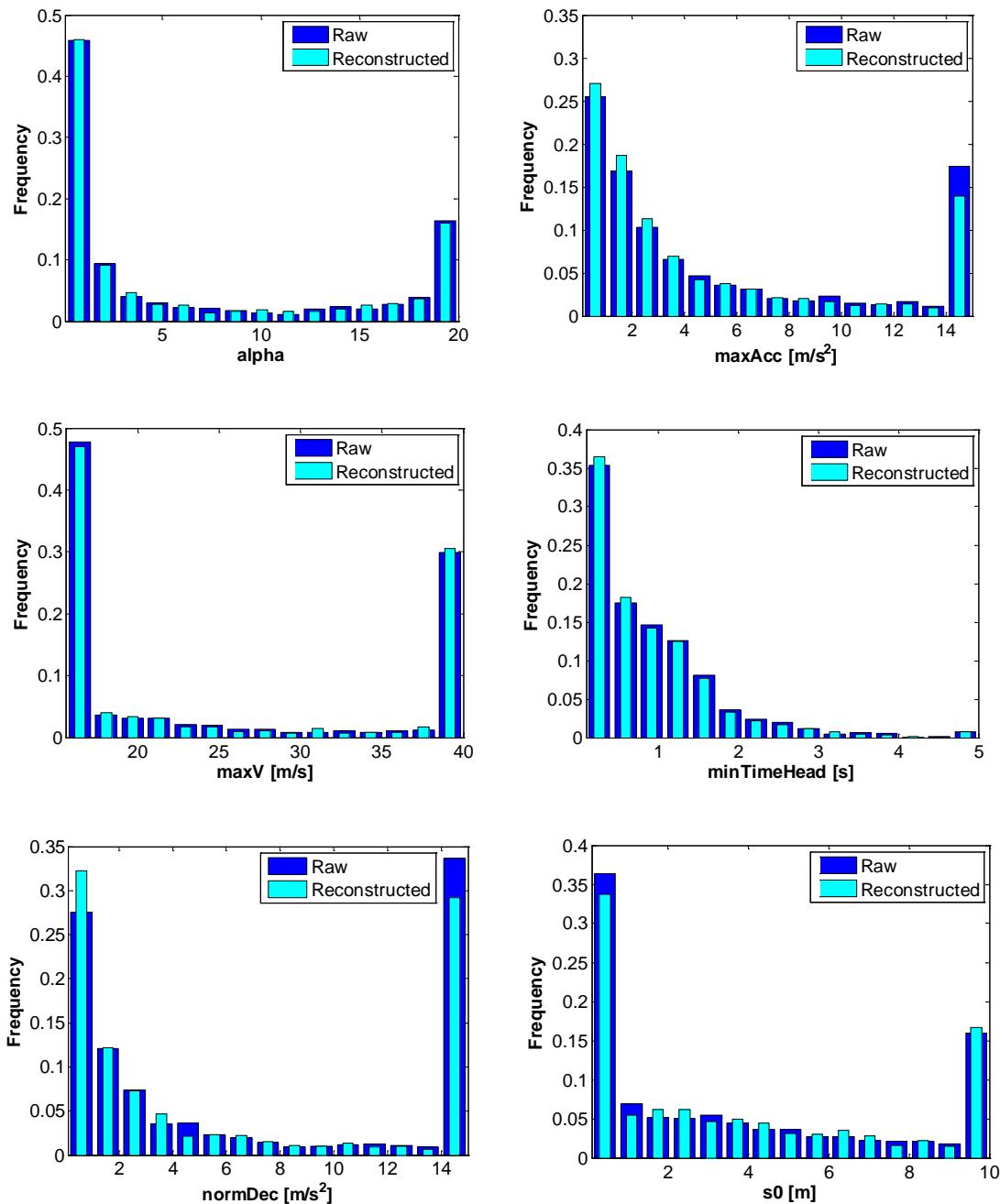
Figure 4.21: Comparison of distributions of IDM model parameters estimated using the RMSE on speed. Blue bars refer to the estimation against *raw* trajectory data, while cyan ones against *reconstructed* data.
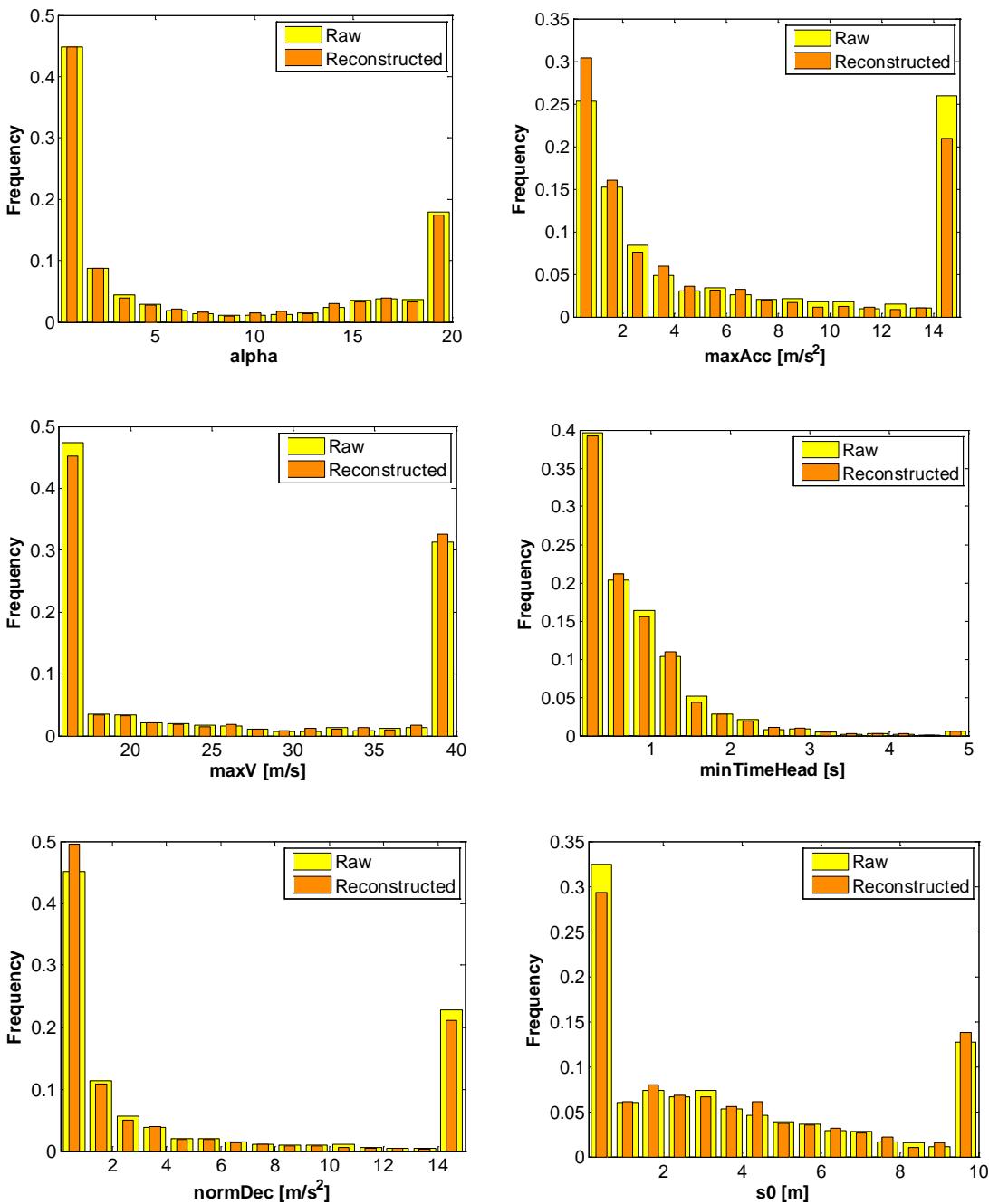
Figure 4.22: Comparison of distributions of IDM model parameters estimated using the RMSE on inter-vehicle spacing. Yellow bars refer to the estimation against *raw* trajectory data, while orange ones against *reconstructed* data.

a)   **Correlation Matrix - Estimation on Speed against Raw data**

Correlation Matrix - Estimation on Speed againt Reconstructed data

b)
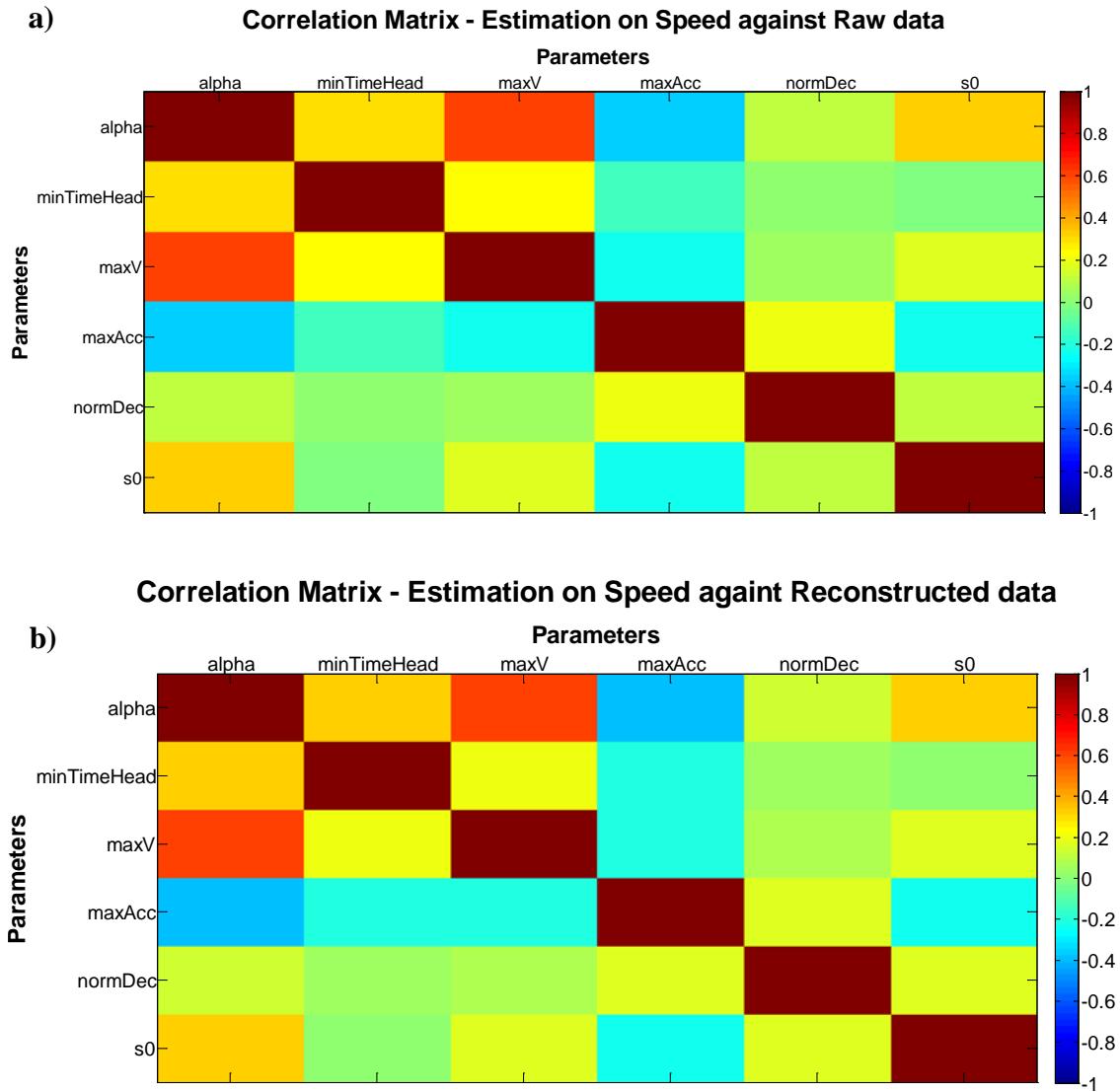
Figure 4.23: Comparison of correlation structures among IDM model parameters estimated on speed against *raw* (a) and *reconstructed* (b) data.
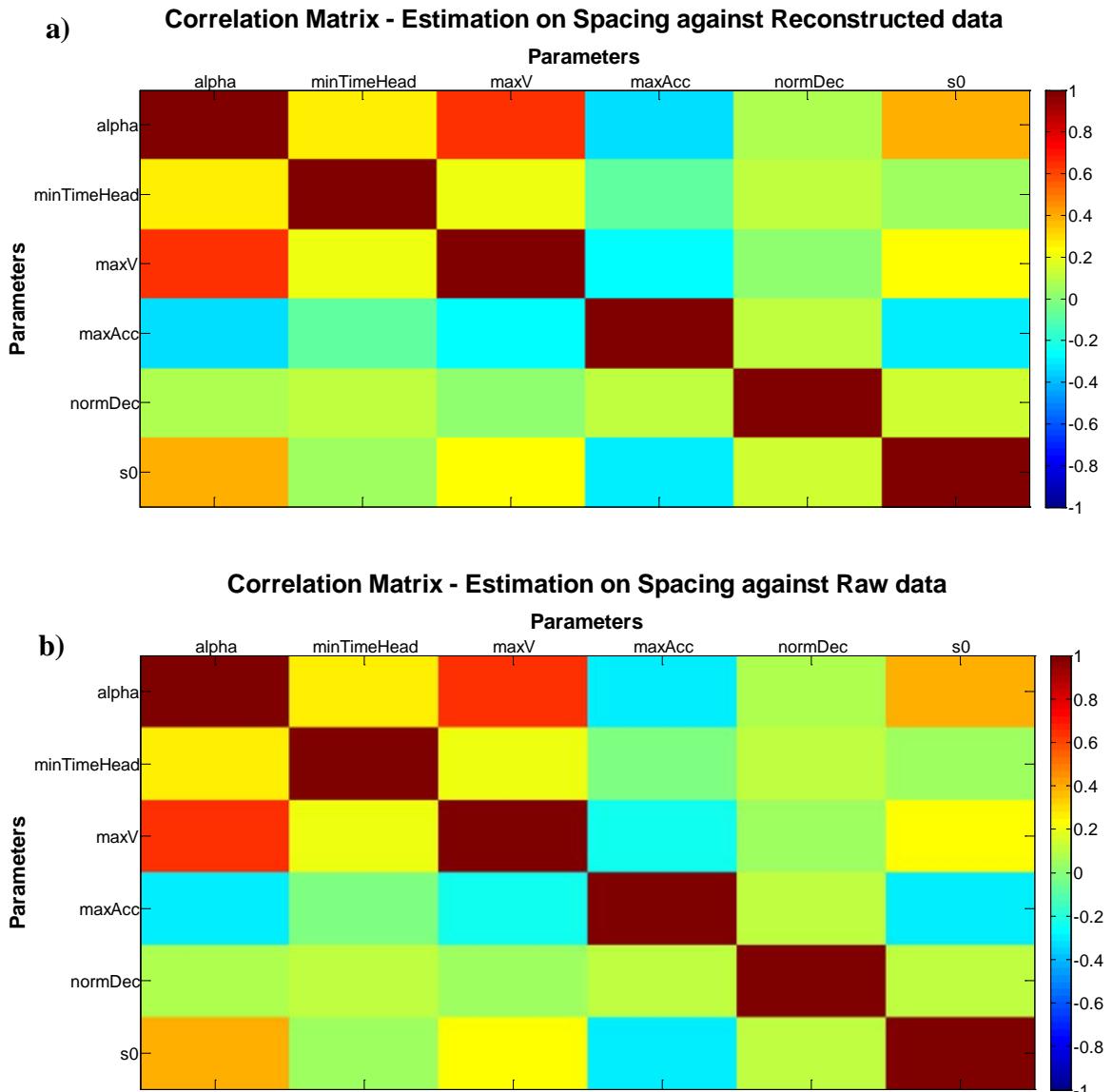
Figure 4.24: Comparison of correlation structures among IDM model parameters estimated on spacing against *raw* (a) and *reconstructed* (b) data.

a)

b)

Figure 4.25: Comparison of distributions of simulation errors (with respect to reconstructed follower vehicle trajectories). Label "raw" refers to simulations with parameters calibrated against *raw* data, while label "reconstructed" refer simulations with parameters calibrated against *reconstructed* data. (a) refers to the calibration on speed, while (b) on inter-vehicle spacing.
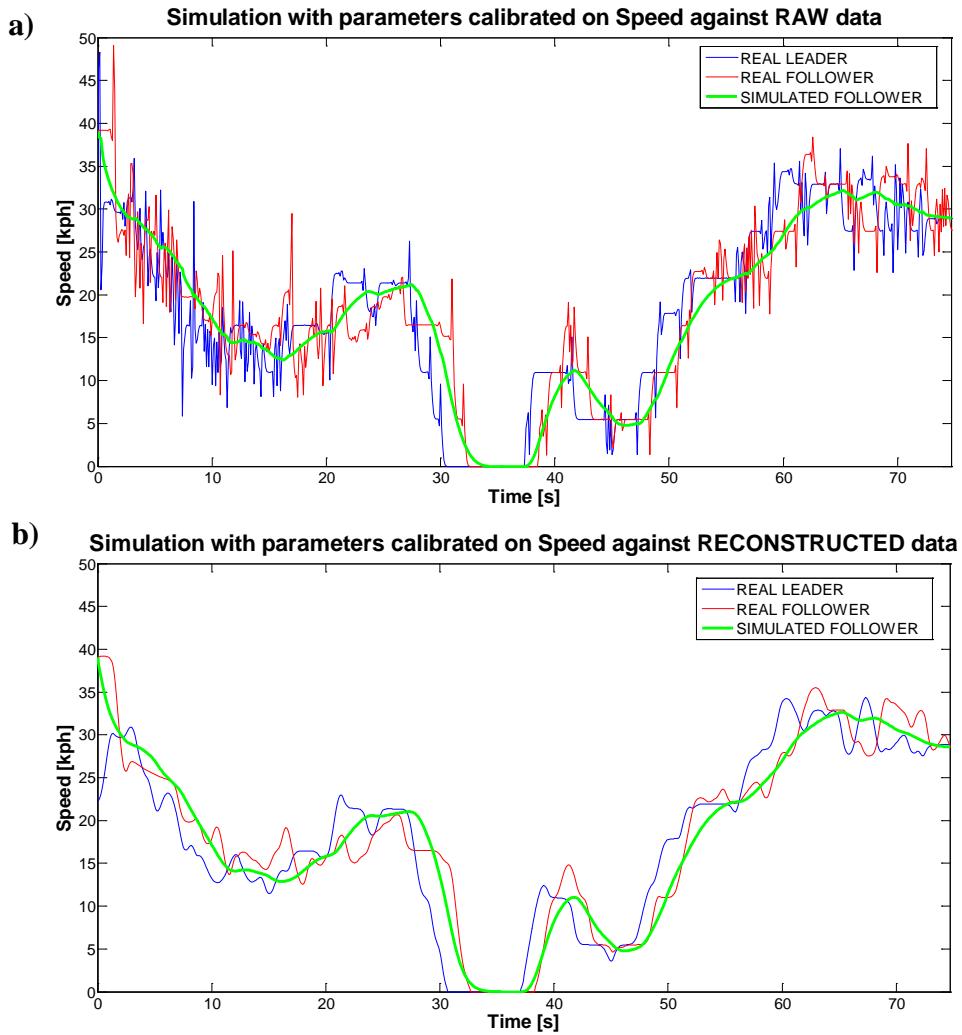
Figure 4.26: Comparison of simulated follower trajectories with parameter estimated on speed against *raw* (a) and *reconstructed* (b) data.
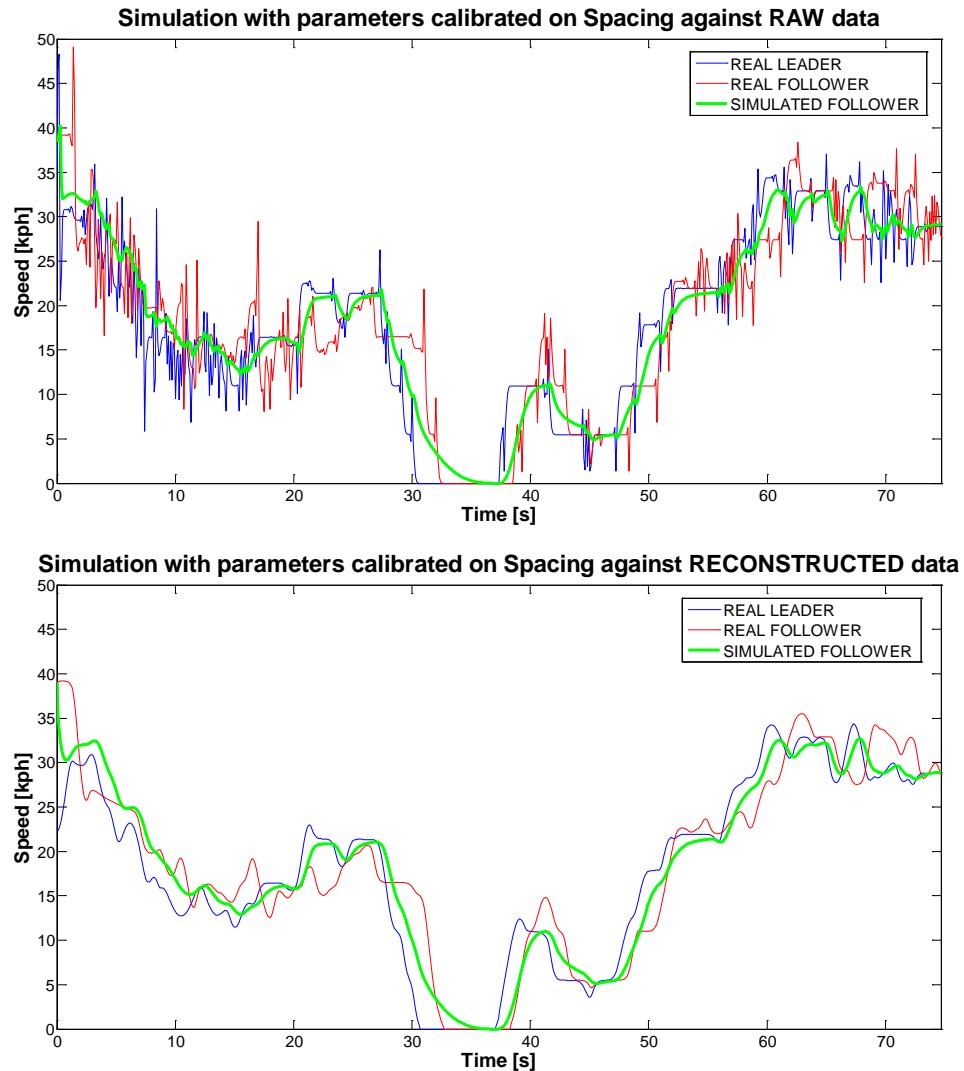
Figure 4.27: Comparison of simulated follower trajectories with parameter estimated on spacing against *raw* (a) and *reconstructed* (b) data.

4.6.2 Lane-Changing model parameters distributions

The three MOBIL model parameters (pf, $\Delta a^{threshold}$, b_{Safe}) were estimated for each individual vehicle in the NGSIM I80-1 dataset (excluding those of type "motorcycle"), against both *raw* and *reconstructed* trajectory data.

As clarified in Appendix D, the benefit of a lane-change choice for a vehicle i at time t depends on the accelerations of the vehicle i and of its follower vehicles in the current and target lanes. Therefore, the estimation of MOBIL model parameters for each individual vehicle is conditioned *i)* to the choice of the acceleration model to simulate car-following behavior, and *ii)* to the adopted value of the car-following model parameters.

Therefore, in this study, to simulate lane-changing choices of vehicle i, we adopted the estimated IDM model parameters for vehicle i and its followers in the current and target lanes. Consequently, MOBIL model parameter estimated distributions were conditioned to the results of IDM model parameters estimation presented in Section 4.6.1. More details on the criteria adopted for the estimation of lane–changing model parameters can be found in Appendix D.

Figures 4.28 and 4.29 reported on the comparison of MOBIL model parameters distributions estimated against *raw* and *reconstructed* data. Figure 4.28 refers to model parameters distributions conditioned to the IDM model parameters estimated on speed, while Figure 4.29 conditioned to IDM calibration on spacing. In the figures, we indicated pf with p, $\Delta a^{Threshold}$ with *minThresholdLC*, and b_{Safe} with *estSafeFollDec*.

From the figures, as in the case of car-following model calibration, the impact of the measurement errors on MOBIL parameter estimation is very limited. Indeed, two-sample Kolgomorov-Smirnov tests revealed that estimated parameter distributions against *raw* and *reconstructed* data were not significantly different at the level of significance of 5%. Further, also model parameter correlation structures did not change between estimates against *raw* and *reconstructed* data, as shown in Figure 4.30 and 4.31.
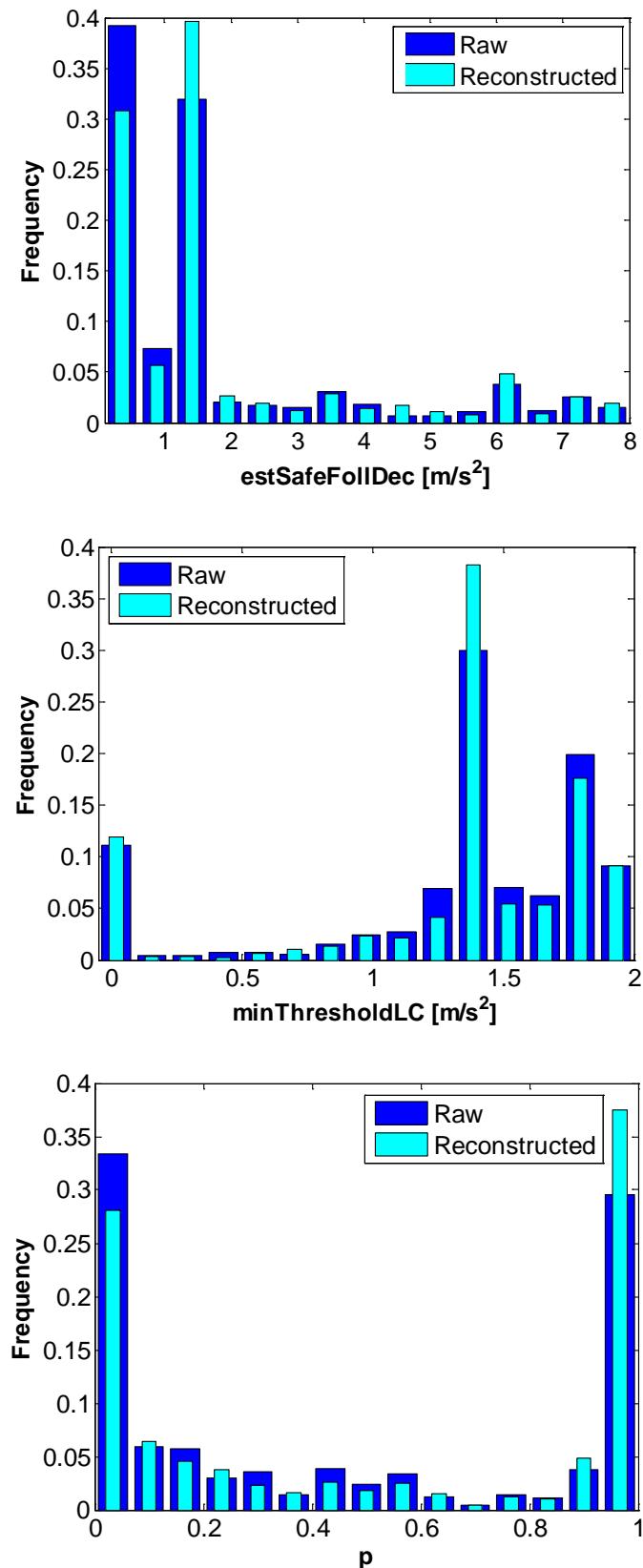
Figure 4.28: Comparison of distributions of MOBIL model parameters estimates conditioned to IDM model parameters calibrated on speed. Blue bars refer to the estimation against *raw* trajectory data, while cyan ones against *reconstructed* data.
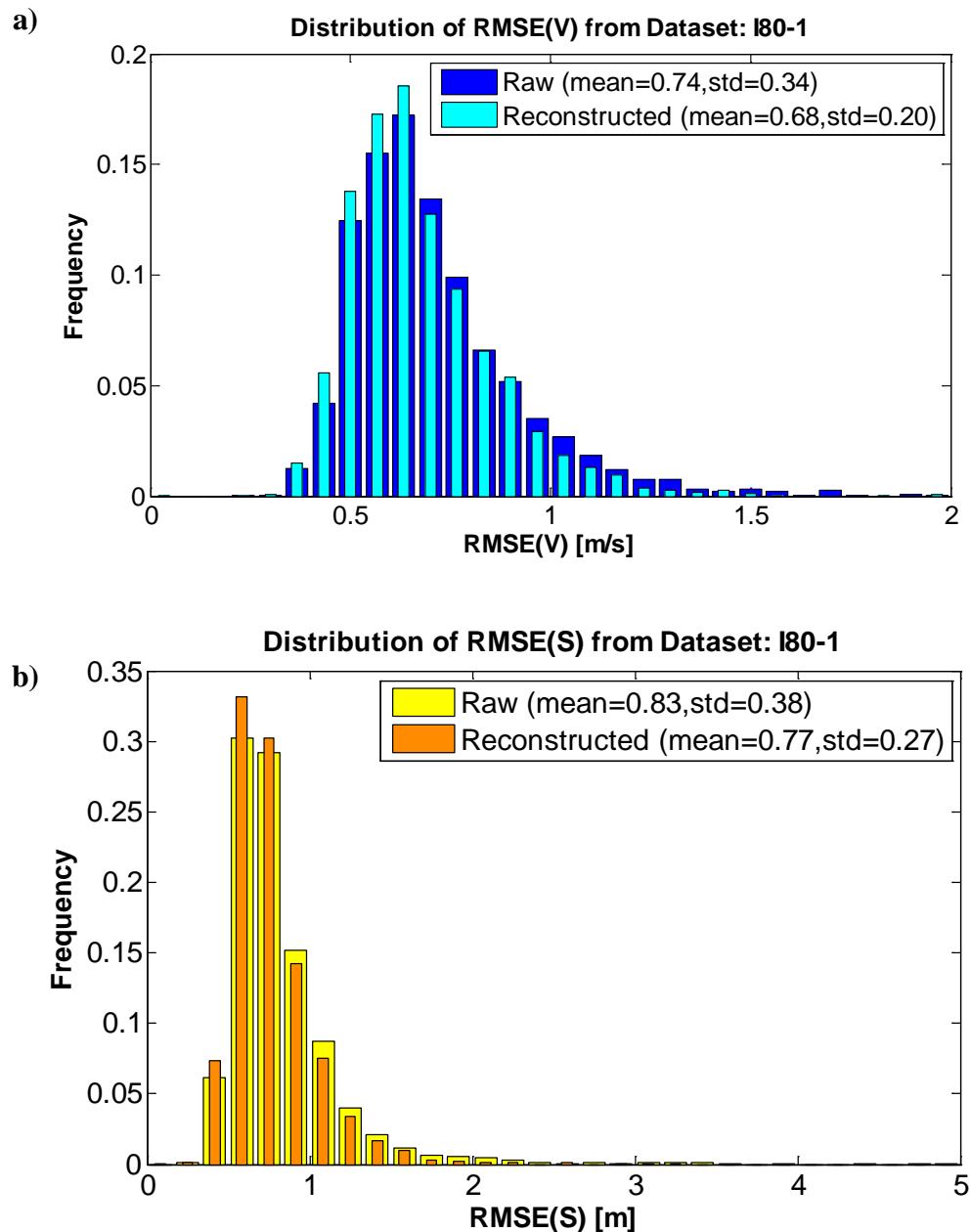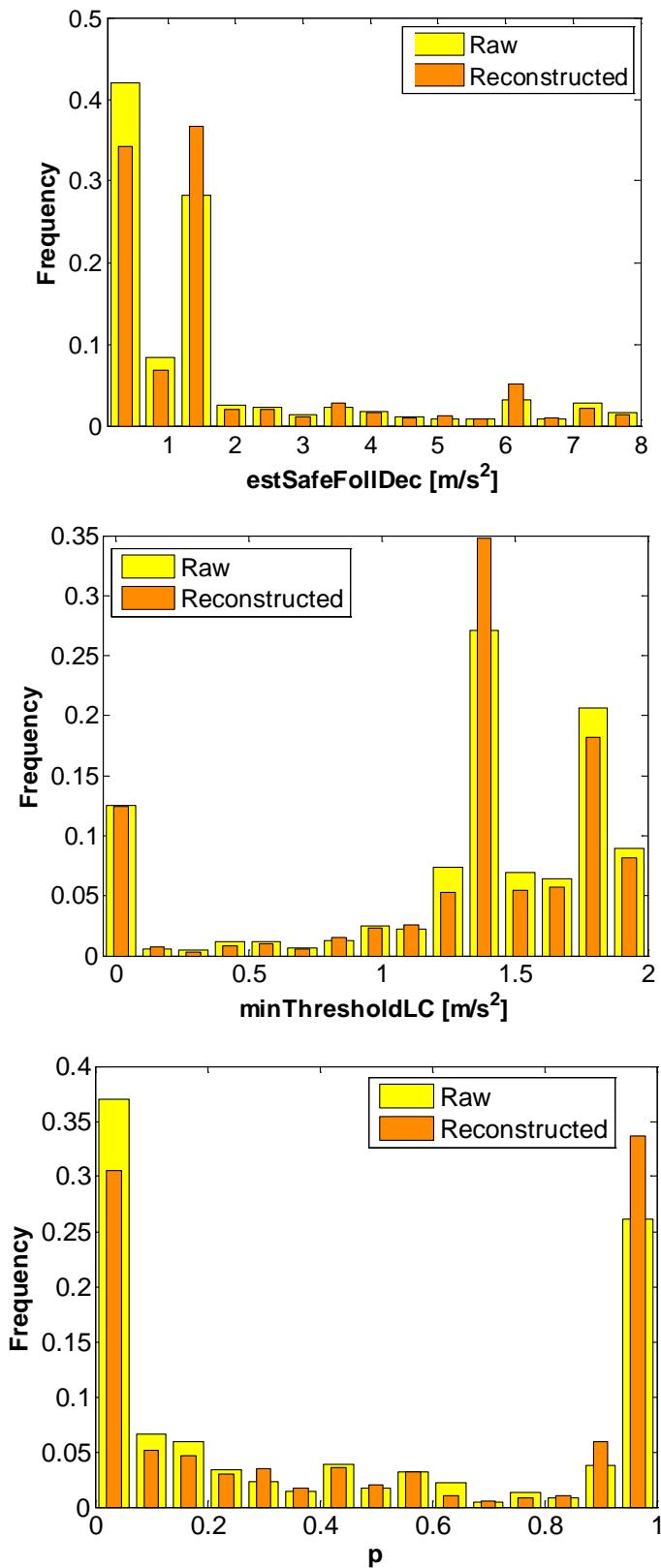
Figure 4.29: Comparison of distributions of MOBIL model parameters estimates conditioned to IDM model parameters calibrated on inter-vehicle spacing. Yellow bars refer to the estimation against *raw* trajectory data, while orange ones against *reconstructed* data.

a)

Correlation Matrix - Estimation against Raw data (IDM parameters estimated on Speed)

b)
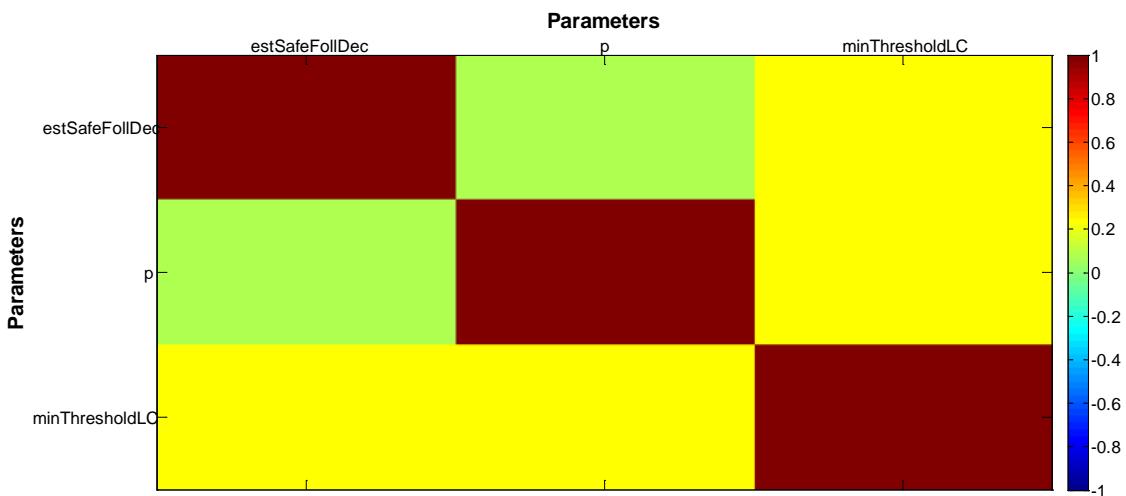
Correlation Matrix - Estimation against Reconstructed data (IDM parameters estimated on Speed)
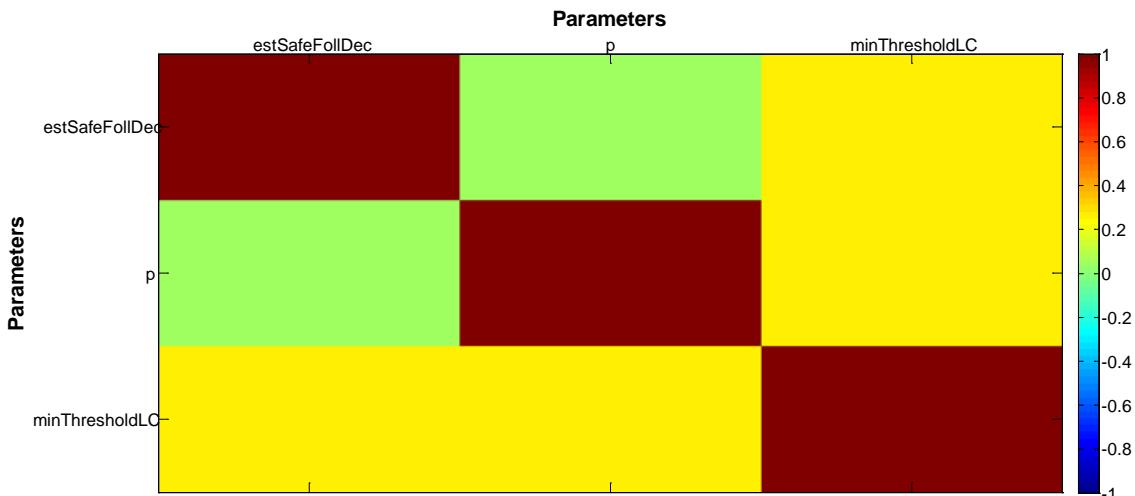
Figure 4.30: Comparison of correlation structures among MOBIL model parameters estimated against *raw* (a) and *reconstructed* (b) data. Related IDM model parameters were estimated on speed.

a)
Correlation Matrix - Estimation against Raw data (IDM parameters estimated on Spacing)
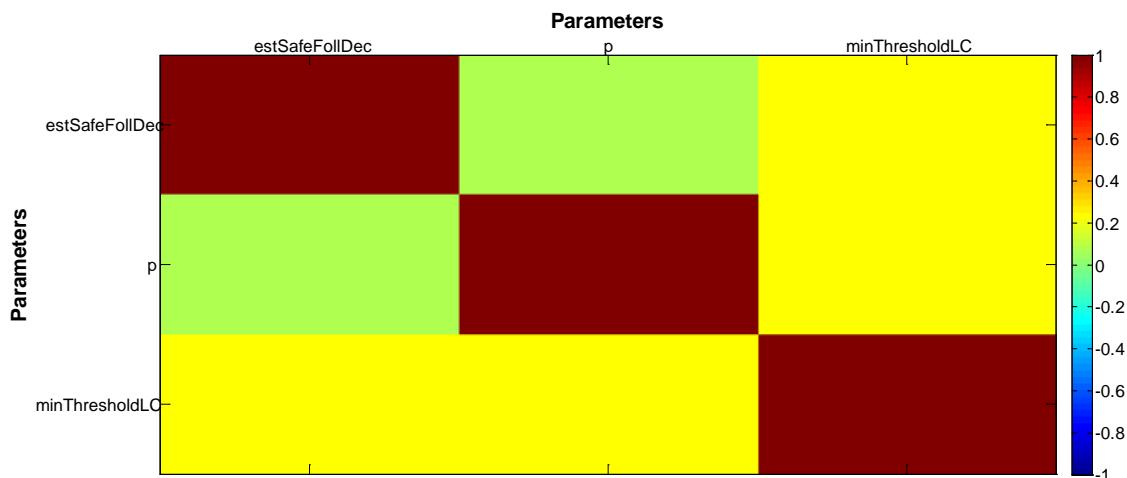
b)
Correlation Matrix - Estimation against Reconstructed data (IDM parameters estimated on Spacing)
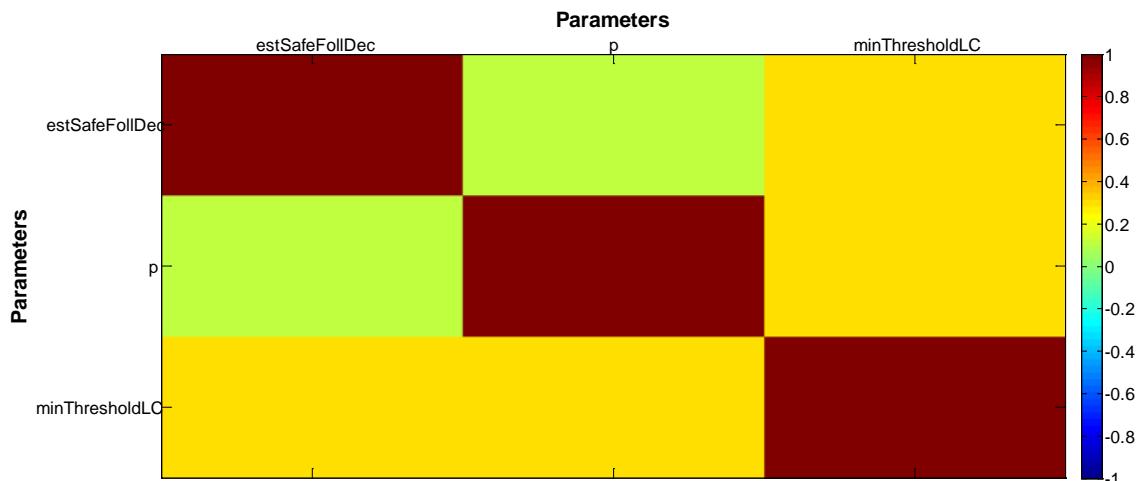
Figure 4.31: Comparison of correlation structures among MOBIL model parameters estimated against *raw* (a) and *reconstructed* (b) data. Related IDM model parameters were estimated on spacing.

4.6.3 Correlation structures among car-following and lane-changing model parameters

Finally, in this section the joint correlation structures among microscopic traffic flow model parameters (car-following and lane-changing models) estimated against *raw* and *reconstructed* data are presented. Figure 4.32 shows the results from IDM model estimation on speed (and, thus, related MOBIL parameters calibration) against *raw* (a) and *reconstructed* data, while Figure 4.33 refers to estimation on speed.
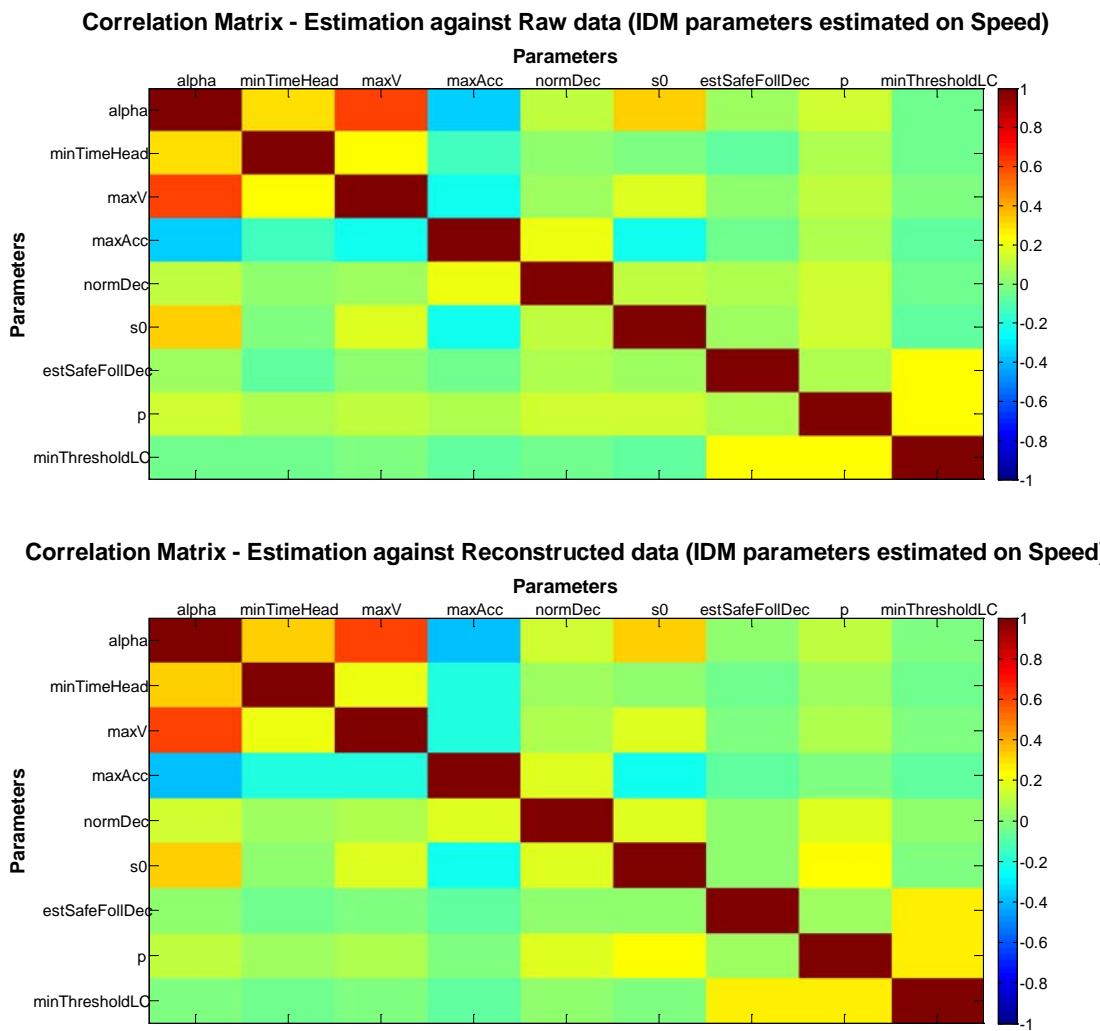
Figure 4.32: Comparison of correlation structures among IDM and MOBIL model parameters estimated against *raw* (a) and *reconstructed* (b) data. IDM model parameters were estimated on speed.
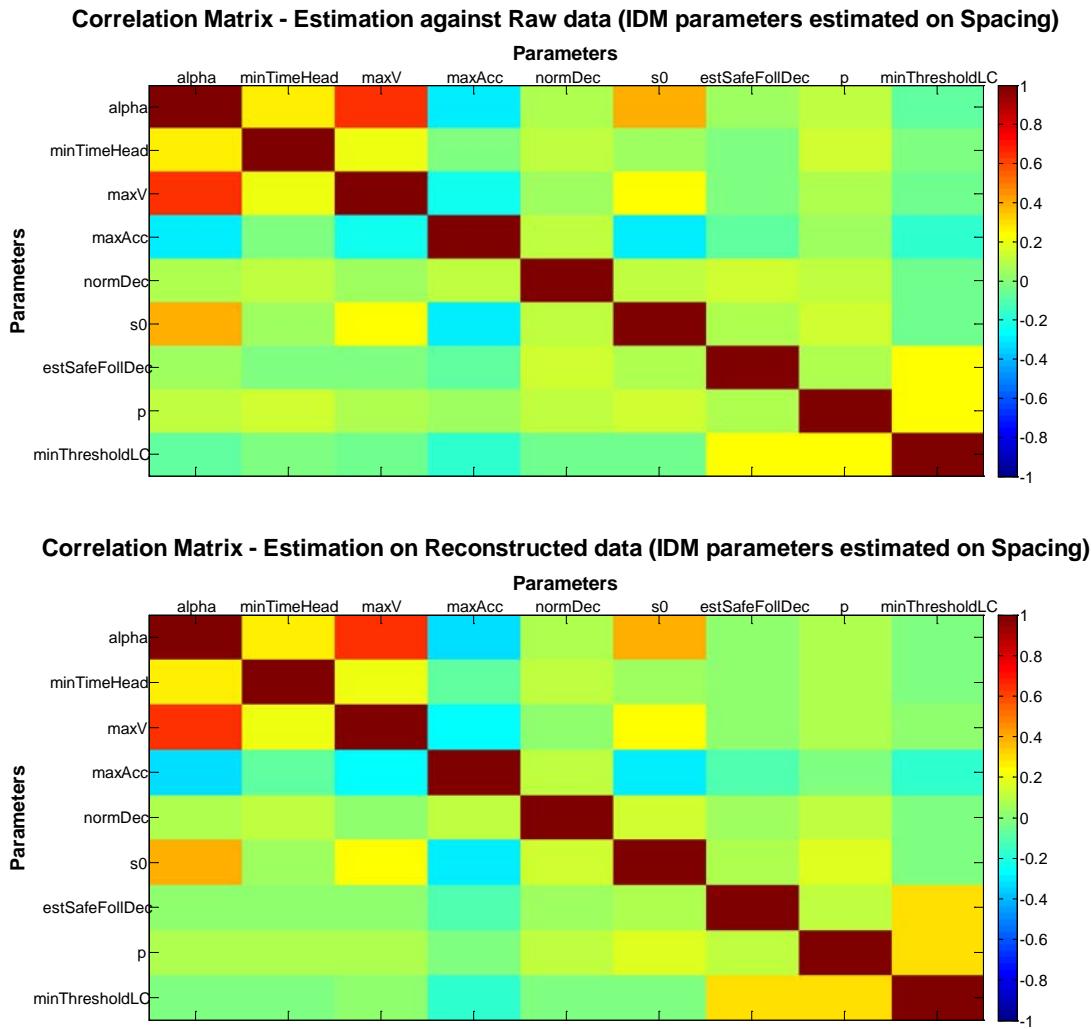
Figure 4.33: Comparison of correlation structures among IDM and MOBIL model parameters estimated against *raw* (a) and *reconstructed* (b) data. IDM model parameters were estimated on spacing.

From the figures, it is clear that the impact of measurement errors in trajectory data on the joint correlation structures of car-following and lane-changing model parameters is practically negligible.

4.7   Summary

In the literature of traffic flow theory, many studies made use of trajectory data to perform experimental analysis and/or support theoretical findings. To this end, data from the NGSIM program are a precious resource, as they depict the traffic behavior of the whole stream over an entire time-space domain. The free availability to the entire traffic community, that gives the opportunity to anyone to reproduce the results or to compare/validate models calibrated against the same data, coupled with the vast amount of gathered data, made this open-source database the most extensively used by researchers.

Despite of the undoubted importance, they were proved to be massively affected by measurement errors in the spatial coordinates of the vehicle, further amplified in the differentiation process when calculating speeds and accelerations. If not properly accounted for, these errors would make NGSIM data not usable for any study on traffic flow theory.

In addition, very few studies attempted to quantify the impact of measurement errors on model estimation.

In the first part of this Chapter, a multi-step procedure for reconstructing vehicles' trajectories is presented. The proposed methodology aimed at eliminating the main inconsistencies and noise from *raw* measurements while preserving *i)* the actual driving dynamics (vehicle stoppages, shifting gears, etc.), *ii)* the *internal* consistency of trajectories (i.e. the consistency among space travelled, speed and acceleration) and *iii)* the platoon consistency (i.e. the actual inter-vehicle spacing).

A comparison between *raw* and *reconstructed* trajectory data is presented. Results from the application to the entire NGSIM I80-1 dataset confirmed, on one hand, the inconsistency of *raw* data, and, on the other hand, the restored consistency in the *reconstructed* data in terms of accelerations' distribution and frequency spectrum, speeds' distribution and inter-vehicle spacings' distribution.

In the second part of the work, provided both *raw* and *reconstructed* trajectory data, we evaluated the impact of real measurement errors on the results of the estimation of car-following and lane-changing model parameters. Results showed that the model operates like a "filter", and the impact of the measurement errors on parameter estimation (and

on their correlation structures) is very limited. These findings are not in line with the results of Ossen and Hoogendoorn (2008a, 2009) where calibration experiments were performed using synthetic data with *normally distributed* error structures added ex-post. A possible explanation of such difference could be due to the substantially different distribution model of the error structure, which is, instead, derived here *empirically* by comparing *raw* and *reconstructed* data.

Based on these findings, in Chapter 6 we will show if and how measurement errors affect the simulation performances of the "aggregate" microscopic traffic flow model.

Chapter 5

Uncertainty Quantification and Sensitivity Analysis of Microscopic Traffic Flow Simulation Models[1]

5.1   Introduction

As pointed out in Chapter 2, model calibration is usually undertook to reduce, in one shot, the impact of both *model* and *parameter* uncertainties, by incorporating them alongside the parametric inputs. In this view, all the commercial software for traffic simulation allow parameters values to be customized by the users in order "to fit" the traffic model to the system at hand.

However, the increasingly high number of parameters in the software, the exponential computational complexity of "black-box" optimization, and the unavailability of dedicated tools in such software, make the automated search for optimal parameter values impracticable for most of the practitioners.

Further elements that hinder calibration are *i)* the improper set up of the calibration problem (for details on this topic, see Chapter 3), *ii)* the quality of measured data (for details on this topic, see Chapter 4), and *iii)* the asymmetry in the importance of model parameters.

---

[1] Regarding the contents of this Chapter, the reader can refer also to Punzo et al. (2014a).

The last issue, in particular, represents both an obstacle to the calibration and a way for its solution. In fact, often, (law-driven over-parameterized) models present a pronounced asymmetry of the parametric inputs in influencing the outputs, with a small subset of parameters accounting for most of the output uncertainty and the others playing little or no role. The inclusion in calibration of such non-influential parameters makes the model response surface flat and the solution search for any optimization algorithm arduous. Therefore, calibration would result much easier if the non-influential parameters of a model could be identified and left out of the calibration itself. Reducing the number of parameters to calibrate would alleviate the computational burden (the CPU time is exponential in the number of parameters) and solve the issue of flat response surfaces (Punzo and Ciuffo, 2011).

These considerations call for methodologies that allow identifying unambiguously the non-influential parameters, and are able to quantify the cost paid – in terms of the model ability to describe reality – of fixing those parameters to arbitrary values. Such methodologies belong to the area of sensitivity analysis, generally intended as "*the study of how the uncertainty in the output of a mathematical model or system can be apportioned to different sources of uncertainty in its inputs*" (Saltelli et al., 2004).

Factor fixing setting, in particular, is the name generally given to the specific setting in which the analysis is framed to answer the question of which parameter can be fixed at whatever value without affecting the output uncertainty. For details on other possible settings, please refer to Saltelli et al. (2004).

The main objective of this study, therefore, is to verify for a well-known microscopic traffic flow simulation model, whether it is possible to reduce the number of parameters to calibrate without sensibly affecting the capacity of the model to reproduce the true output variance. To this aim, variance-based sensitivity analysis is applied, in a *factor fixing* setting, to the Intelligent Driver Model (IDM, Treiber et al., 2000).

Since results of a sensitivity analysis are conditioned to the values of the *fixed* inputs, one could argue that the ranking of importance of car-following model parameters is specific to the selected leader/follower vehicle trajectory, which is respectively used for model simulation and estimation error calculation.

Therefore, in the present analysis, we considered also the input trajectories as *uncertain*, and the investigation has been extended by including all the *reconstructed* trajectories of the NGSIM I80-1 dataset, as detailed in Chapter 4.

Recalling results from Chapter 4, traffic conditions are moderately congested and the trajectories comprehend a wide range of dynamics including stops.

Besides the robustness of the analysis as regards to the *factor fixing* setting, the inclusion of more than two thousand input trajectories allowed us to investigate the model against a significant variety of driver behaviors. To the best of our knowledge this is the first time that such a comprehensive analysis is carried out on a traffic flow model. In addition, though applied to the IDM in this study, the methodology is absolutely general.

The Chapter is organized as follows. In Section 5.1, a review of variance-based techniques for global sensitivity analysis is presented, followed by the description of the IDM model. In Section 5.2., the methodology applied throughout the work is described, while the results of the application are presented in Section 5.3. Brief conclusions end the Chapter.

5.2   Background

In this section, a review of variance-based techniques for global sensitivity analysis in *factor fixing* setting is presented in Section 5.2.1., followed by a synthetic review of the IDM car-following model in Section 5.2.2.

5.2.1 *Variance-based sensitivity analysis in factor fixing setting*

The sensitivity analysis technique applied in this work belongs to the family of the so-called variance-based techniques, which were first employed by Cukier et al. (1973), then generalized by Sobol' (1993, 2007) with a Monte Carlo-based implementation of the concept, and finally enhanced by Saltelli et al. (2010) for computation efficiency. For a detailed explanation on the topic the reader can refer to Saltelli et al. (2010), and to Appendix A.

Those methods were proved to overcome most of the limitations of other common adopted approaches, such as One-At-Time (OAT) analysis, differential methods and

regression/correlation analysis. On this topic, the reader may refer to Saltelli et al. (2008).

In the following a synthetic description of the method is provided, while implementation details and further information are reported in Appendix A.

The basic idea of the method derives from the well-known variance decomposition formula (Mood et al., 1974). Given a model $Y = f(X_1, X_2, ..., X_k)$, where $X_i \; \forall i \in [1,k]$ are the input stochastic variables, i.e. the *uncertain factors*, and Y the *output* stochastic variable, the variance of the output can be decomposed as follows:

$$V(Y) = V_{X_i}\left(E_{\overline{X}_{\sim i}}(Y/X_i)\right) + E_{X_i}\left(V_{\overline{X}_{\sim i}}(Y/X_i)\right) \tag{5.1}$$

where X_i is the i-th factor and $\overline{\mathrm{X}}_{\sim i}$ denotes the vector of all factors but X_i. The first component $V_{X_i}\left(E_{\overline{X}_{\sim i}}(Y/X_i)\right)$ is the variance-based *first-order* effect. The meaning of the inner expectation operator is that the mean of Y is taken over all possible values of $\overline{\mathrm{X}}_{\sim i}$ while keeping X_i fixed, the outer variance being taken over all possible values of X_i. From a visual perspective, if we plot the values of the output Y against the values of the i-th input factor (i.e. in a scatter plot), and then we cut the plane into thin vertical slices, it is possible to calculate the mean value of the output Y in each slice, that is $E_{\overline{X}_{\sim i}}(Y/X_i)$. Therefore, the first-order effect is the variation over the slices of the expected value of Y within each slice. The associated sensitivity measure, called "first order sensitivity index", is equal to the first-order effect normalized over the total (or unconditional) variance:

$$S_i = \frac{V_{X_i}\left(E_{\overline{X}_{\sim i}}(Y/X_i)\right)}{V(Y)} \tag{5.2}$$

It can be interpreted as the portion of the output variance which is due to the variation of the input factor X_i. In fact, it measures the first-order (additive) effect of the i-th factor on the model output. Therefore, the first-order effect captures only the "stand-alone" effect of the input factor on the model output. However, for non-additive models, the input factor X_i contributes to the output variance also in its interaction with the other input factors. In other words, the joint variation of X_i with all (or some of) the input factors may influence the variation of the output. This influence is called higher-order effect related to X_i. The sum of the first-order and higher-order effects for all the input

factors explains all the output variance. Therefore, when the terms are normalized over the unconditional variance such summation is equal to 1:

$$\sum_{i=1}^{k} S_i + \sum_{i=1}^{k}\sum_{\substack{j=1\\j\neq i}}^{k} S_{i,j} + \sum_{i=1}^{k}\sum_{\substack{j=1\\j\neq i}}^{k}\sum_{\substack{l=1\\l\neq\{i,j\}}}^{k} S_{i,j,l} + ... + \sum_{i=1}^{k}\sum_{\substack{j=1\\j\neq i}}^{k}...\sum_{\substack{m=1\\m\neq\{i,...\}}}^{k} S_{i,j,...,m} = 1 \tag{5.3}$$

where $\sum_{i=1}^{k} S_i$ is the contribution of all the first-order effects, while $1 - \sum_{i=1}^{k} S_i$ is the contribution of all the interaction effects across all the input factors. It is worth noting that in case of additive models there are no interaction effects and $\sum_{i=1}^{k} S_i = 1$, while in case of non-additive models it results: $\sum_{i=1}^{k} S_i < 1$.

According to this decomposition, the number of higher order effects to calculate would be very high (i.e. 2^k-1-k with k the number of factors). Therefore, to quantify the *total effect* of a factor, the so-called "total sensitivity index" is introduced:

$$ST_i = \frac{E_{\overline{X}_{\sim i}}\left(V_{X_i}\left(Y/\overline{X}_{\sim i}\right)\right)}{V(Y)} = 1 - \frac{V_{\overline{X}_{\sim i}}\left(E_{X_i}\left(Y/\overline{X}_{\sim i}\right)\right)}{V(Y)} \tag{5.4}$$

That is the sum of the first-order effect of X_i and of all the higher-order effects that involve X_i. As higher-order effects are computed more times, i.e. in the *ST* of each factor involved in the interaction (e.g. $S_{i,j} = S_{j,i}$ is included in both ST_i and ST_j) it results: $\sum_{i=1}^{k} ST_i \geq 1$, where the equality holds only for perfectly additive models (for which $S_i = ST_i$, $\forall\ i = 1,.., k$.)

From a computational point of view, the calculation of the indices can be performed within a Monte Carlo Sampling (MCS) framework, where different sampling strategies can be adopted (see Section 5.3).

Following the above considerations, it is clear that the total sensitivity index is the appropriate measure in a *factor fixing* setting, where the question to be answered is: "which are the factors that can be fixed at whatever value without affecting the output uncertainty?". Indeed, $ST_i = 0$ is a necessary and sufficient condition for X_i to be non-influential.

PROOF. If $ST_i = 0$ for factor X_i, then $E_{\overline{X}_{\sim i}}\left(V_{X_i}\left(Y/\overline{X}_{\sim i}\right)\right) = 0$. As the variance can only be positive, the above relation implies that $V_{X_i}\left(Y/\overline{X}_{\sim i} = \overline{x^*}_{\sim i}\right)$ is identically zero for any value of $\overline{x^*}_{\sim i}$. That is there is no point in the hyperspace of $\overline{X}$ where X_i has an effect.

5.2.2 *The Intelligent Driver Model (IDM)*

The car-following model analyzed in this work is the Intelligent Driver Model (IDM), which belongs to the class of social force models (Treiber et al., 2000). The social force concept states that the driving behaviour is driven by a sum of social forces, including both the force that pushes the vehicle to reach the driver's desired speed, and the interaction force that compels the vehicle to keep a suitable distance from the leading vehicle (Wang et al., 2010). For further details on the model, please refer to Treiber et al. (2006). The model formulation is the reported in Eq. 5.1:

$$a_f(t) = a_f^{Max} \cdot \left\{ 1 - \left[\frac{v_f(t)}{V_f^{Max}}\right]^{alpha} - \left[\frac{\Delta S^*(t)}{\Delta s(t)}\right]^2 \right\}$$

$$\Delta S^*(t) = \Delta S_0 + max\left\{ \Delta S_1 \cdot \sqrt{\frac{v_f(t)}{V_f^{Max}}} + T \cdot v_f(t) + \frac{v_f(t) \cdot \left[v_f(t) - v_l(t)\right]}{2 \cdot \sqrt{a_f^{Max} \cdot \left|b_f\right|}}, 0 \right\}$$

(5.5)

where:

- $v_f(t)$ [m/s] and $a_f(t)$ [m/s²] are, respectively, the follower's speed and acceleration at time *t*;

- V_f^{Max} [m/s] is the follower's maximum desired speed (*default value*: 33.3);

- a_f^{Max} [m/s²] is the follower's maximum acceleration at rest when the distance from his/her leader is much bigger than the distance ΔS_0 (*default value*: 0.73);

- b_f [m/s²] is a sort of deceleration rate between normal and emergency conditions (Treiber et al., 2000) (*default value*: 1.67);

- $v_l(t)$ [m/s] is the leader's speed at time *t*;

- $\Delta s(t)$ [m] is the rear end-front bumper distance between the follower and his/her leader, calculated as follows: $\Delta s(t) = x_l(t) - L_l(t) - x_f(t)$, where $x_l(t)$ and $x_f(t)$ [m], are the positions at time t of the leader's and the follower's front bumpers, respectively, and $L_l(t)$ [m] is the physical length of the leader's vehicle at time t. It is worth noting that it is time-dependent as the leader vehicle can change over time;

- $\Delta S^*(t)$ [m] is the rear end-front follower's desired distance from the leader;

- ΔS_0 is the rear end-front follower's desired distance from the leader at stop [m] (*default value*: 2);

- ΔS_1 [m] is a non-zero parameter necessary for features requiring an inflection point in the equilibrium flow-density (Treiber et al., 2000): in this study, we fixed its value to zero;

- T [s] is the minimum time headway between leader and follower (*default value*: 1.6);

- *alpha* is an additional model parameter (*default value*: 4).

Default values reported in parenthesis are those suggested in Treiber et al. (2000).

It is worth noting that the *max(...)* operator in Eq. 5.5 is necessary to avoid that the follower's desired distance from the leader becomes lower than ΔS_0, for negative speed differences (i.e. $v_f(t) < v_l(t)$).

5.3   Methodology

In this section, the methodology developed in the present study is described. Section 5.3.1 presents the procedure for uncertainty quantification and sensitivity analysis in a Monte Carlo framework, while Section 5.3.2 details the implementation of the adopted *factor fixing* setting. Finally, Section 5.3.3 recalls the framework for car-following model parameters estimation.

5.3.1 *Uncertainty Quantification and Sensitivity Analysis in Monte Carlo framework*

In Figure 5.1, the Monte Carlo framework adopted in this work for the calculation of the sensitivity indexes is outlined.
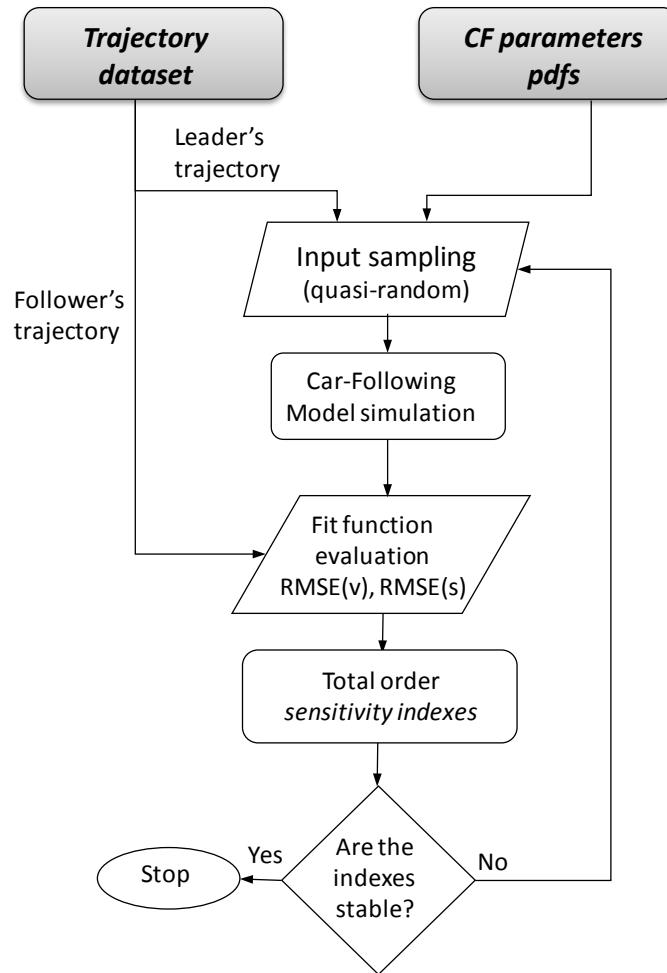
Figure 5.1: Flowchart of the Monte Carlo framework for the calculation of sensitivity indexes.

Once the model inputs are drawn by means of a sampling scheme, the traffic model is evaluated and the distance between the measured and the simulated trajectory is calculated in terms of the Root Mean Square error of the instantaneous speed, i.e. RMSE(v), or inter-vehicle spacing, i.e. RMSE(s). The process is iterated until the number of evaluations is sufficient for the calculated indexes to be stable (for details, see Appendix A).

The Sobol' low-discrepancy quasi-random sequences were adopted here for Monte

Carlo sampling, as approximated formulas for the indices calculation are available only for such sampling strategy. In particular, the formula by Saltelli et al. in (2010) for the computation of *first order* indices, and that by Jansen (1999) for the computation of the *total order* indices were applied in this work. For details, see Appendix A.

The peculiarity of the framework depicted in Figure 5.1 with respect to previous works (Punzo and Ciuffo, 2011) is that, not only the model parameters, but also the input trajectories were chosen as uncertain factors of the analysis. That is, not only parameter values, but also the leader's and follower's trajectories are sampled at each iteration.

In fact, it was conjectured that the variance of the model error explained by model parameters was not independent by the input trajectories, but could sensibly vary with it. If such hypothesis was verified the analysis would also return precious indications on the behavior of the model when facing with different kinematic inputs.

Therefore, the identification number of each of the 2035 leader/follower couples of vehicle trajectories from the NGSIM I80-1 dataset was set as an additional factor with the name of *PairID*. Conversely, the six parameters of the IDM were assumed uniformly distributed over the following intervals: $alpha \in [0.5,\ 10]$, $T \in [0.1,\ 3]$, $V_f^{Max} \in [15.6,\ 29.0]$, $a_f^{Max} \in [0.5,\ 10]$, $b_f \in [0.5,\ 10]$, $\Delta S_0 \in [0.1,\ 5]$.

The assumption of uniform distribution being customary in absence of a priori information on the parameter probability density functions. Indeed, such hypothesis could have an impact on the analysis results but it was not investigated in this study.

On the other hand, as the amplitude of intervals affects the analysis results – too tight intervals fictitiously limit the model error variance whereas too large ones, including unrealistic parameter values, increase it unlikely – the interval limits were set in a trial-and-error manner by leaving out of the intervals the parameter values returning unlikely high variance of the outputs, from the visual observation of the scatter-plots. For this reason also non-physical values were kept (e.g. acceleration up to 10 m/s^2).

As a "measure of uncertainty" to base the sensitivity analysis on, the RMSE between the measured and the simulated trajectory was chosen (i.e. the higher the influence of a parameter, the higher the error in reproducing the observed trajectory). Such statistic was calculated both on the speed and on the spacing.

A total number of 2^{17} model evaluations was necessary in order to have clearly stable

sensitivity indices.

5.3.2 Factor Fixing setting

In Section 5.2.1 it was shown that an appropriate measure for such a setting is the "total sensitivity index", ST_i. In fact, $ST_i = 0$ is a necessary and sufficient condition for the factor X_i to be non-influential (for details, see Section 5.2.1).

In practical applications, however, a threshold on ST_i higher than zero is generally set, under which the parameter is considered non-influential. The choice of the threshold value depends on the approximation accepted by the analyst for the study at hand. In this work a value of 2% was considered as an acceptable threshold.

The study conjecture is therefore that parameters of the IDM model with a total sensitivity index lower than 2%, i.e. explaining less than the 2% of the output unconditional variance, could be fixed at any value without affecting ("too much") the uncertainty in the model outputs.

5.3.3 Calibrations of "full" and "reduced" models

Once the non-influential parameters were identified, a *reduced* model version was obtained by fixing such parameters at arbitrary values; in particular, those suggested in Treiber et al. (2000), and reported in Section 5.2.2, were adopted here.

Both the *reduced* model and the *full* model versions were calibrated against all the trajectories in the *reconstructed* NGSIM I80-1 dataset. Then, in order to verify the study conjecture, the performances resulting from the two series of calibrations were compared, both in terms of the goodness of fit values of the calibrated model and of the computational effort required.

IDM model parameters were calibrated for each individual vehicle (excluding those of type "motorcycle") following the approach reported in Chapter 3. Calibrations experiments were run both on speed and on spacing, in order to analyze the effect of using different measures of performance (MoPs) on the estimation. The goodness of fit functions (GoF) were the RMSE(v) and the RMSE(s), while the optimization algorithm was the OptQuest Multistart (for details, see Chapter 3). Upper and lower bounds for *full*

model parameters calibration were set to the following values: : $alpha \in [0.1, 20]$, $T \in [0.1, 5]$, $V_f^{Max} \in [15.6, 40.0]$, $a_f^{Max} \in [0.1, 15]$, $b_f \in [0.1, 15]$, $\Delta S_0 \in [0.1, 10]$. The calibration of the *reduced* model was performed using the same bounds above, though only for the influential parameters (the others being fixed). For a discussion on the robustness of the adopted calibration procedure, please, refer to Chapter 6.

5.4 Results

In this section, results from the application of variance-based global sensitivity analysis techniques are first presented (Section 5.4.1). Successively, Section 5.4.2 presents the comparison of *full* and *reduced* model calibration results.

5.4.1 Uncertainty Quantification and Sensitivity Analysis

Figure 5.2. presents the uncertainty quantification of the model outputs, from the Monte Carlo-based simulation framework adopted (Figure 5.1).
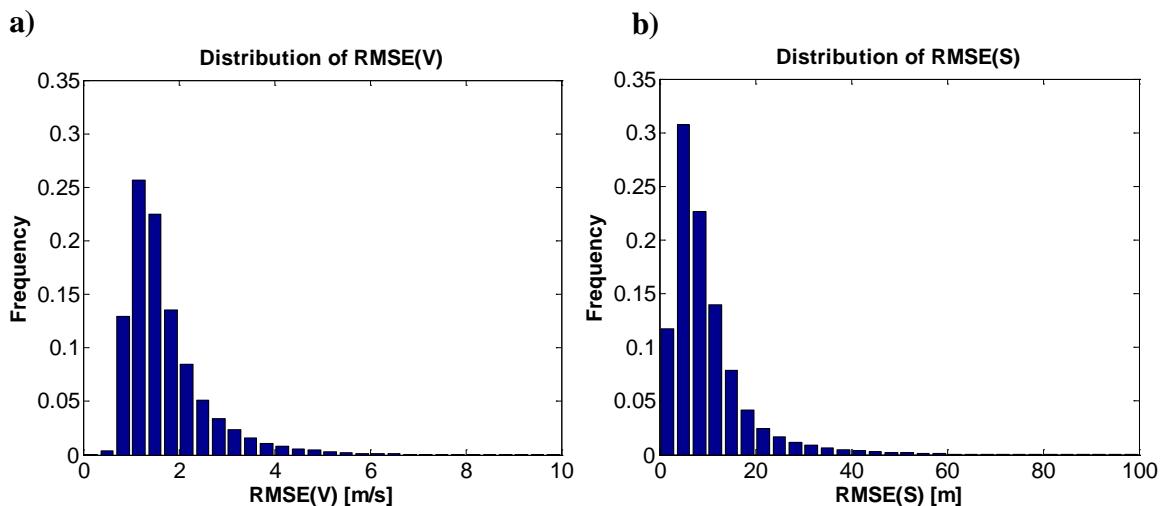
Figure 5.2: Distribution of simulation errors from the Monte Carlo-based simulation. (a) refers to RMSE on speed, while (b) on inter-vehicle spacing.

Though nicely positive-skewed, these distribution of the uncertainty in the simulation errors do not provide any additional information on the degree of importance of model parameters with regards to the error variance.

Conversely, the scatter plots of the two measures of uncertainty against each input factor are suitable for the scope, for a preliminary screening investigation (Figure 5.3). Figure 5.3(a) relates to the RMSE(v), while Figure 5.3(b) to the RMSE(s). In the figure, we indicated T with *minTimeHead*, V_f^{Max} with *maxV*, a_f^{Max} with *maxAcc*, b_f with *normDec* and, ΔS_0 with *s0*.
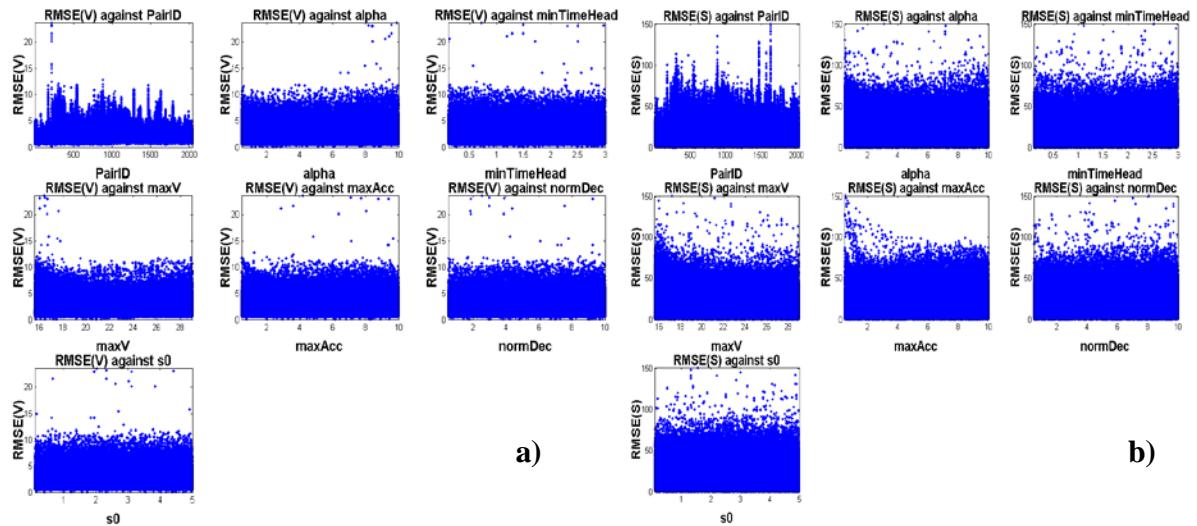
Figure 5.3: Scatter plots of the model output against input factors. (a) refers to the RMSE(v), (b) to the RMSE(s).

The visual inspection of the scatter plots is an important operation, complementary to the results of the sensitivity analysis. In general, scatter plots can be used to investigate (mainly qualitatively) the behavior of a model.

Following the physical interpretation given in Section 5.2.1, the variability of the output in the space of each input factor gives graphical information of the first-order effect of the input factor. In other words, the existence of a clear "shape" or "pattern" in the points (i.e. a not uniform distribution of Y-points over the factor X_i) identifies an important factor, while a uniform cloud is a symptom (though not a proof) of a non-influential one.

From the visual inspections of scatter plots in Figure 5.3 a clear pattern in the variance of the RMSE can be appreciated only for the *PairID*. This shows that the input leader/follower trajectories are an influential factor. Concerning the other factors instead, the scatter plots are not meaningful as the high number of points per plot could hide

possible patterns. In this case the influence of a factor can be thoroughly evaluated only by means of the sensitivity indices (Figure 5.6 and Table 5.1, presented later).

Coming back to the scatter plots of the *PairID*, for any given value of the *PairID*, that is for any given trajectory, each Y-point represents the RMSE of the model for a specific parameter combination. We can therefore think of the variance of the RMSE, conditional on the trajectory, as a measure of the "risk" of not calibrating the model. The higher the RMSE variance, the higher the chance that the model yields high RMSE for non-optimal parameter combinations. On the other hand, a more in-depth analysis of the lower boundary reveals that also the minimum error (the one that corresponds to "best" parameter combination), sensibly varies with the input trajectory (see Figure 5.3). In other words, the model is not able to reproduce all the observed follower trajectories with the same degree of fidelity.

The reason why model performances vary so much with the leader trajectory could depend on many causes. For instance, the un-modeled details of the phenomenon could be significant for some drivers and not for others. The different length of the trajectory could also explain such variability.

The last guess, in particular, was tested in Figure 5.4, where the trajectories are ordered by their duration and not by their *PairID* (points were also coloured according to the duration intervals reported in the legend). The result confirms the guess surprisingly. On the one hand, in both the plots, the RMSE variance – the average height of each colored vertical stripe – decreases as duration increases. Thus, the "risk" of not calibrating the model is lower for longer (in time) trajectories where, the longer exposition to car-following dynamics prevents, even an uncalibrated model, to yield very high errors. On the other hand, looking at the lower boundary of the RMSE on the spacing (zoomed in Figure 5.5), the minimum error value increases as the trajectory duration increases. Therefore, calibration is expected to provide higher errors on longer trajectories. The reason why such condition holds apparently for spacing but not for speed (Figure 5.4 vs. 5.5) can be explained by observing that the spacing is an integral measure so that errors are not compensated rapidly and tend to cumulate.
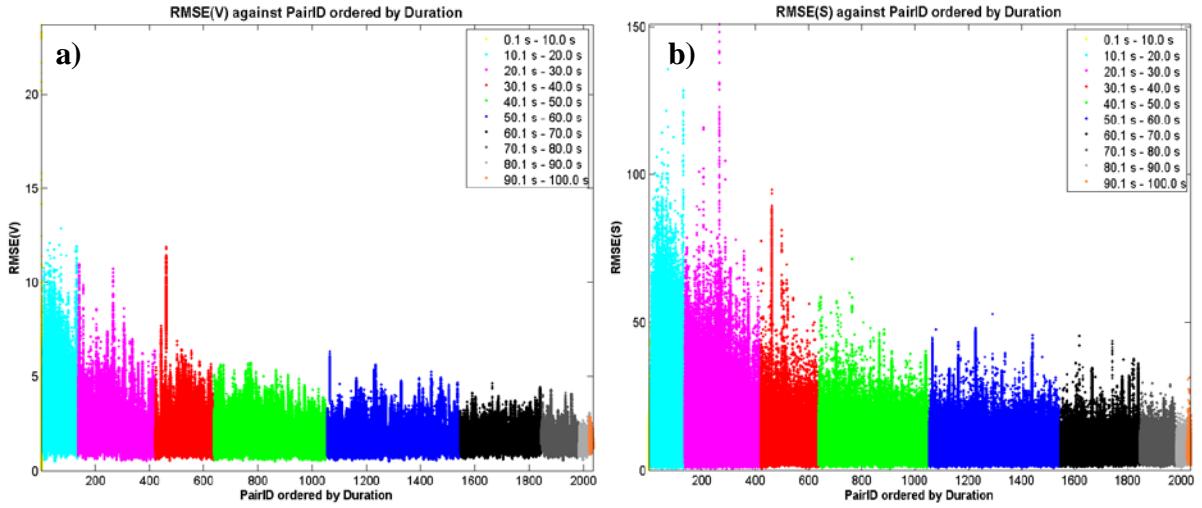
Figure 5.4: Scatter plots of the RMSE against the PairID. (a) refers to the RMSE(v), (b) to the RMSE(s). The PairID values were ordered by duration of the trajectory on the color scale.
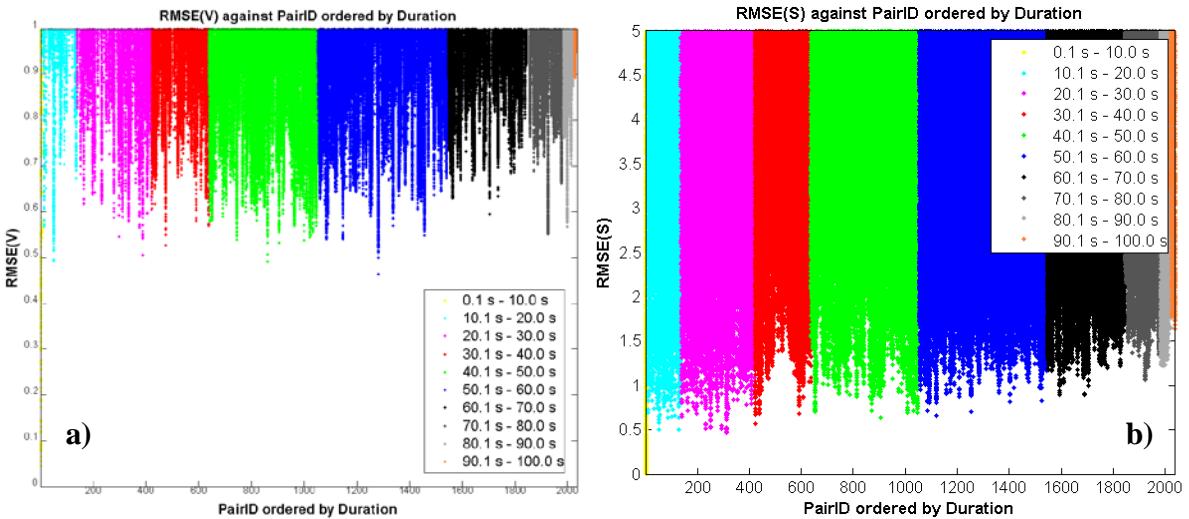
Figure 5.5: Zoom on the lower boundary of the scatter plots reported in Figure 5.4. (a) refers to the RMSE(v), (b) to the RMSE(s).

Moving to the analysis of the sensitivity indices, Figure 5.6 reports the values of the *first order* and *total* sensitivity indices (jointly with their 90% confidence intervals) of all the input factors, related to both the RMSE(v) and the RMSE(s). The numerical values are reported, instead, in Table 5.1.
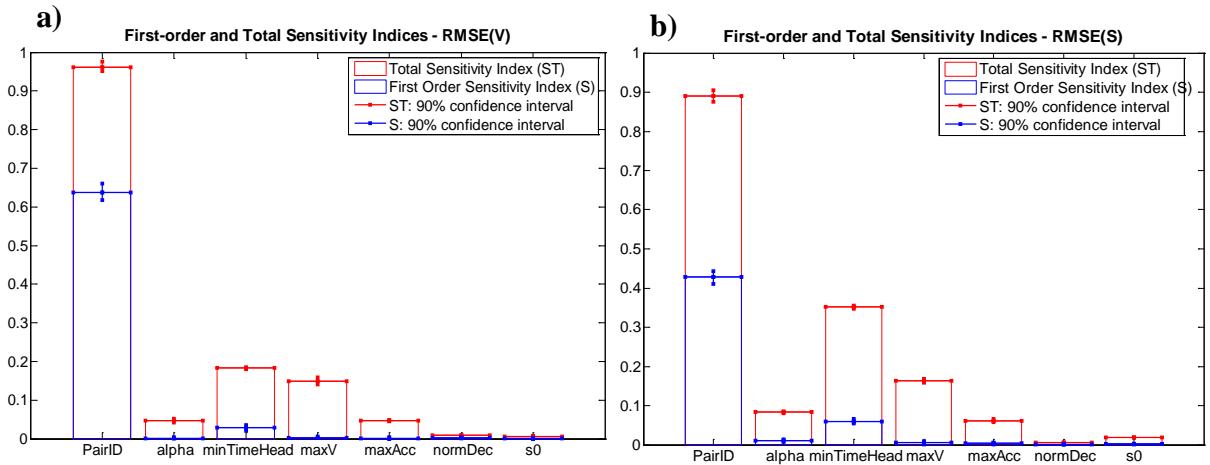
Figure 5.6: *First order* and T*otal* Sensitivity Indices of input factors with respect to (a) the RMSE(v) and (b) the RMSE(s).

Table 5.1: *First order* and *Total* Sensitivity Indices of input factors.

<div align="center">

TABLE I
FIRST-ORDER AND TOTAL SENSITIVITY INDICES

</div>

RMSE(V)		
Parameter	*First-Order (S)* [%]	*Total (ST)* [%]
PairID	63.75	96.19
alpha	0.13	4.67
minTimeHead	2.87	18.27
maxV	0.18	14.87
maxAcc	0.16	4.71
normDec	0.24	0.84
s0	0.03	0.63
RMSE(S)		
Parameter	*First-Order (S)* [%]	*Total (ST)* [%]
PairID	42.88	89.09
alpha	1.11	8.35
minTimeHead	6.01	35.23
maxV	0.58	16.32
maxAcc	0.40	6.15
normDec	0.12	0.55
s0	0.17	1.85

The variance of the output variables is explained by the *PairID* ("stand-alone" and in combination with the model parameters). Comparing the value of the variance explained by the *PairID* (63.75 or 42.88) with those of all the other factors, it comes out clearly that the model performances are very low if the model is not calibrated against the single

trajectory (this does not want to be a critique to the IDM, as it holds for all the car-following models). In other words, if the validity of a car-following model is judged through the capacity of reproducing a trajectory, individual calibration on that trajectory is necessary.

Regarding the model parameters, the first-order effects of all the parameters explained less than the 10% of the total variance, meaning that their effect on the output variable is due to their mutual interactions: in this case, the total sensitivity indices are a measure of the higher-order effects.

According to a *factor fixing* setting, the total sensitivity measures define the rank of influence of model parameters. In this view, the minimum time headway with its interaction effects explained 18% and 35% of the total output variance of the RMSE(v) and the RMSE(s), respectively, followed by the maximum desired speed with 14% and 16%. The remaining portion of error variance is finally explained by alpha (5% and 9%) and maxAcc (5% and 6%). The interaction effects of the remaining model parameters (normDec and s0), instead, explained about 2% of the output variance. This means that fixing them to any value it is expected to reduce the capability of the model to explain about the 2% of the error variance This was deemed to be an acceptable threshold for the parameters to be considered not influential in calibration.

Following the above considerations, in the second part of the work, we tested the impact on the calibration performances of fixing the two not influential parameters to values commonly used in the literature.

By *full model* estimation we indicated the calibration experiment where all the 6 IDM parameters were estimated, and by *reduced model* estimation the calibration experiment where the minimum time headway (T), the maximum desired speed (V_f^{Max}), the maximum acceleration (a_f^{Max}) and *alpha*, i.e. the most sensitive parameters, were estimated.

5.4.2 Model parameters' estimation

Figures 5.7 and 5.8 present the comparison of estimated empirical distributions of model parameters between *full* and *reduced* models. Plots relate to the distribution only of those parameters that were calibrated for both the *full* and *reduced* model, i. e. *alpha*, *T*, V_f^{Max} and a_f^{Max}. Figure 5.7 relates to the estimation on speed, while Figure 5.8 on spacing.
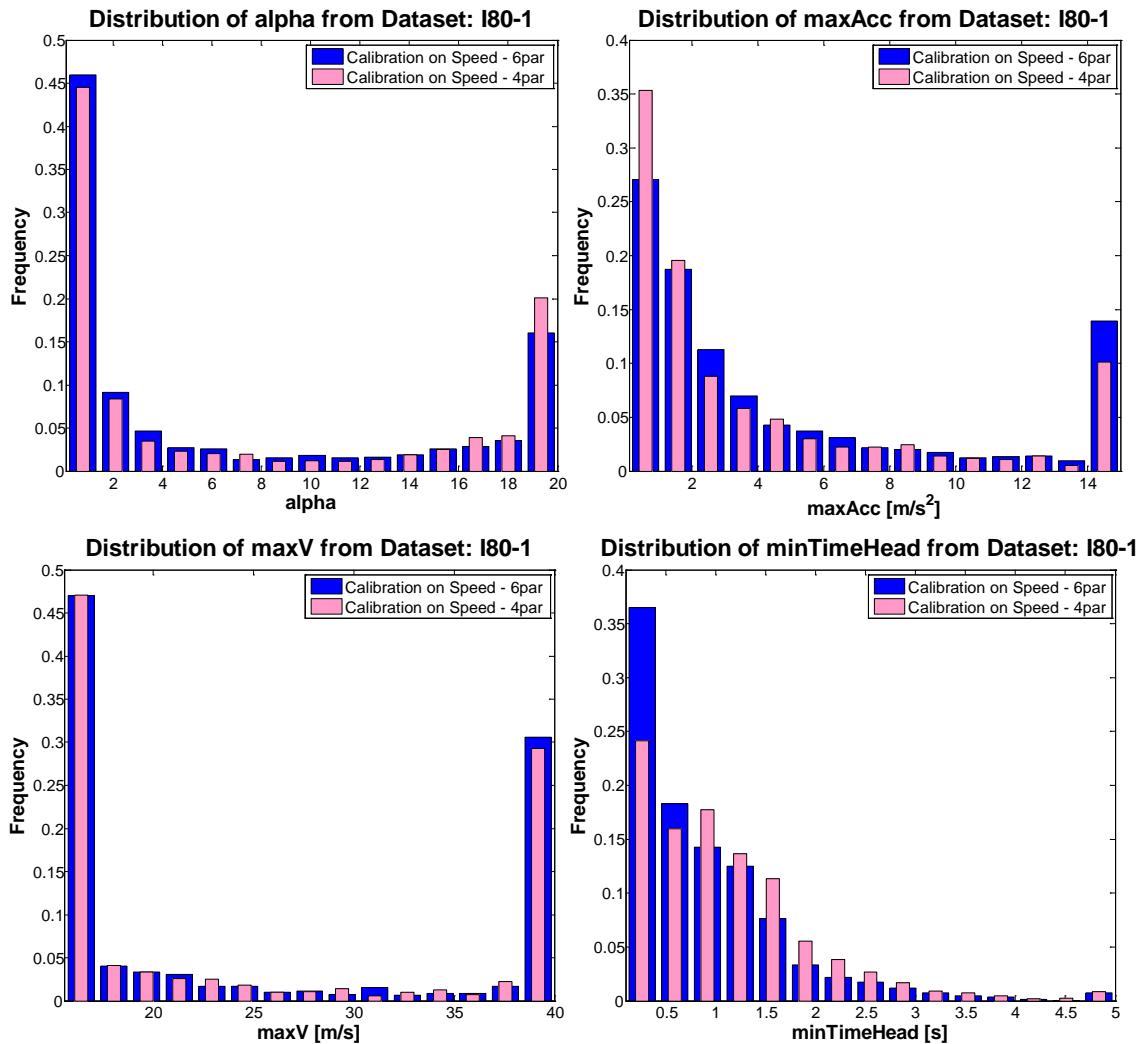
Figure 5.7: Empirical distributions of estimated model parameters for the *full* (tagged as "6par") and *reduced* (tagged as "4par") models. Model estimation was performed using the *speed* as measure of performance.
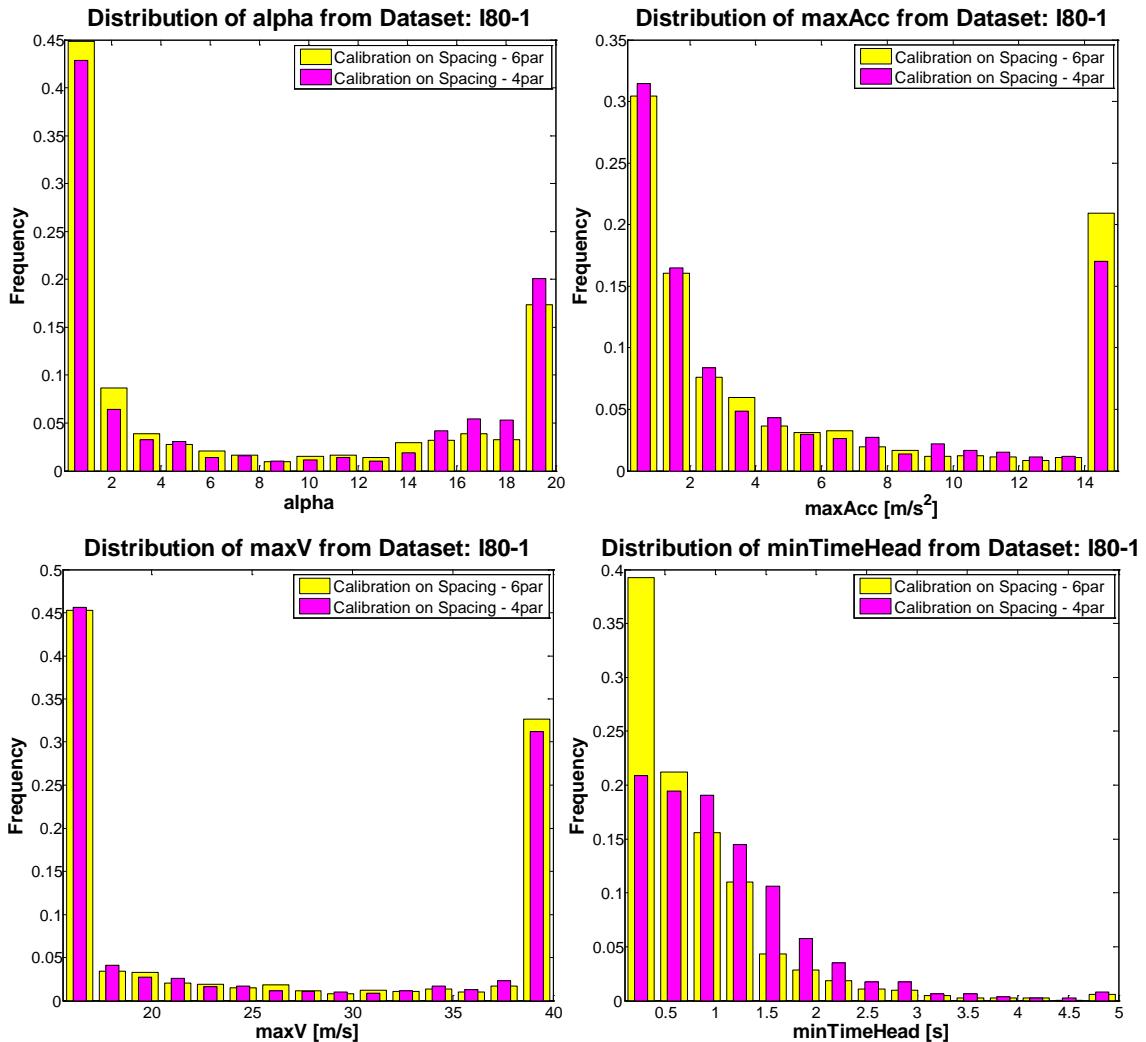
Figure 5.8: Empirical distributions of estimated model parameters for the *full* (tagged as "6par") and *reduced* (tagged as "4par") models. Model estimation was performed using the *inter-vehicle spacing* as measure of performance.

From the figures, it can be appreciated that differences in estimated model parameter distributions are relevant only for the *minimum time headway* of the IDM model, which turned out to be the most influential model parameter in the sensitivity analysis (see Figure 5.6).

In order to compare estimation performances between the two models, Figure 5.9 shows the distributions of the minimum error achieved in the *full model* (tagged as "6par") and *reduced model* (tagged as "4par") calibrations against the entire NGSIM I80-1 dataset. Figure 5.9(a) refers to the calibration on speed (the GoF function was the RMSE(v)), while Figure 5.9(b) to the calibration on spacing (the GoF function was the RMSE(s)).

Figure 5.9: Empirical distributions of the optimal value of the GOF function after the calibration of model parameters against (a) the RMSE(v) and (b) the RMSE(s), for all the vehicles in the NGSIM I80-1 dataset. In (a), the blue bars refer to the *full* model calibration (i.e. all 6 IDM parameters were estimated) while the pink ones refer to the *reduced* model calibration (i.e. where T, V_f^{Max}, a_f^{Max} and *alpha* are estimated). In (b), the yellow bars refer to the *full* model calibration while the magenta ones to the *reduced* model calibration.

In the case of calibration on speed (see Figure 5.9(a)), fixing non-influential model parameters produced very little effects on the capability of the model to reproduce the follower trajectory, as compared to the full model. Indeed, the mean value of the RMSE(v) shifted from 0.68 m/s, in the full model, to 0.75 m/s, in the reduced one, with an average increase of 10%. Also the increase in the standard deviations was minimal (from 0.20 m/s to 0.23 m/s).

Though distributions are statistically different at the level of confidence of 5%, we can

assert that, from a practical point of view, the *reduced* model performed mostly the same as the *full* model when estimated on speed.

In the case of calibration on spacing (see Figure 5.9(b)), results are somewhat different. The average increase in RMSE(s) was about 30%. This is due to the fact that the sum of the total order effects on spacing of the two last parameters is higher than that on speed, and that this increase (apparently negligible) has non-linear effects on the calibration results. Unlike usual sensitivity analysis, indeed, in this case we are not looking at the variance of the (*reduced*) model results but, at the variance of the calibration results (of the *reduced* model), which is a strict subset of the first one. That would explain such non linear effect. However, it is worth noting that, in absolute value, such effect corresponds to an average additional error of *50 cm*, which is not relevant in practical applications.

In addition, in both the cases, computational improvement is evident as the number of iterations in calibration halves for the reduced model to *10,000*, on average.

Indeed, the same analyses was repeated on a further *reduced* model where only the most two influential parameters, according to Table 5.1, were estimated, i.e. the minimum time headway (T) and the maximum desired speed (V_f^{Max}). Estimated model parameter distributions are reported in Figure 5.10, while estimation errors are presented in Figure 5.11. In the figures, (a) refers to the estimation results on speed, while (b) on spacing.

Also in this case, the empirical distribution of the estimated minimum time headway presents differences between the case of *full* and *reduced* model calibrations, confirming the importance ranking highlighted by the sensitivity analysis (see Figure 5.10).

On the other hand, when comparing fitting performances (Figure 5.11), ignoring the contributions of a_f^{Max} and *alpha* on the estimation error variance produced an average variation of the RMSE(v) of about 32%, though with a tremendous benefit in terms of computational effort with an average decrease of about 90% in the number of model evaluations needed. On the other hand, when calibrating on spacing, though the improvement in computational performances is evident (-88%), the average minimum estimation error was more than doubled. This confirmed the fact that neglecting even a small part of the uncertainty in the model inputs (i.e. by fixing a_f^{Max} and *alpha*) may have an impact on calibration results.
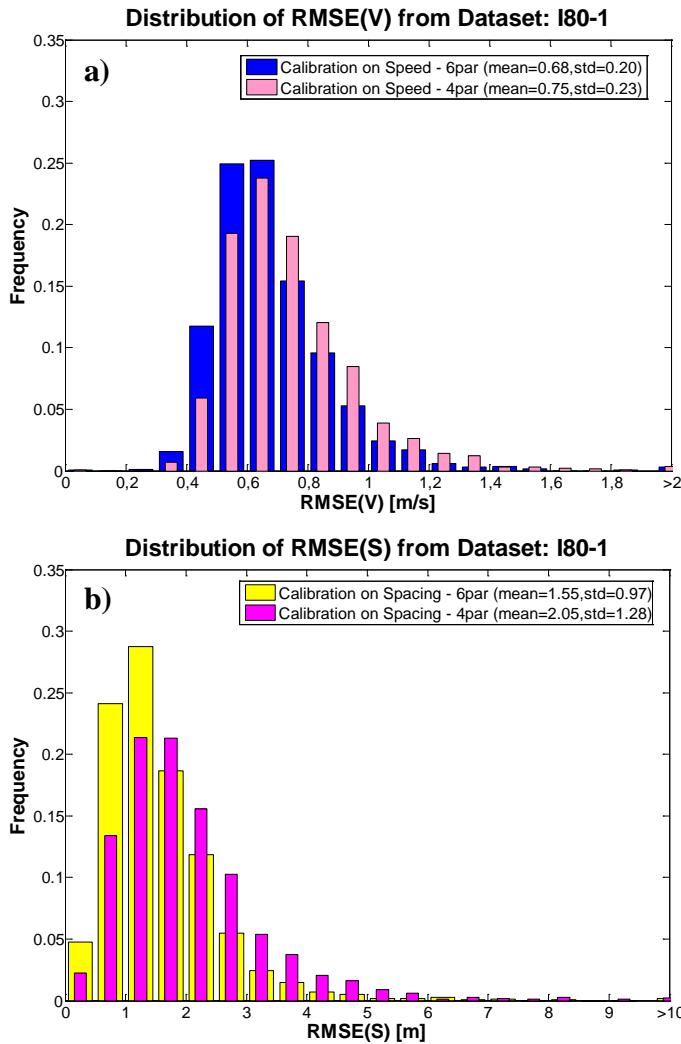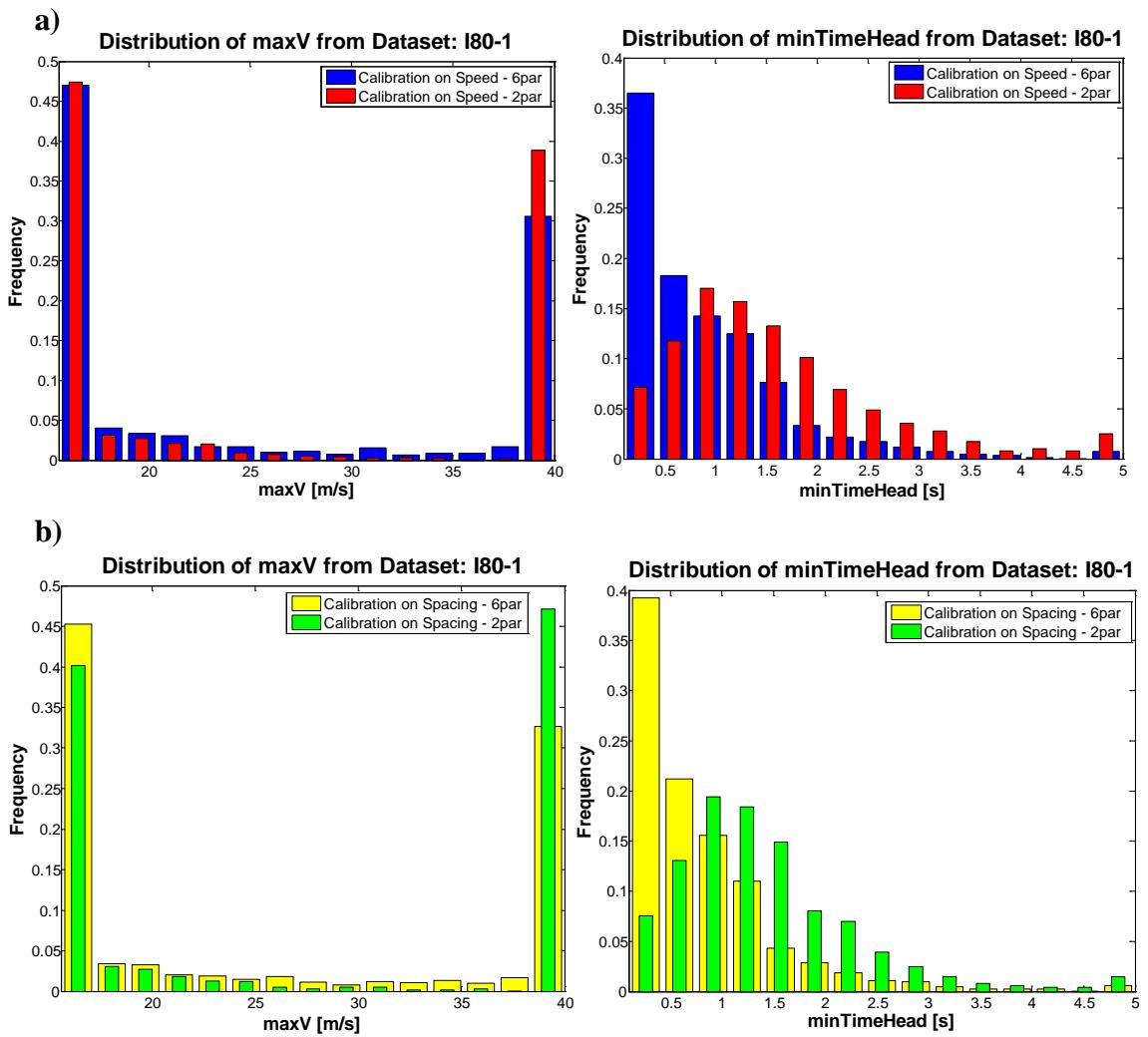
Figure 5.10: Empirical distributions of estimated model parameters for the *full* (tagged as "6par") and *reduced* (tagged as "2par") models. (a) refers to model estimation on speed, while (b) on spacing.
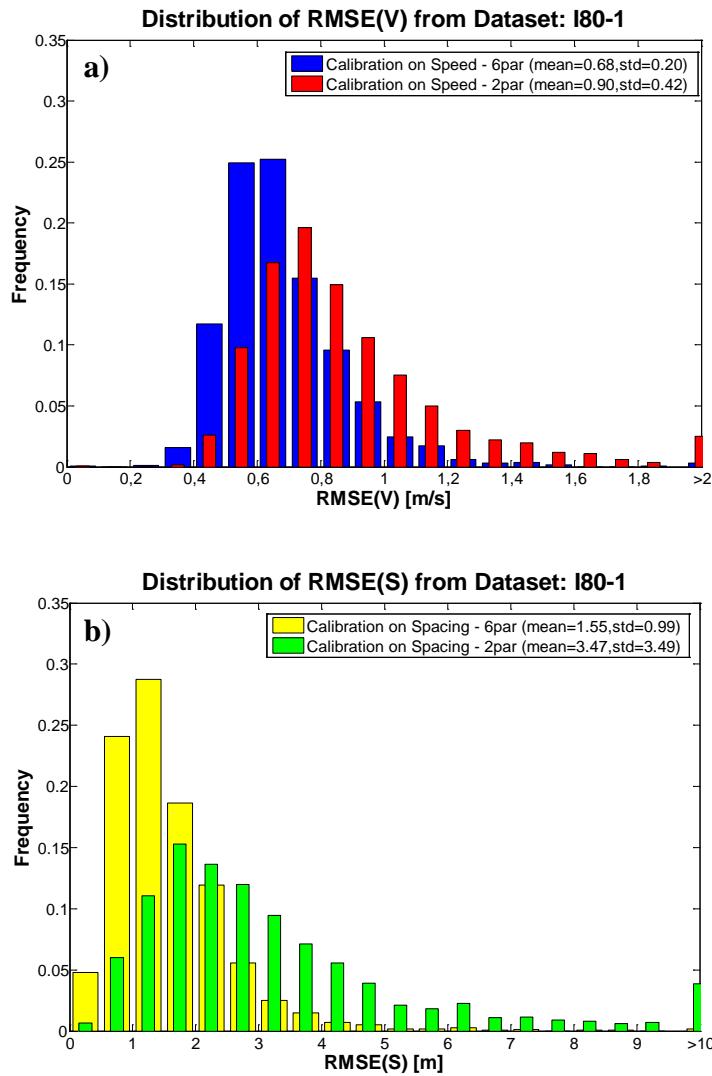
Figure 5.11: Empirical distributions of the optimal value of the GOF function after the calibration of model parameters against (a) the RMSE(v) and (b) the RMSE(s), for all the vehicles in the NGSIM I80-1 dataset. In (a), the blue bars refer to the *full* model calibration (i.e. all 6 IDM parameters were estimated) while the red ones refer to the *reduced* model calibration (i.e. only T and V_f^{Max} were estimated). In (b), the yellow bars refer to the *full* model calibration while the green ones to the *reduced* model calibration.

5.4.3 Calibration on speed vs. spacing

The experimental design applied in this work allowed us to empirically study the impact of the adopted Measure of Performance (MoP) on the parameter estimation results.

Indeed, as detailed in Chapter 3, most of the studies in the field research adopted, as MoP, either the speed profile of the measured follower vehicle, or the time-series of the inter-vehicle spacing between the measured leader and follower trajectories.

However, at the best of our knowledge, there is a gap in the research literature on this topic, as the choice of the MoP is not supported by any scientific proof or empirical evidence on which is the one that should be adopted for a robust specification of the model estimation problem.

On this basis, in this section we attempted to investigate this issue by providing empirical evidence on the differences of the two measures in the estimation performances.

Recalling the experimental design for the estimation of the *full* model (though similar considerations hold also on the *reduced* model calibration), we estimated the IDM model parameters of each individual vehicle from the *reconstructed* NGSIM I80-1 dataset adopting, in one case, the RMSE(v), and, in the other, the RMSE(s).

Therefore, we could evaluate cross-performances on spacing and speed, respectively, at the optimal values of model parameters, i.e. we evaluated the error on spacing using the model parameters estimated on speed, and vice versa.

Results are presented in Figure 5.12. In Figure 5.12(a), each dot corresponds to the calibration of model parameters of an individual vehicle from the *reconstructed* NGSIM I80-1 dataset. The coordinates of each blue dot are the estimation error on speed (*x*-axis) and the cross-simulation error on spacing with estimated model parameters (on speed; *y*-axis). Conversely, the coordinates of each yellow dot are the estimation error on spacing (*y*-axis) and the cross-simulation error on speed with estimated model parameters (on spacing; *x*-axis). Figure 5.12(a) allowed for qualitative investigation on the variance of the *estimation* and *cross-simulation* errors. However, errors on speed and on spacing could not be compared quantitatively. Therefore, in Figure 5.12(b), we normalized the error values on the *x*- and *y*-axis with respect to the maximum *estimation* error achieved when calibrating on speed and on spacing, respectively.

a)

b)
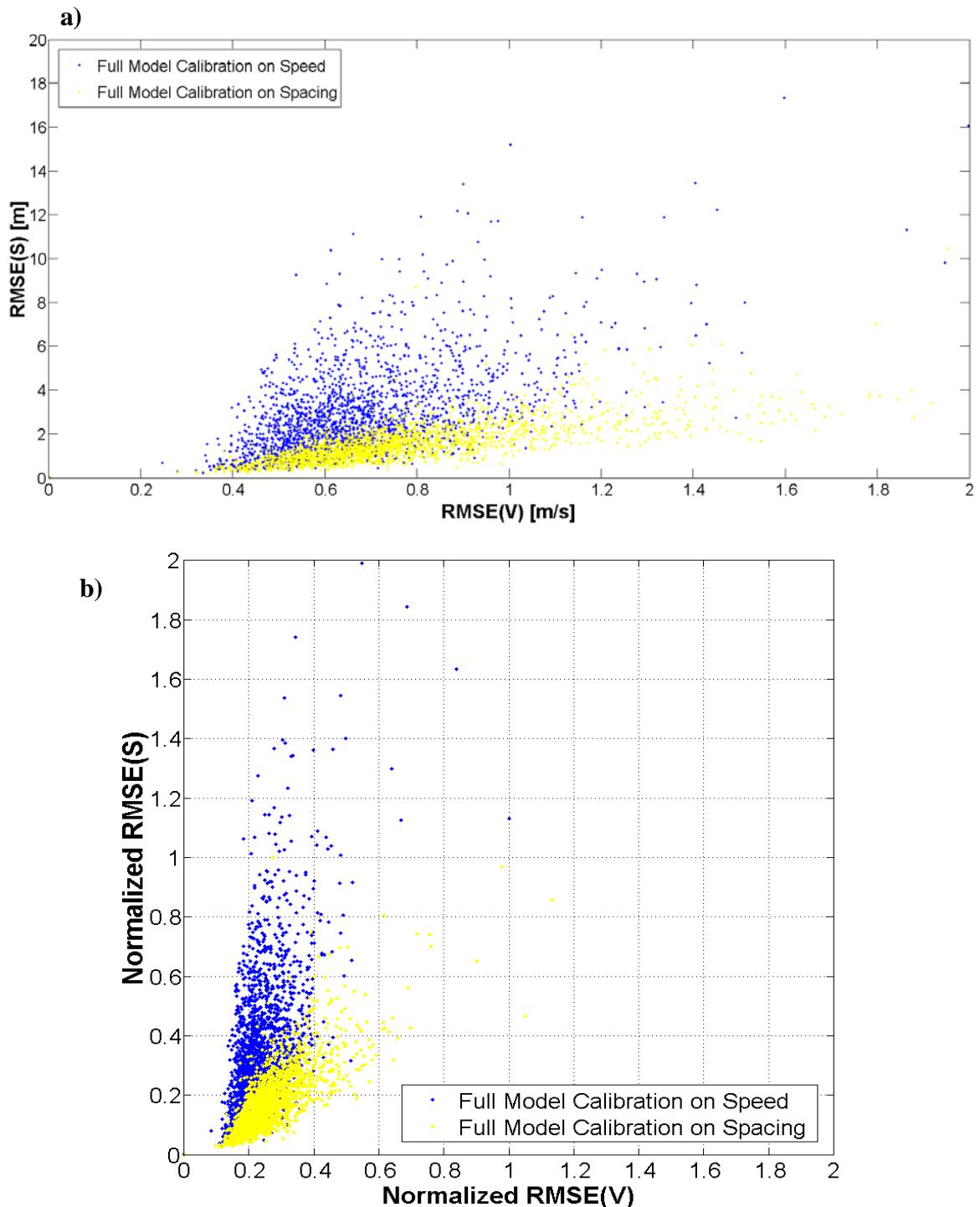
Figure 5.12: Comparison of model calibration performances on speed and on spacing. Each dot corresponds to the calibration of model parameters of an individual vehicle from the NGSIM I80-1 reconstructed dataset. In (a), the coordinates of each blue dot are the *estimation* error on speed (x-axis) and the *simulation* error on spacing with estimated model parameters (y-axis). Conversely, the coordinates of each yellow dot are the *estimation* error on spacing (y-axis) and the *simulation* error on speed with estimated model parameters (x-axis). In (b), values are normalized over the maximum *estimation* error on speed (x-axis) and on spacing (y-axis),

It is worth noting that, in the normalized plot (Figure 5.12(b)), the estimation errors on speed (RMSE(v) of the blue points) are bounded in the range of variation [0, 1]; the same holds for the estimation errors on spacing (RMSE(s) of the yellow points). On the other hand, instead, the *cross-simulation* errors (i.e. simulation error on spacing with parameters estimated on speed, and vice versa) can be greater than the related maximum estimation error (e.g. the *normalized* RMSE(s) of the blue points can be greater than 1).

From the figure, we can appreciate the following:

- the *estimation* error variance (the width of the scatter of *normalized* points over the measure of performance used in the estimation) in the case of calibration on speed (variation of *normalized* RMSE(v) of blue points equal to 0.07) is comparable with that resulting from calibration on spacing (variation of *normalized* RMSE(s) of yellow points equal to 0.10);

- the *cross-simulation* error variance (the width of the scatter of *normalized* points over the measure of performance used in simulation for cross-comparison) is much lower in the case of calibration on spacing (variation of *normalized* RMSE(s) of blue points equal to 0.09, with 95th percentiles equal to 0.43) than the one resulting from calibration on speed (variation of *normalized* RMSE(s) of yellow points equal to 0.25, with 95th percentiles equal to 0.77).

Apart from the (estimation and cross-simulation) error variances, these plots do not provide a quantitative performance evaluation of the two estimation settings. In other words, we need a common base to evaluate the two settings. Therefore, recalling the methodology adopted in Chapter 6 to compare different optimization settings, we used the sum of the Theil's Inequality Coefficients on speed and on spacing as a *validation score* of the model fitting.

Results are presented in Figure 5.13, where the distribution of the validation score from model simulation with parameters estimated against *speed* (blue bars) is compared with the distribution of the validation score from model simulation with parameters estimated against spacing (yellow bars). The figure confirmed previous findings, as calibration on spacing allowed to achieve "better" validation scores (i.e. lower values of the validation measure) than model estimations on speed (40% average reduction of the validation measure from 0.14, on speed, to 0.10, on spacing).
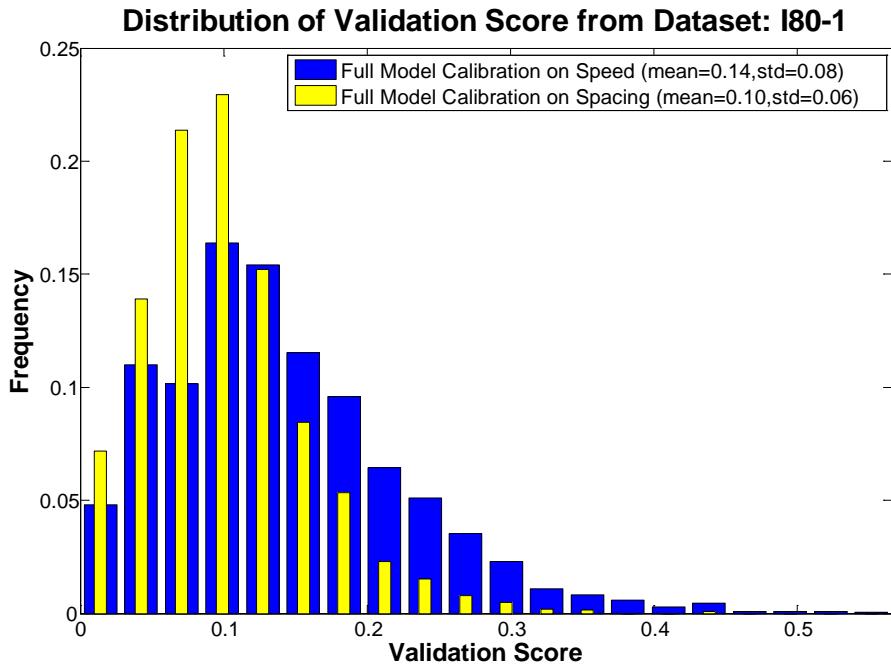
Figure 5.13: Empirical distributions of the optimal value of the Validation Score (i.e. sum of Theil's Inequality Coefficients on speed and spacing) after the calibration of model parameters against (a) the RMSE(v) and (b) the RMSE(s), for all the vehicles in the NGSIM I80-1 dataset.

5.5 Summary

In microscopic traffic simulation, the characterization of the uncertainty in the parametric inputs is referred as "calibration" and basically consists in identifying the parameters values that make the model output as close as possible to the reality. Calibration is therefore the acknowledged way to cope with the approximation of traffic simulation models, and is expected to cover both the uncertainty in the modeling assumptions/formulations and the uncertainty in the inputs.

However, automated calibration of microscopic traffic flow model parameters is arduous for a number of reasons. These include the difficulty to define a proper setting for the calibration problem, the computational complexity of any "black-box" optimization – that is exponential in the number of parameters – and the asymmetry of the parametric inputs in influencing the outputs.

A possible remedy to the problem complexity is to reduce the number of parameters to calibrate. As fixing a model parameter at a constant value means reducing fictitiously the output variance, this operation is generally arbitrary. However, a model sensitivity

analysis in a so-called "factor fixing" setting, can help on this. It allows identifying the parametric inputs that can be fixed at any value without affecting the output variance: a necessary and sufficient condition for this is that the parameter has a total sensitivity index equal to zero. In practical applications, however, a threshold value higher than zero is adopted: it is upon the analyst to verify whether the threshold chosen is consistent with the approximation requested/expected.

In this study, a global sensitivity analysis of the Intelligent Driver Model has been performed in a *factor fixing* setting. Main scope of the investigation was to verify which parameters have to be considered uncertain and therefore calibrated, in order to characterize correctly the uncertainty in the inputs.

To this aim an original framework has been designed for the analysis. As previous literature highlighted a significant dependency of car-following model performances on the input trajectory, and results of a sensitivity analysis are conditional on the inputs that remain fixed, the input trajectory was considered uncertain and sampled from the NGSIM I80-1dataset, as *reconstructed* in Chapter 4. The analysis was carried out considering both the errors on the speed and those on the spacing.

Results showed that the input trajectory is the most influential factor both in terms of first-order effect and in interactions with the model parameters. The variance of the model error conditional on the input trajectory – a function of the parameter combination sampled – has been therefore suggested as a measure of the "risk" of choosing a non-optimal model parameter combination: the higher the variance, the higher the risk of incurring in big modeling errors. It has also been shown, graphically, that such a variance is a decreasing function of the trajectory duration. This is an empirical evidence that car-following models should be calibrated on 'long' trajectories.

Basing on the results of the sensitivity analysis a reduced model has been tested. This model has four uncertain parameters instead of six, as the "desired distance at stop" and the "deceleration parameter" has been fixed at default values being non-influential. In fact, the variance explained by the ensemble of the two parameters is about 2%. Once calibrated on speed, the *reduced* model has almost equivalent performances to the *full* model. An increase in the error of the 30% is recorded instead when the model is calibrated against spacing. This result was not expected given the very low output variance explained by the two parameters in the *full* model (2%), but could be the effect

of the integral nature of the measure "spacing". If focusing on absolute values of the error, however, the 30% correspond to an average increase of 47 centimeters that is negligible in practical applications.

A reduced model with only two parameters has been also tested. Again, the error increase is limited for calibrations made on speed, while it is significant for those made on spacing. This is in line with the previous literature (Punzo and Simonelli, 2005) and suggests again that the calibration made on spacing is more challenging than that on speed.

In conclusion, the methodology allowed us to simplify the Intelligent Driver Model without sensibly affecting model performances. This has a dramatic effect on model calibration as, reducing the number of model parameters by one/third, the calibration time approximately halved.

Ultimately, the analysis allowed us to quantify the dependency of the model performances – as measured by the output variance – on the input trajectory and, in particular, on its duration.

Chapter 6

From Driver Behavioral Models to "Aggregate" Micro-Simulation

6.1   Introduction

In the previous Chapters, we applied the four-steps uncertainty management framework to car-following and lane-changing models, separately, in order to understand the impact of the different sources of *parametric* and *non-parametric* uncertainty on model performances.

However, in these studies, we have deliberately ignored the fact that, when using microscopic traffic flow simulation software, driver behavioral models (such as car-following and lane-changing models) are only "components" which continuously interact with each other, and with other "components" (e.g. route choice model).

Indeed, as pointed out in Chapter 2, focusing on "disaggregate" models would generally allow the analyst to reduce the (mostly computational) complexity of the uncertainty management framework, as well as to study more accurately some model properties.

However, one could question which is the impact of the findings related to "disaggregate" models (e.g. model calibration, model simplifications) on the performances of the micro- simulation software.

For the sake of simplicity, in the following, we would refer to the micro-simulation software as the "aggregate" simulation model, while to the driver behavioral models (and more in general to all the "components") as the "disaggregate" models or sub-models. It is worth noting that the attributes "aggregate" and "disaggregate" do not imply any relationship with the level of detail of the simulation (which is, indeed, microscopic) or with the level of aggregation of traffic measurements.

Therefore, in this Chapter we undertook the following studies to evaluate different impacts on simulation performances of the "aggregate" model. In particular:

- Analysis of the impact of "disaggregate" model calibration in presence of *measurement errors*;

- Analysis of the impact of "disaggregate" model simplifications;

- Analysis of the impact of "disaggregate" model parameters correlation;

For the scope of the analysis, two elements of the design of the experiment are crucial to allow for robust and reliable evaluation.

First of all, a complete set of all individual vehicle trajectories from a traffic stream are needed as a common basis to *i)* perform the "disaggregate" model parameters calibration, and *ii)* evaluate the simulation performances of the "aggregate" model, by space-time aggregation of trajectory data.

Secondly, as to be fair in comparison with measured data, we needed to perform an "aggregate" micro-simulation where all the external inputs (e.g. vehicle generation) as well the initial state of the simulation, were taken directly from the measured data. Indeed, if we ignored this, when looking at simulation performances of the "aggregate" model, we would not be able to distinguish the portion of uncertainty apportioned to the model itself from that depending on the external inputs used to feed the model with.

The Chapter is organized as follows. The methodological framework applied herein to compare measured and simulated performances of the "aggregate" model is depicted in Section 6.2. Section 6.3 is dedicated to the Design of Experiment, with regards to both the description of the case-study, and the design of the "aggregate" microscopic traffic flow simulation model. Then, Sections 6.4 – 6.6 detail the analysis of the impact listed above. Brief summary of the main findings ends the Chapter.

6.2 Methodological Framework

The objective of this work is the evaluation of the performances of the "aggregate" micro-simulation software in reproducing measured quantities, under a variety of modeling assumptions with regards to its sub-models (e.g. different estimated parameter distributions and correlation structures, model simplifications).

To this aim, in order to design a fair comparison between measured and simulated performances, we needed to eliminate the portion of uncertainty in model inputs which is not in the objective of the analysis. For example, the input OD flows or the uncertainty in the vehicle generation model (e.g. probabilistic model and parameters) may have a great impact on the results of the "aggregate" micro-simulation. However, in the present study we would like to isolate the impact of only the uncertainty in driver behavioral models.

To reach this goal, a *trace-driven simulation* was needed to evaluate "aggregate" model performances. Indeed, in a trace-driven simulation, the observed values (also called "trace" in computer science applications) of all the fixed inputs (as defined in Chapter 1) is used to run the simulation. Therefore, the initial state of the simulation was loaded according to the actual measured state (lane, position, speed, and acceleration) of the vehicles on the freeway stretch under analysis. Further, during simulation, vehicles attempted to enter the stretch at the same time instant, and in the same state, as in measured data.

Consequently, the complete set of all individual vehicle trajectories in a time-space domain are needed to accomplish with the requirements of a trace-driven simulation.

6.3 Design Of Experiment

This section is devoted to *i)* the description of the freeway scenario used in this work to conduct the simulation study and *ii)* the description of the "aggregate" microscopic traffic flow simulation model applied herein.

6.3.1 NGSIM I80 case study

The scenario adopted in this study is the northbound stretch of the I80 freeway in Emeryville, California (NGSIM, 2005). The study area is represented in Figure 6.1 (NGSIM, 2005).
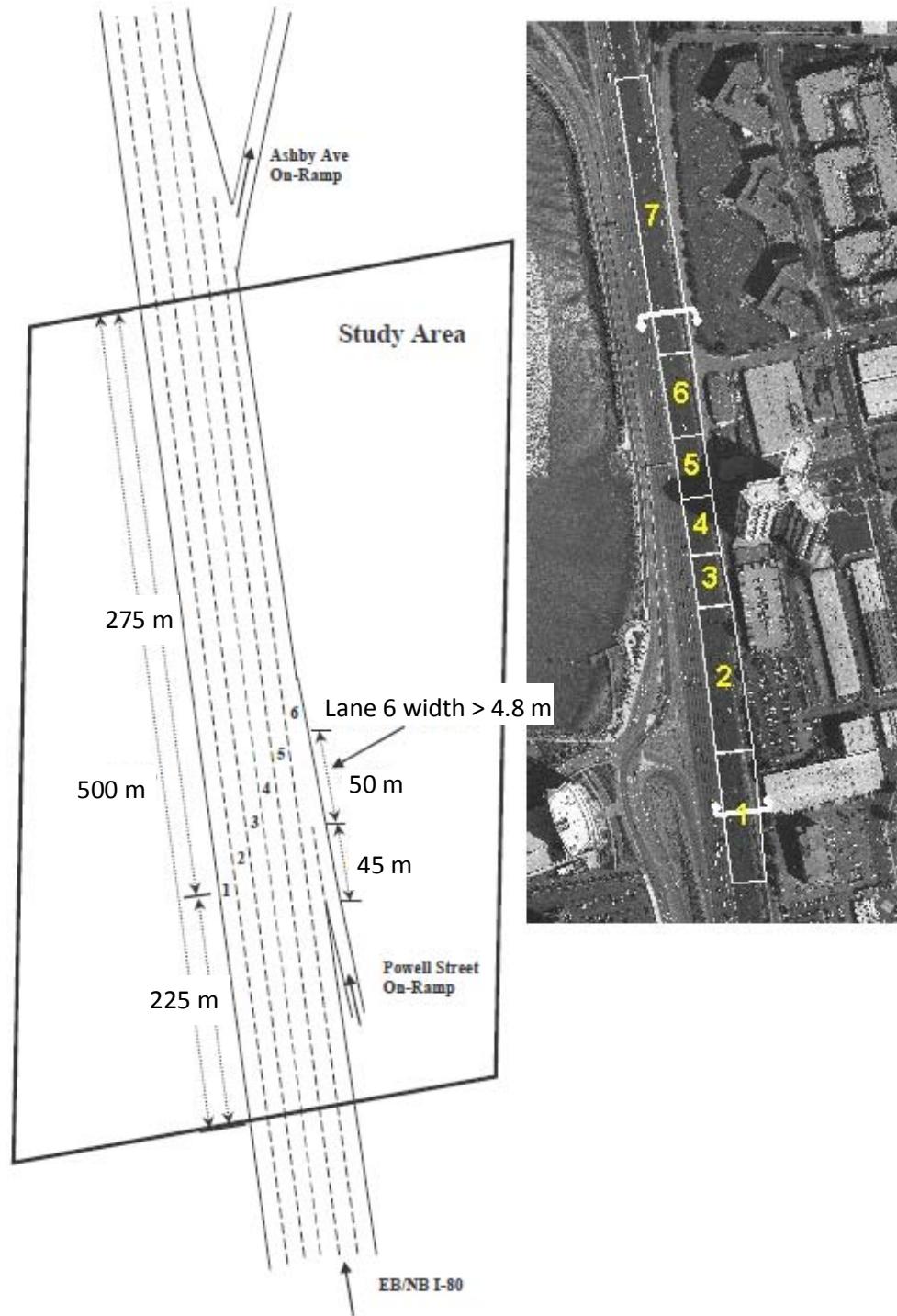
Figure 6.1: NGSIM I80 study area with camera coverage for individual vehicle trajectory tracking.

The section is approximately 500 m. length and has an HOV lane (lane 1) and 5 regular lanes (from lane 2 to 6). An on-ramp merges in the section approximately at approximately 225 m. from the upstream bound. The on-ramp lane (lane 7) is not physically separated from the most right section lane, while the road markings identify the lane limits for about 45 m after the initial merge. Downstream from the on-ramp, lane 6 has a greater width for about 50 m. This geometric precaution is due to safety reason, in order to allow slower vehicles coming from the on-ramp to "forced" merge in lane 6 by the occupation of that lane, side by side with vehicles from the upstream. Ultimately, an off-ramp is located downstream of the study area.

For the study purpose, we used the complete set of all individual vehicle trajectory data recorded from 4:00 p.m. to 4:15 p.m. on April 13, 2005 – in the following referred as *I80-1* – were used to *i)* estimate driver behavioral model parameters and *ii)* evaluate "aggregate" model performances. Provided the large amount of measurement errors in the original data, data used herein were *reconstructed* as detailed in Chapter 4. Further, as pointed out in Chapter 4, data acquired by camera 1 and camera 7, were substantially corrupted by a large amount of measurement errors (and indeed not *reconstructed*). Therefore, in the present simulation study, traffic flow in these areas was not simulated. Figure 6.2 depicts the NGSIM I80 simulation area and its peculiarities.

The complete picture of the traffic dynamics in the monitored period is depicted in Figure 6.3, by means of the space-time evolution of the Edie's mean speeds (Edie, 1974). The figure clearly shows the upstream propagation of three waves, accompanied by intense congestion.

In order to reproduce downstream traffic dynamics, we designed a "*buffer zone*" in the last 50 meters of the simulation area, where vehicles moved accordingly to the measured positions (instead of being actually simulated). Indeed, for each vehicle, a transition to the measured speed levels was set up based on vehicles' maximum acceleration and deceleration parameter values of the car-following model, in order to preserve consistency on accelerations and decelerations.

Finally, in order to compare measured and simulated performances, three virtual detectors were located in the simulation area, located, respectively, 50 meters downstream of the simulation entrance, 25 meters downstream of the merging zone, and at the end of the simulation area, i.e. 50 meters before the buffer zone (where boundary conditions are applied).

NGSIM I80 – Study area
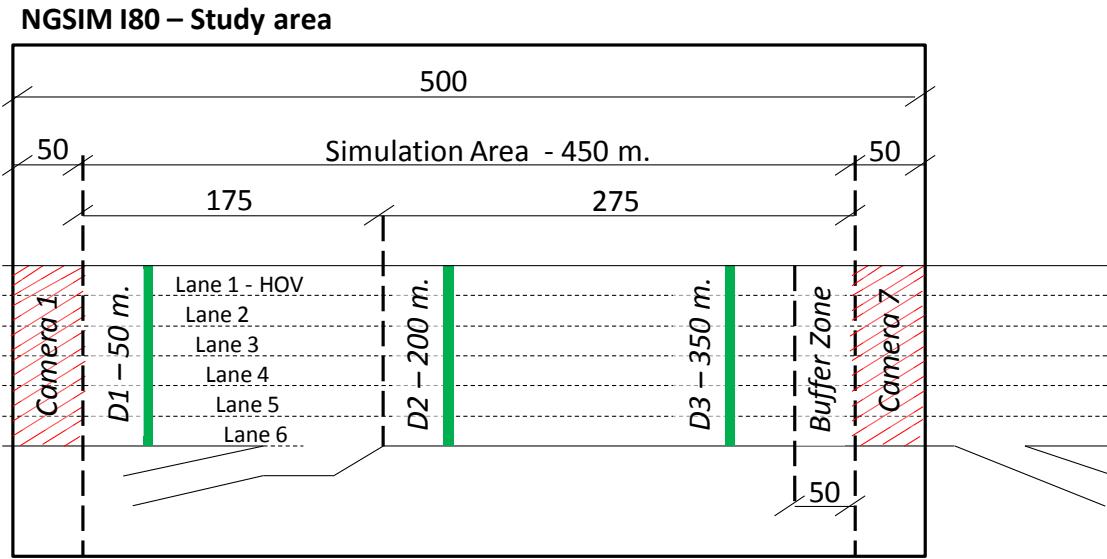
Figure 6.2: NGSIM I80 study area and the simulation area. The graphical representation is not in scale. Distances are in meters. In the study area, lane 1 is reserved to high occupancy vehicles. As trajectory data extracted from camera 1 and 7 were not considered in the vehicle trajectory reconstruction procedure (see Chapter 4 for details), traffic flow in these areas was not simulated. As a consequence, the simulation area was reduced to 450 meters. In the last 50 meters of the simulation area, boundary conditions were applied to preserve the propagation of traffic conditions from downstream (see Section 6.2). Three virtual detectors were located on the main road: D1 located 50 meters after vehicles entered the simulation, D2 located 25 meters after the on-ramp, and D3 located 50 meters before the buffer zone.
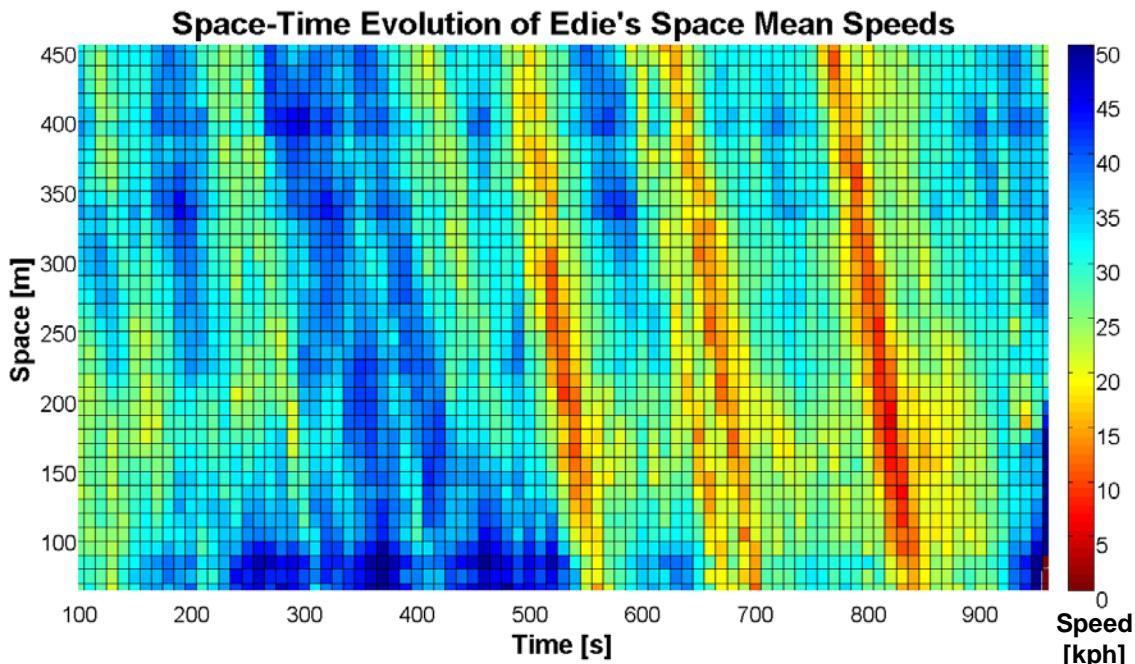
Figure 6.3: Space-Time evolution of space mean speeds in the NGSIM I80-1 dataset.

6.3.2 "Aggregate" Microscopic Traffic Flow Simulation Model

The peculiarities of the *trace-driven* simulation, the requirements of imposing downstream boundary conditions, and the need to fully customized the choice of the driver behavioral models used to simulate car-following, lane-changing and merging dynamics among vehicles, advised against the use of commercial microsimulation software for the study purpose.

Therefore, in order to accomplish the features listed above, an "aggregate" microscopic traffic flow simulation model was designed expressly for the study purpose, and coded in MATLAB (2009).

The selected car-following and lane-changing models were, respectively, the Intelligent Driver Model (IDM) by Treiber et al. (2000) and the MOBIL model by Kesting et al. (2007). The merging behaviour of vehicles from the on-ramp was simulated as a mandatory lane-changing with the MOBIL model, by (locally) setting the politeness factor and the acceleration threshold parameter values equal to zero (for a discussion, see Treiber and Kesting, 2013). For more details on models' formulation and the estimation framework adopted for parameter calibration against individual vehicle trajectory data, please refer to Chapter 3, for the IDM model, and to Appendix D, for the MOBIL model.

With regards to the upstream boundary condition, vehicles were generated according to the measured state (time instant, lane, position, speed, acceleration), and entered the simulation only if no physical constraint (e.g. presence of a vehicle in the target position in the target lane) was violated; otherwise, they were assigned to a virtual queue.

With regards to the simulation of the HOV lane (lane 1), it was necessary to introduce a lane-changing rule that emulated the actual existing traffic regulation, i.e. only vehicles with more than three occupants are allowed to travel in the HOV lane. Therefore, in simulation, we prevented vehicles to perform lane-changing maneuvers from lane 2 to the HOV lane.

6.3.3 Summary of experiments

According to the study objectives, the following topic were investigated:

- Analysis of the impact of "disaggregate" model calibration in presence of *measurement errors*, on "aggregate" simulation model performances (Section 6.4);

- Analysis of the impact of "disaggregate" model simplifications, on "aggregate" simulation model performances (Section 6.5);

- Analysis of the impact of "disaggregate" model parameters correlation, on "aggregate" simulation model performances (Section 6.6);

In the following sections, a detail description of each analysis is followed by the discussion of the main findings.

6.4 Impacts of Measurement Errors in Parameter Estimation

Based on the findings discussed in Chapter 4, the objective of this study is to understand the impact on "aggregate" simulation model performances of running a simulation with driver behavioral model parameters estimated against individual vehicle trajectory data in presence of measurement errors.

The methodology applied here is discussed in Section 6.4.1, while main findings are summarized in Section 6.4.2.

6.4.1 Methodology

The methodology applied in this analysis is described in Figure 6.4.

Based on the results presented in Chapter 4, we carried out two simulation experiments, by assigning each vehicle entering the simulation with its set of car-following and lane-changing model parameters, estimated against *raw* (the left-hand side of Figure 6.4) and *reconstructed* (the right-hand side of Figure 6.4) vehicle trajectory data.

It is worth noting that, the inter-vehicle spacing was adopted as measure of performance in car-following model calibration. For details, please refer to Chapter 5.

On the other hand, individual vehicle trajectory data were aggregated over time and space in order to obtain aggregate measures to compare simulation results with. It is worth noting that, as discussed in Chapter 4, the impact of measurement errors in vehicle trajectory data on aggregate measures (such as time mean speeds, space mean speeds, section density, travel times) is negligible. This allowed us to use the same measured output to compare simulation results from the experiments with both *raw* and *reconstructed* data.
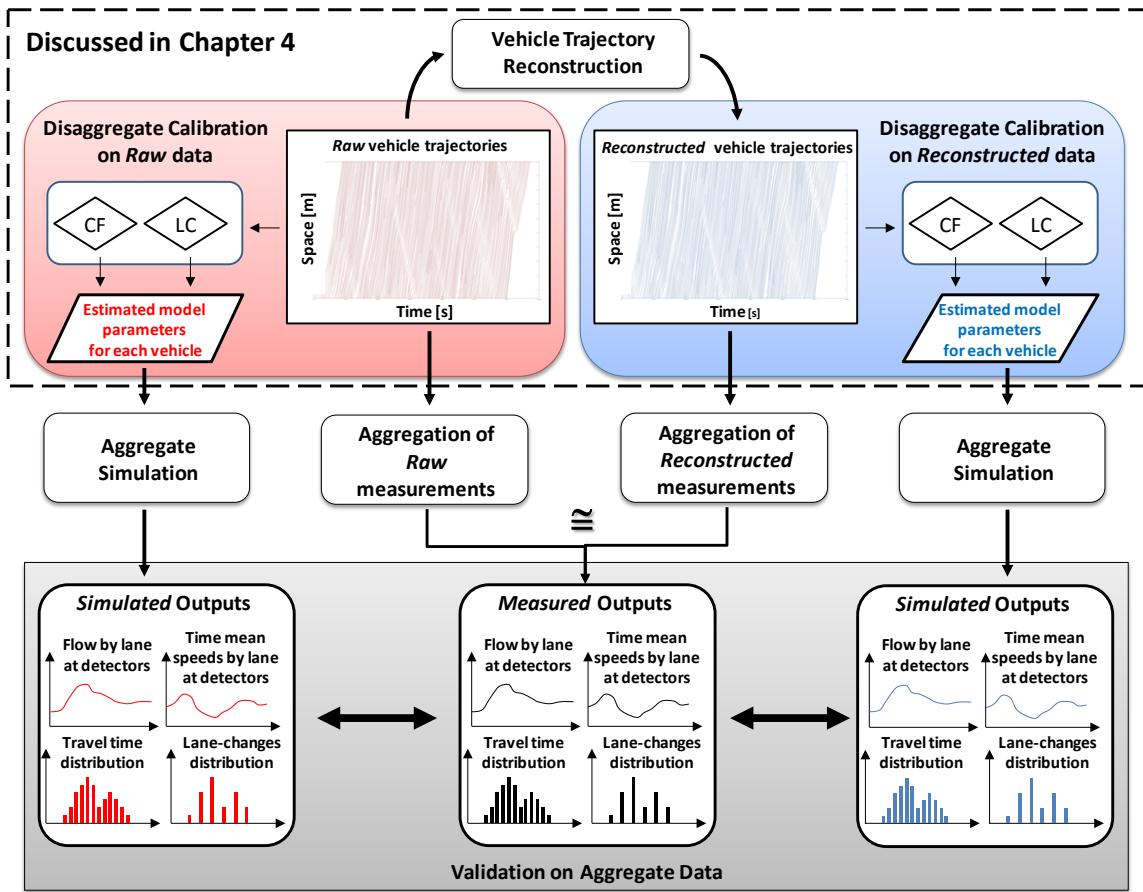
Figure 6.4: Conceptual framework to evaluate the impact of measurement errors in vehicle trajectory data on aggregate traffic flow model simulation. Car-following and lane-changing model parameters are disaggregate estimated against both *raw* and *reconstructed* trajectory data. Successively, traffic flow is simulated using the overall microscopic simulator, where each vehicle is assigned with its estimated model parameters. Finally, aggregate simulation results based on parameters estimated against both *raw* and *reconstructed* data, are compared with aggregated measurements over time and space.

6.4.2 Results

The comparison between measured and simulation results is conducted with respect to the following outputs:

- time-space speed contour plots (Figure 6.5);

- time-series of section density (Figure 6.6);

- distribution of travel times (Figure 6.7);

- distribution of the number of lane-changes (Figure 6.8).

Results are shown in Figures 6.5 – 6.8.

Figure 6.5 presents the space-time contour plots of the Edie's space mean speeds (Edie, 1974) related the measured data (a), and to the simulated one in case of simulation experiment with model parameter estimated against *raw* (b) and *reconstructed* (c) data.

Figure 6.6 shows the comparison between time evolution of the measured (black line) and simulated section density in case of simulation experiment with model parameter estimated against *raw* (red line) and *reconstructed* (blue line) data.

Finally, Figures 6.7 and 6.8 present the comparison of the empirical distributions of section travel times (Figure 6.7) and number of lane changes (Figure 6.8) between measured and simulated data.

From a visual inspection of the time-space speed contour plots (Figure 6.5), we may observe that both the two simulation experiments (Figure 6.5(b-c)) were not able to reproduce the complete back-propagation of the three shock-waves present in the measured data (Figure 6.5(a)).

Further, though model parameters distribution (and correlation structures) were practically identical between estimation against *raw* and *reconstructed* data, it appeared that, at the "aggregate" level, the simulation experiment with model parameters estimated against *raw* trajectory data (Figure 6.5(b)) overestimates traffic congestion within the section more than in the case with parameters estimated against *reconstructed* data (Figure 6.5(c)).

It is worth mentioning here that in both the *trace-driven* simulation experiments, intense congestion occurred at vehicle entrance, from a given instant in time. Indeed, as clarified in Section 6.3, though vehicle entered the simulation according to measured data, their simulated behavior depended on the actual interaction with the surrounding (simulated) vehicles, whose positions in the section might be substantially different from the measured states. Therefore, the back-propagation of the *simulated* shockwaves up to the section entrance, if not dissipated as in the real measurements, may produce a great impact on simulation results, preventing vehicles from entering (i.e. due to the boundary constraint at entrance; for details, see Section 6.3). For example, this can be clearly observed in Figure 6.5(a-b) around instant 500 s.

Based on this considerations, such observed phenomena in simulation – in the following referred as *entrance congestion* – is purely the result of performing a trace-driven

simulation. However, since the objective of this study is to assess the model capability to reproduce measured data, *including traffic states at section entrance*, the occurrence of the entrance congestion should be read as a failure of the model.

On the other hand, the increase of section congestion, can be seen also in Figure 6.6 in terms of section density, where, in the simulation experiment with parameters estimated against *raw* data, higher densities (about 300 veh/km) are reached during the back-propagation of the first shock-wave (in the time interval between 400 and 500s), compared to the simulation with parameters estimated against *reconstructed* data (which are closer to the measure data).

In terms of travel time distributions (Figure 6.7), we may see that both the models are able to reproduce average congested regimes, i.e. the peak in the measured data around 50 seconds. However, the higher speed levels in the HOV lane (which corresponds to the peak in the travel times around 20 seconds) are completely missed in the simulation experiments, which, conversely, as already pointed out, overestimate congestion (i.e. right tail of the distribution).

Finally, with respect to the distribution of the number of lane-changes, Figure 6.8 reveals that the simulation model largely overestimates the number of lane-changing maneuvers, although lane-changing model parameters were individually estimated against measured data.

a)

b)

c)
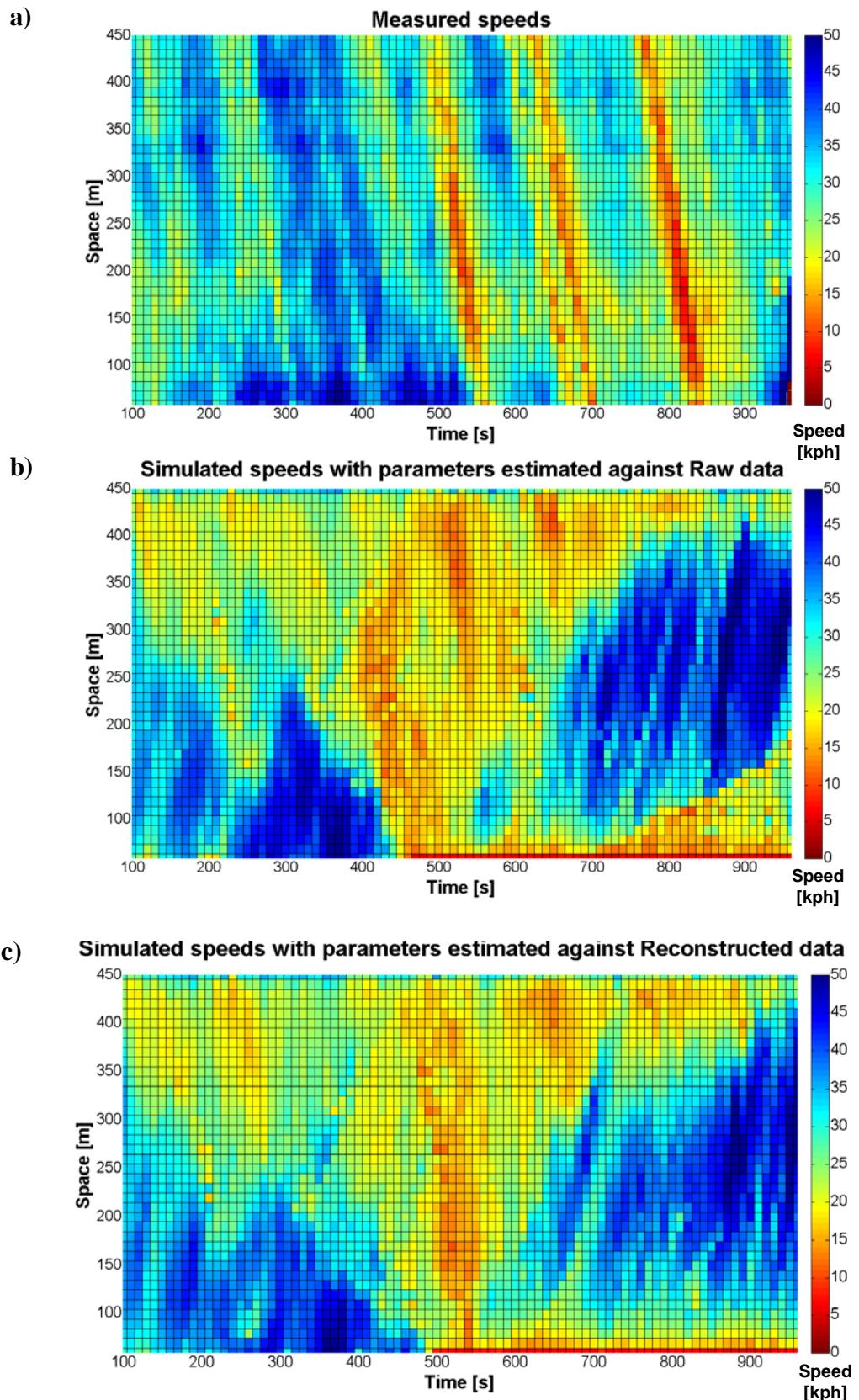
Figure 6.5: Space-Time contour plots of Edie's space mean speeds. (a) refer to measured data, while (b) and (c) to simulated data with parameters estimated against *raw* and *reconstructed* data, respectively.
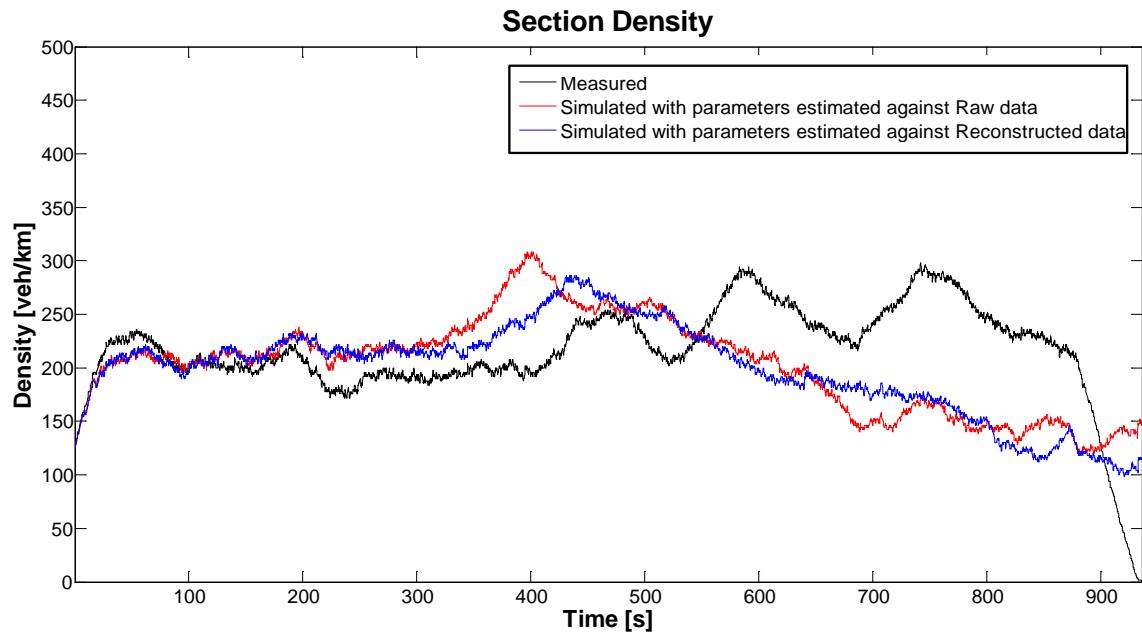
Figure 6.6: Time-series of the section density related to measured (black line) and simulated data, in case of simulation experiments with parameters estimated against *raw* (red line) and *reconstructed* (blue line) data.
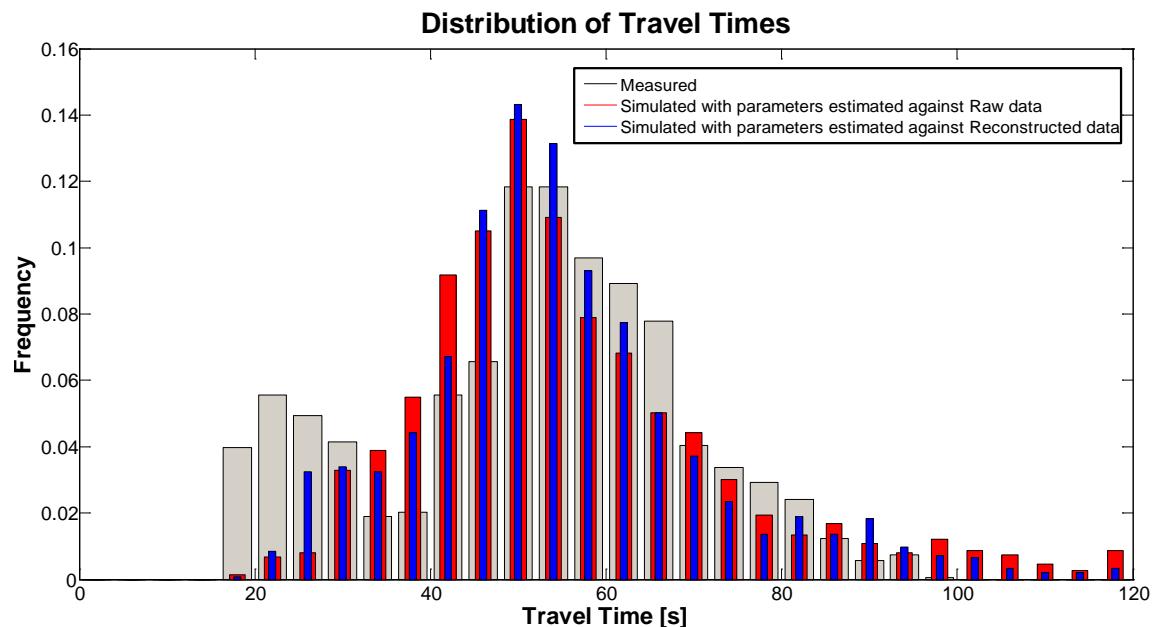
Figure 6.7: Empirical distribution of section travel times for measured (grey bars) and simulated data, in case of simulation experiments with parameters estimated against *raw* (red bars) and *reconstructed* (blue bars) data.
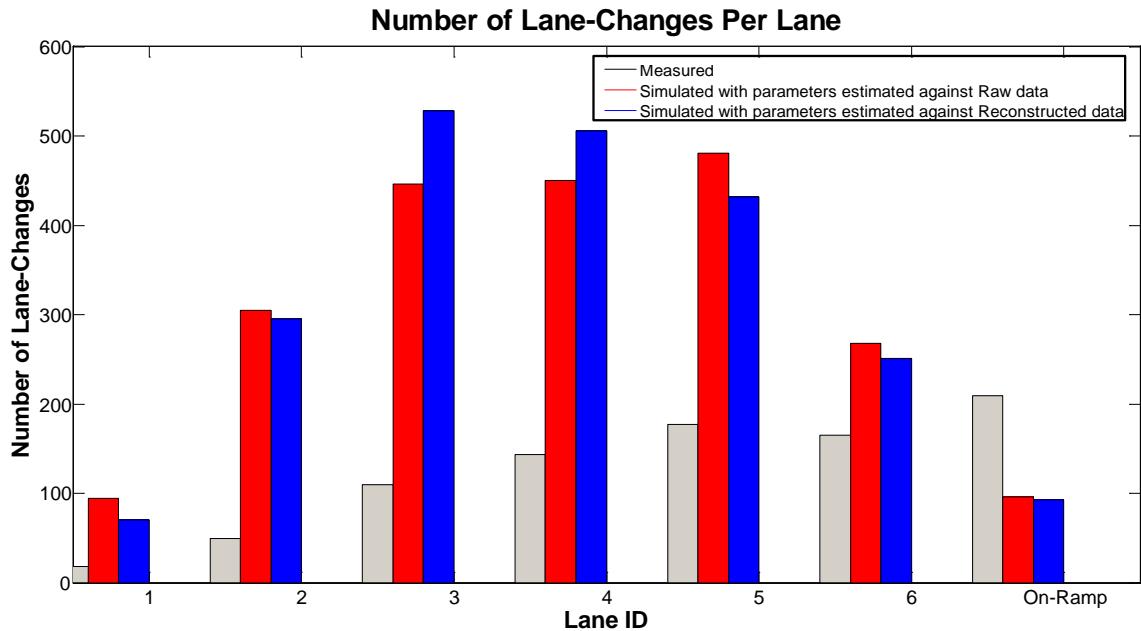
Figure 6.8: Empirical distribution of the number of lane-changes for measured (grey bars) and simulated data, in case of simulation experiments with parameters estimated against *raw* (red bars) and *reconstructed* (blue bars) data.

6.5   Impacts of Model Simplification Assumptions

According to the results of car-following model sensitivity analysis performed in Chapter 5, in this section we addressed the question regarding the impact of model simplifications on the ability to reproduce aggregate traffic flows.

On this basis, the methodology here applied is discussed in Section 6.5.1, while main findings are summarized in Section 6.5.2.

6.5.1 Methodology

The methodology applied in this analysis is described in Figure 6.9.

Similarly to the methodological approach depicted in Section 6.5, we based our investigation on the results of car-following model sensitivity analysis presented in Chapter 5. Therefore, we made use of the car-following model parameters estimated in case of the *full* model calibration experiment (i.e. all six IDM model parameters were calibrated) and of the *reduced* model calibration experiment (i.e. only two IDM model parameters were calibrated: the minimum time headway and the maximum desired speed). Lane-changing model parameter calibration was performed based on the *full* and *reduced* car-following model parameter estimations (the dependency of lane-changing model parameter estimation from car-following model estimation is detailed in Appendix D).

Successively, we carried out two simulation experiments, where each vehicle entering the simulation was assigned with its set of car-following and lane-changing model parameters, estimated in case of *full* (the left-hand side of Figure 6.9) and *reduced* (the right-hand side of Figure 6.9) car-following model calibrations.

It is worth noting that, as in the previous analysis, the inter-vehicle spacing was adopted as measure of performance in car-following model parameter calibration. For details, please refer to Chapter 5.

On the other hand, individual vehicle trajectory data were aggregated over time and space in order to obtain aggregate measures to compare simulation results with. This allowed us to evaluate the impact of model simplification assumption of the "aggregate" model performances, by comparing simulated outputs with measured ones.
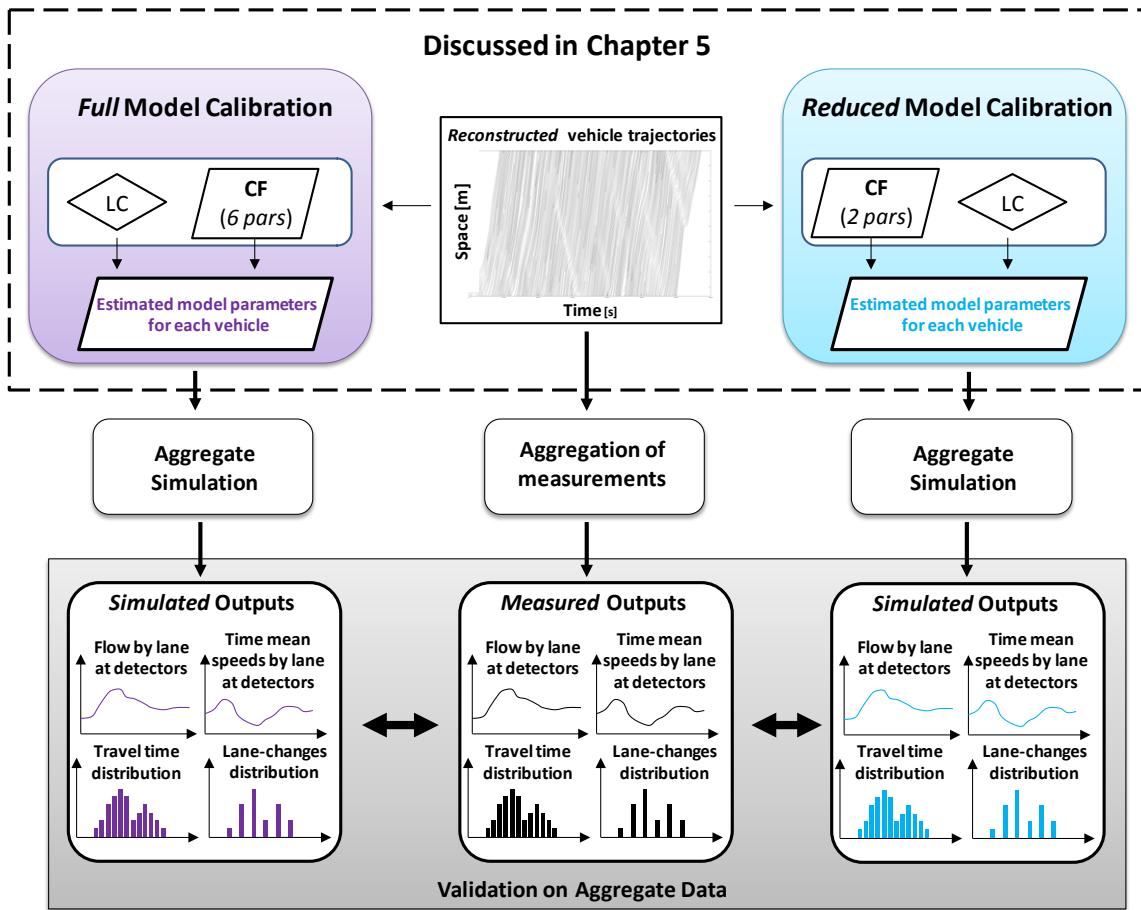
Figure 6.9: Conceptual framework to evaluate the impact of car-following model simplifications on aggregate traffic flow model simulation. *Full* (6 pars) and *reduced* (2 pars; minimum time headway and maximum desired speed) car-following and lane-changing model parameters are disaggregate estimated against *reconstructed* trajectory data. Successively, traffic flow is simulated using the overall microscopic simulator, where each vehicle is assigned with its estimated model parameters. Finally, aggregate simulation results with *full* and *reduced* model parameters estimations are compared with aggregated measurements over time and space.

6.5.2 Results

The comparison between measured and simulation results is conducted with respect to the following outputs:

- time-space speed contour plots (Figure 6.10);

- time-series of section density (Figure 6.11);

- distribution of travel times (Figure 6.12);

- distribution of the number of lane-changes (Figure 6.13).

Results are shown in Figures 6.10 – 6.13.

Figure 6.10 presents the space-time contour plots of the Edie's space mean speeds (Edie, 1974) related the measured data (a), and to the simulated one in case of simulation experiment with model parameter estimated in case of *full* (b) and *reduced* (c) car-following model.

Figure 6.11 shows the comparison between time evolution of the measured (black line) and simulated section density in case of simulation experiment with *full* (magenta line) and *reduced* (cyan line) car-following model parameter estimations.

Finally, Figures 6.12 and 6.13 present the comparison of the empirical distributions of section travel times (Figure 6.12) and number of lane changes (Figure 6.13) between measured and simulated data. It is worth noting that, as calibrations were performed against *reconstructed* data, measured and simulated (*full* model) outputs in Figure 6.10 – 6.13 coincides with those in Figures 6.5 – 6.8.

The comparison of the simulated speed contour plots with the measured one (Figure 6.10) shows that the simulation results of the experiment with estimated parameters from the *reduced* car-following model calibration reproduced more accurately the shockwaves present in the measured data, in terms of both propagation and wave speed. Further, also the overall level of congestion is lower than in case of the experiment with estimated parameters from the *full* car-following model calibration, and closer to the measurements, as can be appreciated from the time-series of the simulated section densities in Figure 6.11.

It is worth noting that in both the experiments emerged the so-called *entrance congestion*, highlighting a failure in the model to reproduce measured entrance conditions. For the interpretation of this phenomena, please refer to Section 6.4.2.

In terms of travel time distributions (Figure 6.12), the higher speed levels in the HOV lane (i.e. around 20 seconds in the measured travel time distribution; see the grey bars) is "better" captured by the *reduced* model simulation, while the *full* model simulation results are closer in reproducing the measured distribution in the range of travel times between 40 seconds and 60 seconds. On the other hand, as pointed out above, the *reduced* model does not show the high congestion levels as in case of the *full* model (see the magenta bars in the left tale of the travel time distribution).

a)

Measured speeds

b)

Simulated speeds with parameters estimated from full model calibration

c)

Simulated speeds with parameters estimated from reduced model calibration
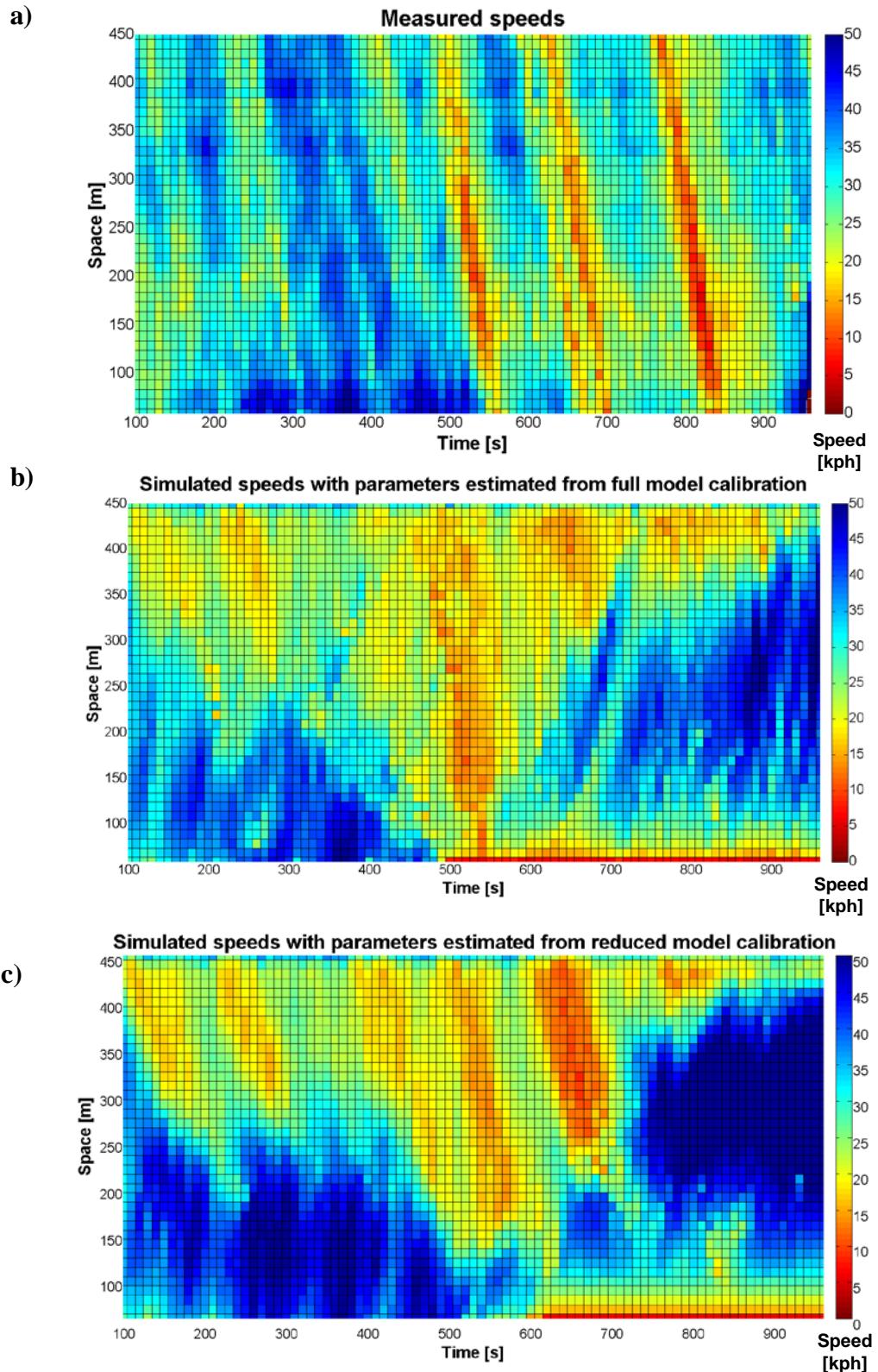
Figure 6.10: Space-Time contour plots of Edie's space mean speeds. (a) refer to measured data, while (b) and (c) to simulated data with parameters estimated in case of *full* and *reduced* model calibrations.
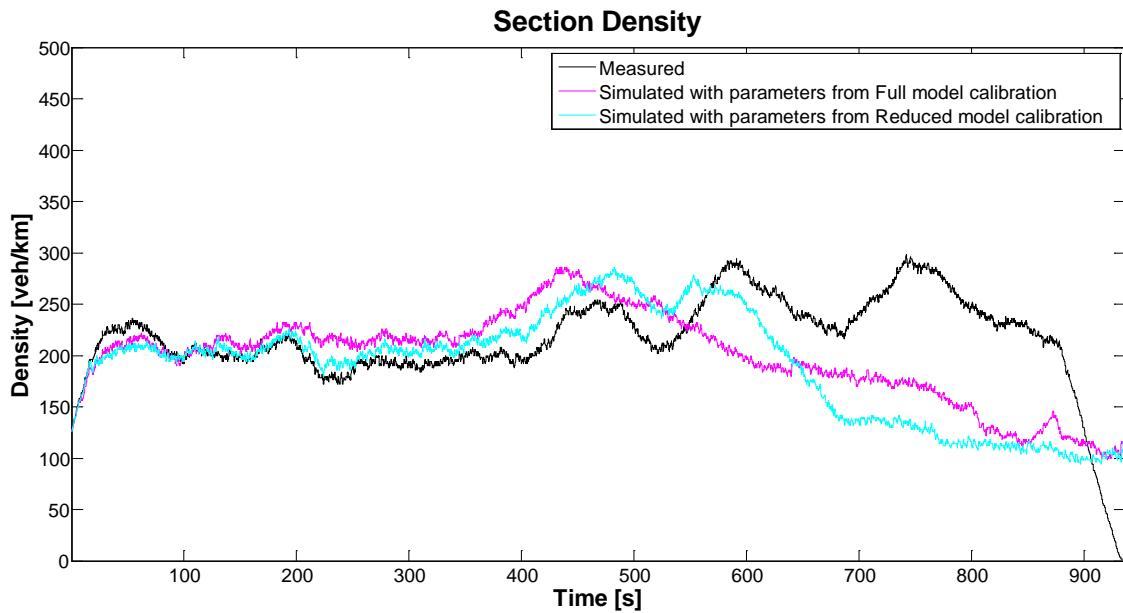
Figure 6.11: Time-series of the section density related to measured (black line) and simulated data, in case of simulation experiments with parameters estimated from *full* (magenta line) and *reduced* (cyan line) model calibration experiments.
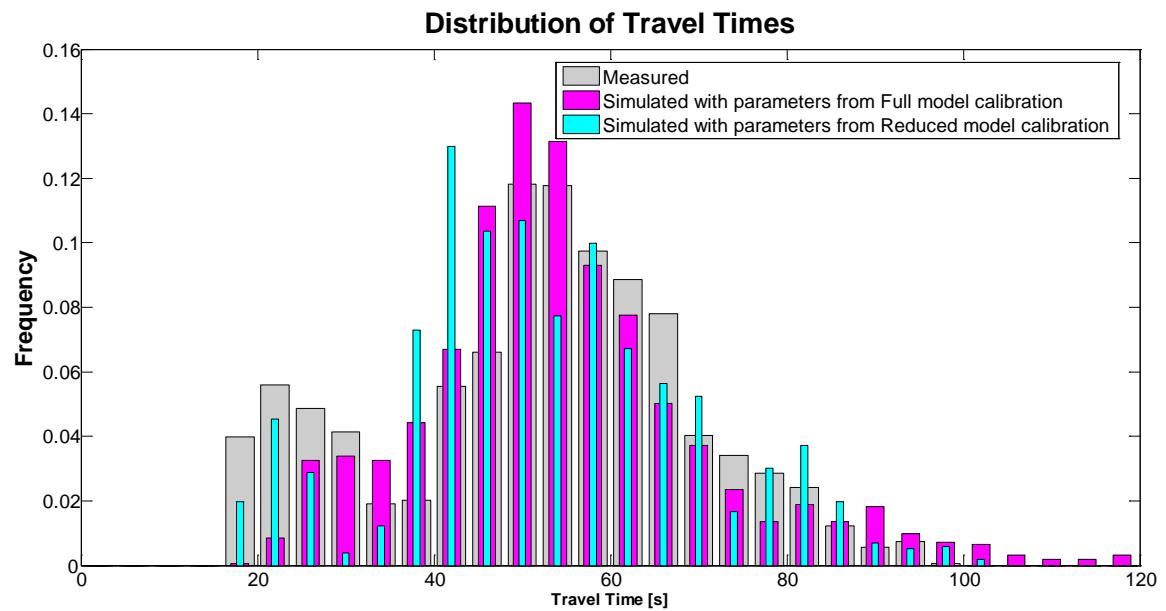
Figure 6.12: Empirical distribution of section travel times for measured (grey bars) and simulated data, in case of simulation experiments with parameters estimated from *full* (magenta bars) and *reduced* (cyan bars) model calibration experiments.
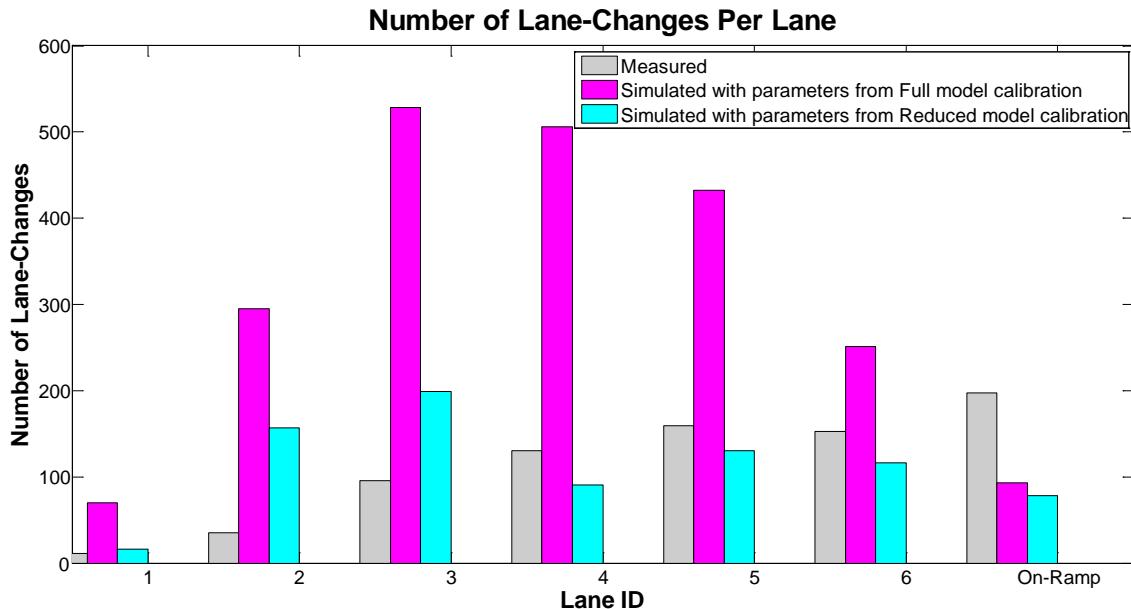
Figure 6.13: Empirical distribution of the number of lane-changes for measured (grey bars) and simulated data, in case of simulation experiments with parameters estimated from *full* (magenta bars) and *reduced* (cyan bars) model calibration experiments.

Finally, a quite surprising results emerges from the comparison of the distribution of the number of lane changes per lane (Figure 6.13). From the figure, it appears that the lane-changing model parameters estimated conditionally to the *reduced* car-following model (i.e. where only two parameters were calibrated) allowed the "aggregate" model to reproduce measured lane-changes distribution much better that in case of the estimation conditioned to the *full* car-following model calibration.

Therefore, interesting conclusions can be drawn from the presented results. Indeed, the comparison between the simulation results highlights that the *reduced* car-following model is likely to outperform the *full* car-following model at the "aggregate" level. This finding is quite surprising, as the *reduced* model had a greater estimation error, compared to the *full* model, at the "disaggregate" level (i.e. car-following model calibration against individual trajectory data). Although not a proof, this finding may be a symptom of model *overfitting* in the *full* model estimation against individual vehicle trajectory data, which, conversely, can be interpreted as a consequence of *over-parameterization*.

6.6   Impacts of Parameter Correlation

In this section, the inner objective is the evaluation of the impact of different driver behavioral model parameter correlation structures on "aggregate" model performances. Indeed, at the best of our knowledge, almost all commercial simulation packages (e.g. AIMSUN, 2012; VISSIM, 2011; PARAMICS, 2003), made the underlying assumption of uncorrelated normal distributed model parameters. However, as shown in Chapter 4, empirical distributions of driver behavioral model parameters (i.e. distribution of estimated model parameters) are far from being normal and present a clear correlation structure.

Indeed, based on the findings discussed in Chapter 4, simulation experiments were carried out assuming the following different distribution models for model parameters:

- *Vehicle-specific estimated parameters*, where each vehicle is assigned with its set of estimated (car-following and lane-changing) model parameters;

- Sampling of model parameters from the *joint empirical distribution* of estimated model parameters (i.e. preserving the correlation structure);

- Sampling of model parameters from the *marginal empirical distributions* of each estimated model parameters (i.e. considering model parameters uncorrelated);

- Sampling of model parameters from *marginal normal distributions* of each estimated model parameters (i.e. neglecting estimated model distributions and considering model parameters uncorrelated);

The methodology here applied is discussed in Section 6.6.1, while main findings are summarized in Section 6.6.2.

6.6.1 Methodology

The methodology applied in this analysis is described in Figure 6.14.

Similarly to the methodological approach depicted in the previous sections, we based our investigation on the estimation results of car-following and lane-changing model parameter distribution (against *reconstructed* data) presented in Chapter 4.
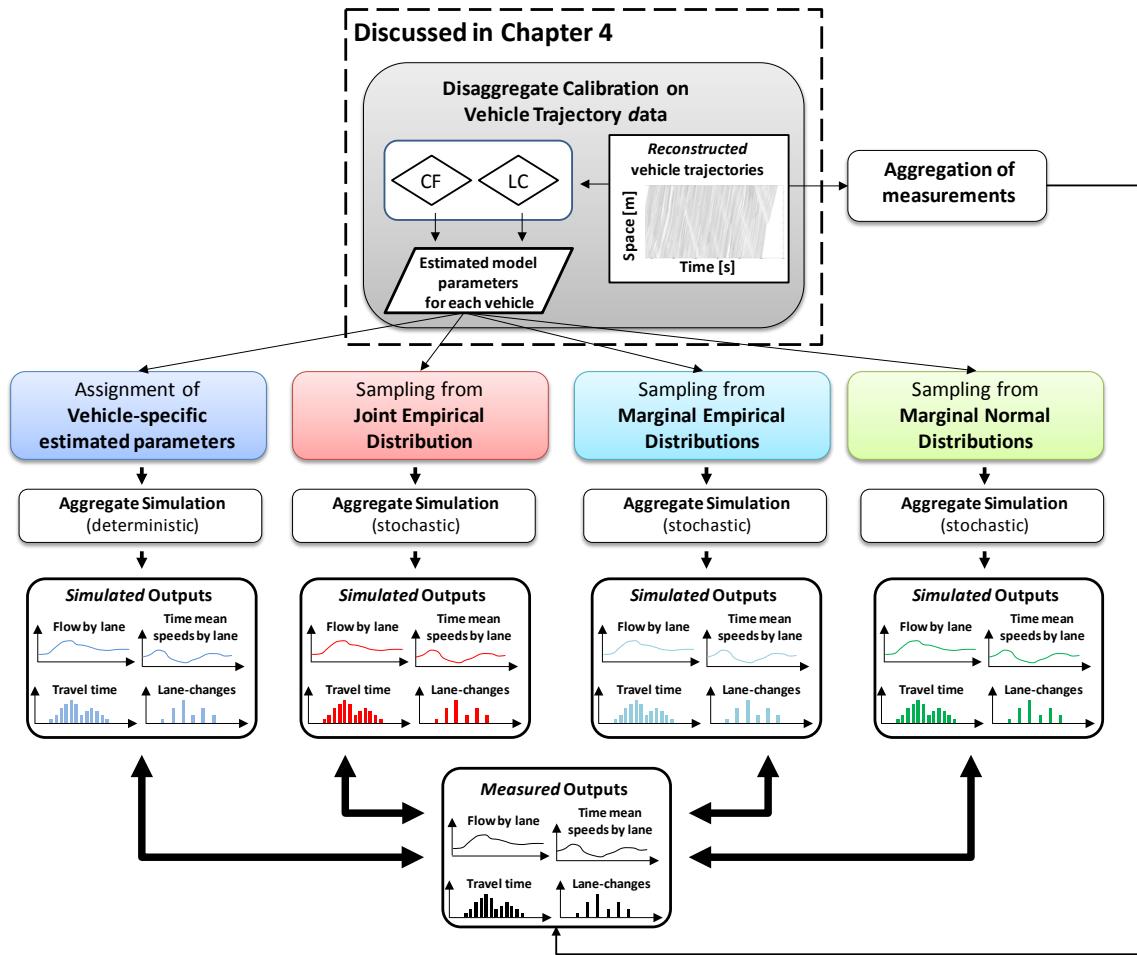
Figure 6.14: Conceptual framework to evaluate the impact of different assumptions on parameter correlation structures on aggregate traffic flow model simulation. Based on the estimated model parameters (and correlation structures) against reconstructed data for each individual vehicle trajectory in the dataset, four different aggregate simulation experiments are performed: *i)* each vehicle is assigned with its set of model estimated parameters, *ii)* the set of model parameters is sampled from the joint empirical distribution, i.e. preserving correlation structures, *iii)* each model parameters is sampled from its marginal empirical distribution, i.e. neglecting parameter correlation, and *iv)* each model parameters is sampled from a marginal normal distribution (estimated on empirical sample). Aggregate simulation results from the four experiments are compared with aggregated measurements over time and space.

It is worth noting that, as in the previous analyses, results from car-following model calibration on the inter-vehicle spacing (and related lane-changing model calibration) were applied herein. For details on this choice, please refer to Chapter 5.

On this basis, we designed four aggregate simulation experiments, where vehicle parameters were assigned based on different modeling assumptions.

Given the design of the experiment, the simulation with *vehicle-specific estimated parameters* (the blue box in Figure 6.14) is exactly the same as the one performed in Section 6.4 (tagged there as "simulation with parameters estimated on *reconstructed* data") and in Section 6.5 (tagged there as "simulation with parameters estimated from *full* model calibration"). In this experiment, each vehicle entering the simulation was assigned with its set of estimated model parameters. Therefore, the simulation experiment is deterministic as no source of variability due to parameter sampling is introduced.

Then, provided estimated model parameter distributions, we designed the remaining simulation experiments by sampling model parameters according to different sampling schemes. It is worth noting that, due to parameters' sampling, the following simulation experiments were stochastic. Therefore, in order to account for such source of variability in model parametric inputs, we ran each simulation experiment 10 times (at fixed random seed), and adopting the same 10 different random sequences for all the simulation experiments.

Further, in order to compare simulation outputs from these stochastic experiments with both the deterministic (i.e. vehicle-specific parameters assignment) simulation outputs and with measurements, we required a unique criteria to indentify the "representative" replication of a stochastic simulation experiment.

In the literature, the definition of "representative" replication for a stochastic simulation experiment is still an open research topic and not conclusive indications were available. Indeed, from a practical point of view, one could perform a deterministic evaluation of model performances (considering the best replication or the median one) or even a probabilistic assessment, thus presenting results in terms of distributions and confidence intervals on time-series. Further, as several simulation outputs are available (e.g. quantities at detectors, section measurements, and so on), the choice of the measure of performance (MoP) at the "aggregate" level could have a great impact on analysis results. Finally, the uncertainty in the choice of the goodness of fit (GOF) function added increase the complexity of the analysis.

However, since our intention was *comparative* and not exploratory, we decided to adopt a selection criteria common to all stochastic experiments, neglecting the degree of influence that the adopted MoP and GOF function could have on the results. Indeed, for

the study purpose, we adopted the sum of the Root Mean Square Errors (RMSE) between the simulated and measured time-series of the time mean speed at all detectors. Different aggregation interval were tested (30, 60 and 120 seconds) and the replication was selected based on the minimum value of the GOF function.

On this basis, we performed different simulation experiments, sampling model parameter according to the following schemes:

- Sampling from the *joint empirical distribution* of estimated model parameters (red box in Figure 6.14): in this case, a *set* of estimated (car-following and lane-changing) model parameters was sampled and assigned to the vehicle entering the simulation, thus preserving the inner correlation structure among estimated model parameters;

- Sampling from the *marginal empirical distributions* of each estimated model parameters (cyan box in Figure 6.14): in this case, each model parameter was *independently* sampled from its estimated distribution and assigned to the vehicle entering the simulation, thus assuming uncorrelated model parameters;

- Sampling from *marginal normal distributions* fitted on the empirical distributions of each model parameters (green box in Figure 6.14): in this case, both the empirical distribution model and the estimated parameter correlation structure is neglected.

On this basis, simulation results from the four experiments were compared with the measured outputs resulting from time-space aggregation of the (*reconstructed*) individual vehicle trajectory from the NGSIM I80-1 dataset.

6.6.2 Results

The comparison between measured and simulation results is conducted with respect to the following measures of performances:

- time-space speed contour plots (Figure 6.15);

- time-series of section density (Figure 6.16);

- distribution of travel times (Figure 6.17);

- distribution of the number of lane-changes (Figure 6.18).

Results are shown in Figures 6.15 – 6.18.

Figure 6.15 presents the space-time contour plots of the Edie's space mean speeds (Edie, 1974) related the measured data (a), and to the simulated one in case of simulation experiment with *vehicle-specific* assigned parameters (b), parameters sampled from *joint empirical* (c), *marginal empirical* (d) and *marginal normal* (e) distributions.

Figure 6.16 shows the comparison between time evolution of the measured and simulated section density in case of the four simulation experiments, while Figures 6.17 and 6.18 present the comparison of the empirical distributions of section travel times (Figure 6.17) and number of lane changes (Figure 6.18).

Results presented in Figures 6.15 – 6.18 highlighted important research outcomes.

The visual inspection of the speed contour plots (Figure 6.15) clearly highlights that the simulation experiment with parameters sampled from marginal normal distributions (Figure 6.15(e)) reproduce observed traffic phenomena quite accurately. Indeed, the back-propagation of the three shockwaves present in the measured data is correctly reproduced, though with less intensity (see the absence of red zones, indicating speed levels lower than 10 kph, as compared to Figure 6.15(a)). Further, also the overall level of congestion in simulated outputs is much closer to the observed one. Finally, no sign of *entrance congestion* (for details, please refer to Section 6.4.2) is present in the simulated outputs.

Similar trends, and even with more emphasis, can be appreciated in the time-series of the section density (Figure 6.16), where the simulated values in case of normal parameters sampling (the green profile) practically matches the observed density (the black profile) almost everywhere.
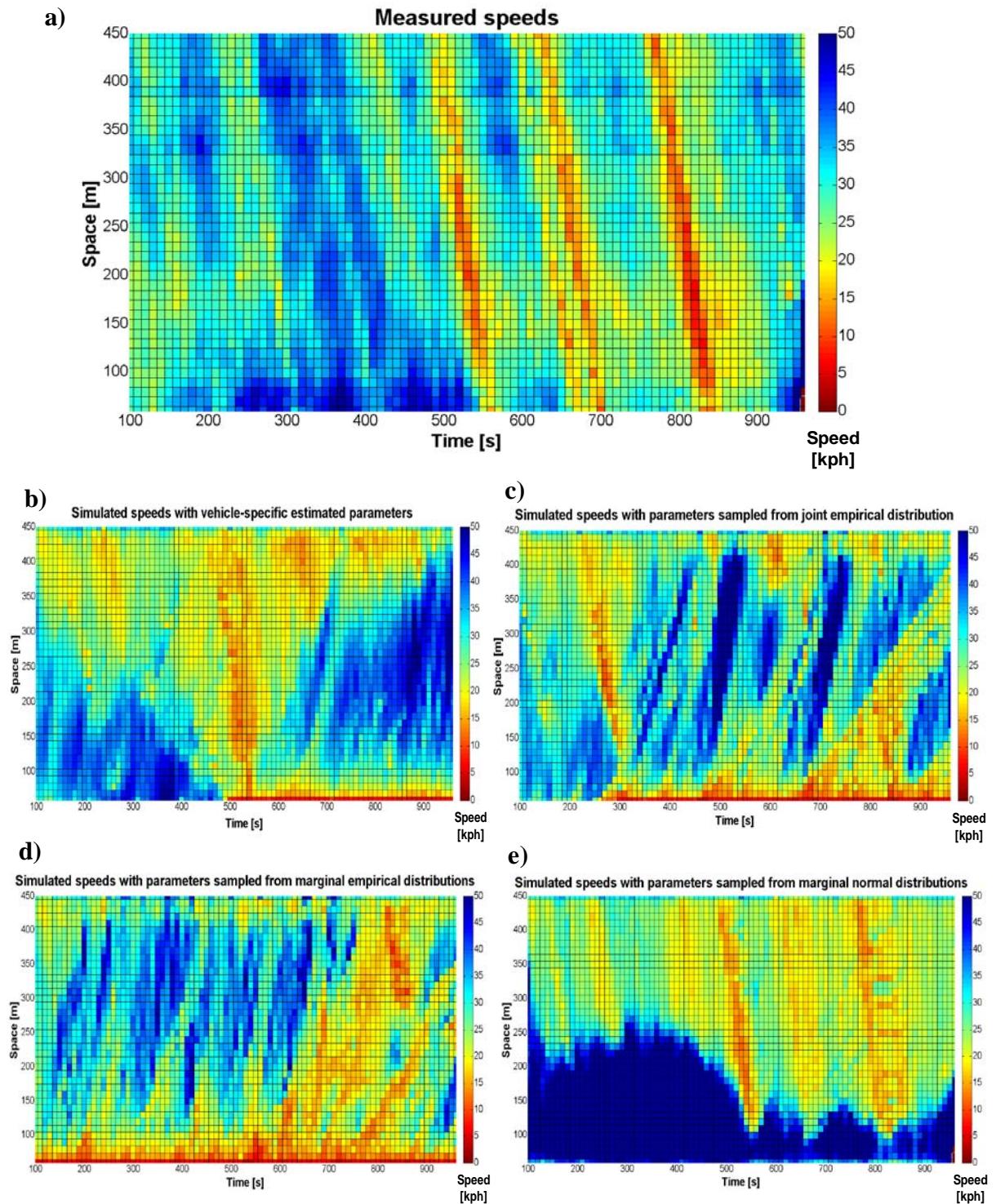
Figure 6.15: Space-Time contour plots of Edie's space mean speeds. (a) refer to measured data, (b) to the simulated data from the experiment with vehicle-specific estimated parameters, (c) to the simulated data from the experiment with parameters sampled from joint empirical distribution, (d) to the simulated data from the experiment with parameters sampled from marginal empirical distributions, and (e) to the simulated data from the experiment with parameters sampled from normal marginal distributions.
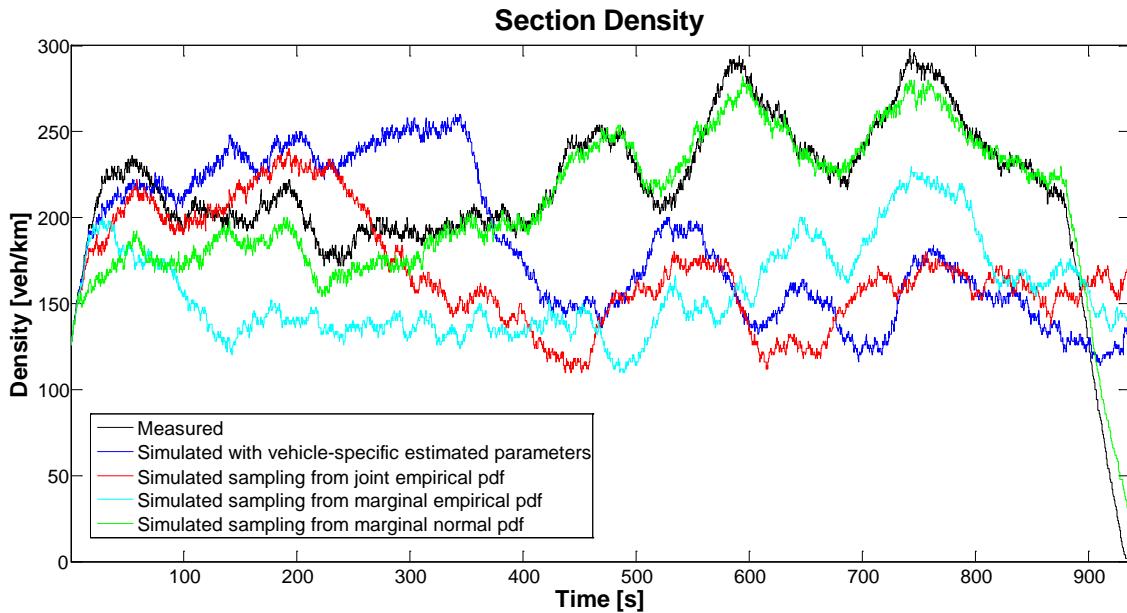
Figure 6.16: Time-series of the section density related to measured (black line) and simulated data, in case of simulation experiments with vehicle-specific estimated parameters (blue line), and with parameters sampled from joint empirical (red line), marginal empirical (cyan line) and marginal normal (green line) distributions.
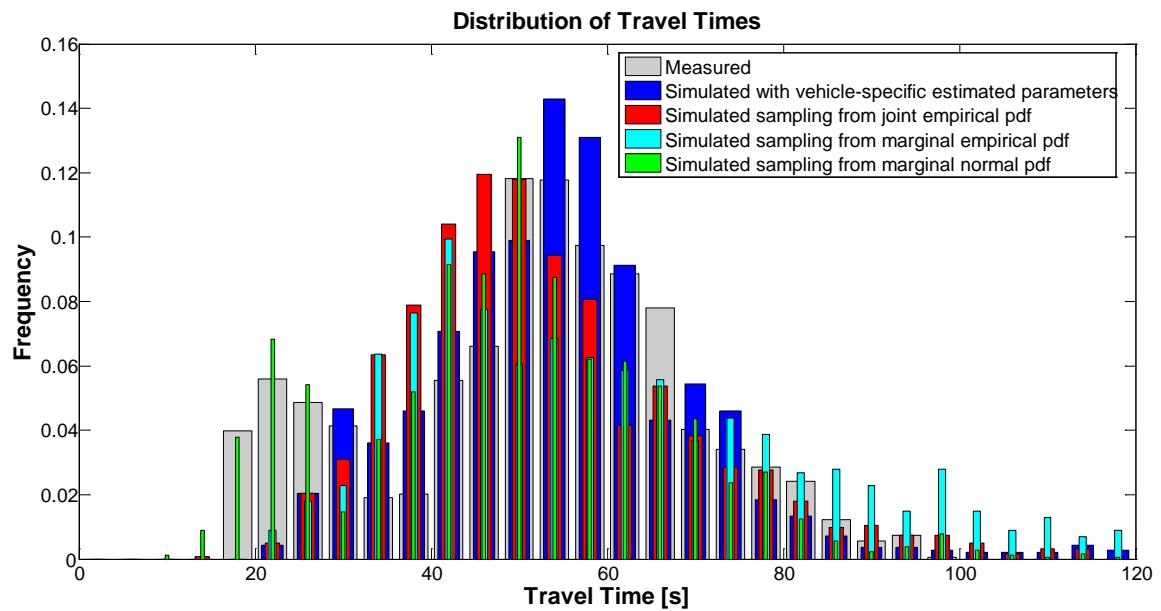
Figure 6.17: Empirical distribution of section travel times for measured (grey bars) and simulated data, in case of simulation experiments with vehicle-specific estimated parameters (blue bars), and with parameters sampled from joint empirical (red bars), marginal empirical (cyan bars) and marginal normal (green bars) distributions.

Number of Lane-Changes Per Lane
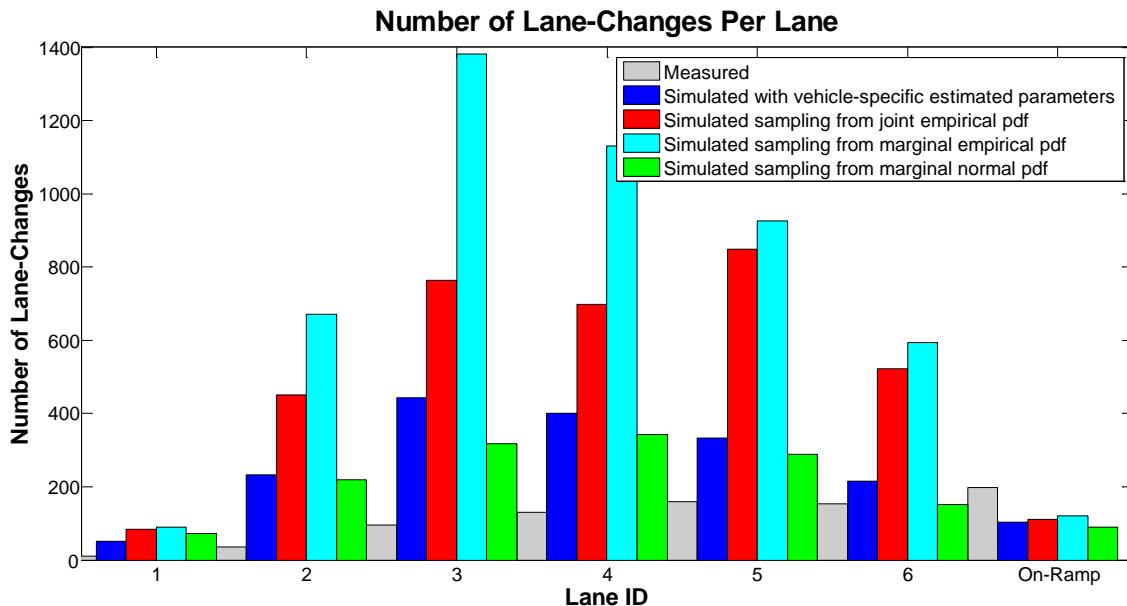
Figure 6.18: Empirical distribution of the number of lane-changes for measured (grey bars) and simulated data, in case of simulation experiments with vehicle-specific estimated parameters (blue bars), and with parameters sampled from joint empirical (red bars), marginal empirical (cyan bars) and marginal normal (green bars) distributions.

In terms of travel times distribution (Figure 6.17), simulated outputs from this experiment (green bars) outperformed the others and are able to match even the higher speed levels in the HOV lane (the peak in the frequency of lower travel times, around 20 seconds) which, instead, is not captured in the other simulation experiments.

Similar results can be also appreciated on the distribution of the number of lane-changes per lane (Figure 6.18).

The findings described above are quite surprising and strongly unforeseen. Indeed, the analysis performed in Chapter 4 at the "disaggregate" level showed that the estimated distributions of model parameters were far from being normal, and with a clear correlation structure. Therefore, we expected to appreciate even more this difference at the "aggregate" level. However, results shown in Figure 6.15 – 6.18 did not confirm our guess.

From a general point of view, more than the results, this preliminary analysis pointed out an important lack in the field research on the study of microscopic traffic flow simulation modeling. Indeed, it is not yet clear which is the degree of relationship

between the findings at the level of model components (sub-models) and their impact on aggregate traffic flow simulation.

What emerges from the analysis is that *the sum of components may be less than the aggregate model*, meaning that the results of analysis at the disaggregate level (sub-models) may not capture important aspects of aggregate traffic.

Therefore, though "disaggregate" analysis deemed to have a great importance for modeling, there might be no relationship between the results at the "disaggregate" levels with those at the "aggregate" one. Indeed, preliminary findings shown in this section suggest that the study of sub-models could not be undertaken as a mean to infer on "aggregate" model performances.

6.7 Summary

In the previous Chapters, we applied the four-steps uncertainty management framework to car-following and lane-changing models, separately, in order to understand the impact of the different sources of *parametric* and *non-parametric* uncertainty on model performances.

However, one could question about the impact of this findings at the level of the "aggregate" traffic flow simulation model.

Therefore, in this Chapter we investigated the impacts on "aggregate" simulation model performances of the results from previous studies on sub-models. In particular, we focused on the following three topics:

- Analysis of the impact of "disaggregate" model calibration in presence of *measurement errors*;

- Analysis of the impact of "disaggregate" model simplifications;

- Analysis of the impact of "disaggregate" model parameters correlation;

For the scope, a trace-driven "aggregate" microscopic traffic flow simulation model was designed in order to evaluate model performances with respect to the measured data from the (*reconstructed*) NGSIM I80-1 dataset.

Main findings are summarized in the following.

Similarly to the findings on sub-models, measurement errors in individual vehicle trajectory data play a little role on "aggregate" simulation model performances, and timidly suggest that the simulation outputs from the experiment with parameters estimated in presence of measurement errors present higher level of congestion.

Moving to the analysis of the impact of model simplifications, the comparative study suggest that the *reduced* car-following model (i.e. where only most influential model parameters were estimated) is likely to outperform the *full* one at the "aggregate" level. This finding is quite surprising, as the *reduced* model had a greater estimation error, compared to the *full* model, at the "disaggregate" level (i.e. car-following model calibration against individual trajectory data). Although not a proof, this finding may be a symptom of model *overfitting* in the *full* model estimation against individual vehicle trajectory data, which, conversely, can be interpreted as a consequence of *over-parameterization*.

Finally, results from the analysis of the impact of parameter correlation structures on "aggregate" simulation model performances suggested that the "disaggregate" calibration of sub-models may not capture important aspects that, in turn, can be revealed only when focusing directly on the "aggregate" model. Indeed, though "disaggregate" analysis deemed to have a great importance for modeling itself (e.g. understanding model behavior), there might be no relationship between the results at the "disaggregate" levels and those at the "aggregate" one. Indeed, results suggest that the study of sub-models could not be undertaken as a mean to infer on "aggregate" model performances.

Chapter 7

Conclusions

7.1. Summary

In this dissertation thesis, we proposed and specified a common methodological framework for quantitative management of uncertainty in microscopic traffic flow simulation modeling.

The approach followed in this study (Chapter 1) builds up on techniques, initially established in the industrial practice, that are increasingly applied in many modeling fields including environmental, climate and financial ones, as well as, in system reliability and risk analysis.

In Chapter 2, we presented the specialization of the conceptual framework for quantitative uncertainty assessment to the study of microscopic traffic flow simulation models. In particular we showed that different sources of uncertainty may invest microscopic traffic flow simulation modeling at multiple levels. On this basis, we investigated how parametric and non-parametric sources of uncertainty affect model performances in reproducing measured vehicle trajectory data. It is worth noting that the focus of this study is on driver behavioral models only, and, specifically, on well-known car-following and lane-changing models proposed in the literature. In the following, a summary of each chapter related to the steps of the uncertainty management cycle is presented.

In Chapter 3 we explored the impact of the uncertainty in the calibration procedure on the results of car-following model parameter calibration. The model investigated in this Chapter was the well-known Gipps car-following model (Gipps, 1981). We proposed a methodological framework to verify the suitability of a calibration setting – here intended as the combination of the optimization algorithm, the goodness of fit function and the measure of performance – in finding the global solution of a "black-box" optimization problem. The methodology was based on the use of synthetic vehicle trajectory data, as this is the only way to ascertain the ability of a calibration setting to discover the global optimum. Compact indicators were proposed to evaluate the capability of a calibration setting to find the "known" global solution, in terms of both the accuracy and robustness. Then, a novel graphic inspection method, based on the so called *Cobweb* plots, was used to explore the existence and the nature of the local minima found by the algorithms, as well as to give insights into the measures of performance and the goodness of fit functions used in the calibration experiments.

In Chapter 4, we focused on measurement errors in vehicle trajectory data, and evaluated their impact on the results of driver behavioral model parameter calibration. To this aim, we used vehicle trajectory data from the NGSIM I80 dataset, whose low degree of accuracy in terms of several criteria is widely recognized in the transportation community. Therefore, we first proposed a multi-step procedure for vehicle trajectory reconstruction. The methodology aimed at eliminating the main inconsistencies and noise from *raw* measurements while preserving *i)* the actual driving dynamics (vehicle stoppages, shifting gears, etc.), *ii)* the *internal* consistency of trajectories (i.e. the consistency among space travelled, speed and acceleration) and *iii)* the platoon consistency (i.e. the actual inter-vehicle spacing). Successively, provided both *raw* and *reconstructed* vehicle trajectory data, we evaluated the impact of (real) measurement errors on results of car-following and lane-changing model parameter calibration. In particular, calibration results showed that the model operates like a "filter", and the impact of the measurement errors on parameter estimation (and on their correlation structures) is very limited. These findings are not in line with the results of Ossen and Hoogendoorn (2008a, 2009) where calibration experiments were performed using synthetic data with *normally distributed* error structures added ex-post. A possible explanation of such difference could be due to the considerably different structure of the error ,which was empirically obtained here by comparing *raw* and *reconstructed* data.

In Chapter 5, we focused on the asymmetry in the importance of driver behavioral model parameters in influencing the variability of model performances. Therefore, we proposed a methodological framework to verify whether it is possible to reduce the number of parameters to calibrate without sensibly affecting the capacity of the model to reproduce the true output variance. To this aim, variance-based sensitivity analysis is applied, in a *factor fixing* setting, to the Intelligent Driver Model (IDM, Treiber et al., 2000). Since results of a sensitivity analysis are conditioned to the values of the fixed inputs, one could argue that ranking of importance of car-following model parameters is specific to the selected leader/follower vehicle trajectory, which is respectively used for model simulation and estimation error calculation. Therefore, in the proposed framework we considered also the input trajectories as uncertain, and the investigation has been extended by including all the trajectories of the NGSIM I80-1 dataset, as *reconstructed* in Chapter 4. Besides the robustness of the analysis with regards to the *factor fixing* setting, the inclusion of more than two thousands input trajectories allowed us to investigate the model against a significant variety of driver behaviors. Results showed that the input trajectory is the most influential factor both in terms of first-order effect and in interactions with the model parameters. The variance of the model error conditional on the input trajectory – a function of the parameter combination sampled – has been therefore suggested as a measure of the "risk" of choosing a non-optimal model parameter combination: the higher the variance, the higher the risk of incurring in big modeling errors. It has also been shown, graphically, that such a variance is a decreasing function of the trajectory duration. This is an empirical evidence that car-following models should be calibrated on "long" trajectories. Based on the outlined sensitivity ranking of model parameters, we proposed two alternative model simplifications by fixing the non-influential parameters to common values adopted in the literature, and assessed their performances in comparison with the original model formulation where all parameters were considered uncertain. Comparison results confirmed that performances of the reduced model are very close to the full model ones, with an average increase of about 6% in the estimation error on speeds, and 30% on spacing. On the other hand, model simplifications turned into a remarkable benefit in the computational effort required for model calibration, reducing of about 50% the number of model evaluations needed for convergence of the optimization algorithm.

Finally, in Chapter 6, we explored the relationship between "disaggregate" model calibration and simulation performances of a micro-simulation software. Indeed, in micro-simulation software, driver behavioral models strongly interacts with each other in order to emulate traffic flows. However, there is no explicit relationship between the results of analyses carried out on each model component separately – here indicated as "disaggregate" models or sub-models – and the performances of micro-simulation software – here indicated with "aggregate" model. Therefore, we investigated the impact of different sources of uncertainties in "disaggregate" modeling on "aggregate" simulation performances.

In particular, based on the results of Chapter 4, we evaluated the impact of measurement errors in vehicle trajectory data on "aggregate" performances. Results showed that, similarly to the findings on sub-models, measurement errors in individual trajectory data play a little role on "aggregate" simulation performances.

Further, based again on the results of Chapter 4, we investigated the impact of model parameter correlation structures on "aggregate" simulation performances. Interestingly, results suggested that the analysis of sub-models may not capture important aspects that can be revealed only when focusing directly on the "aggregate" model. Indeed, though "disaggregate" analysis deemed to have a great importance for modeling itself (e.g. understanding model behavior), correlation structures estimated at the "disaggregate" levels may have a great impact on "aggregate" performances, and there might be no relationship between the results at the "disaggregate" levels with those achievable at the "aggregate" one.

Based on results of Chapter 5, we explored the impact of disaggregate model simplifications on the simulation performances of the aggregate model. Results suggest that the reduced model (i.e. where only most influential model parameters were estimated) is likely to outperform the full model at the "aggregate" level. This finding is quite surprising, as the reduced model had a greater estimation error, compared to the full model, at the "disaggregate" level (i.e. car-following model calibration against individual trajectory data). Although not a proof, this finding may be a symptom of model overfitting in the full model estimation against individual vehicle trajectory data, which, conversely, can be interpreted as a consequence of over-parameterization.

7.2   Contributions

In this section we summarized the main contributions of the dissertation thesis to the state-of-the-art in the field of microscopic traffic flow simulation modeling. The following list is provided by research topic, and thus multiple contributions on the same subject were aggregated.

- We presented and applied a methodological framework for the management of both parametric and non-parametric sources of uncertainty in microscopic traffic flow simulation models. Specifically, we focused on driver behavioral models only, i.e. car-following and lane-changing models. The methodological approach followed in this thesis is based on an ensemble of techniques established in the industrial practice and increasingly applied in many modeling fields including environmental, climate and financial ones, as well as, in system reliability and risk analysis. The framework relies on the following four steps: *i)* problem specification, *ii)* uncertainty modeling, *iii)* uncertainty propagation and *iv)* sensitivity analysis.

- We quantified the uncertainty entailed in the calibration procedure, and its impacts on the accuracy and reliability of results. For the scope, we proposed a general framework to verify the goodness of a calibration setting, based on the use of synthetic data. The robustness of the calibration setting was then quantified through a set of synthetic indicators and the use of graphical methods, such as the *Cobweb* plots. The methodology was apply to evaluate the correctness and robustness of most of the calibration settings specified in the literature – i.e. different combination of the choice of the measure of performance, the goodness of fit function and the optimization algorithm.

- We proposed a robust methodology for the "disaggregate" calibration of car-following and lane-changing models against individual vehicle trajectory data. With regards to the calibration of car-following models, the novelty of the approach consists in considering a goodness of fit function evaluated in the frequency domain, that allow a robust comparison between observed and simulated time-series of the measure of performance. With regards to the calibration of non-stochastic lane-changing models, the approach relies on the innovative concept of "scenario" which allows to test the model capability of reproducing observed lane-changing choices preserving correlation among input variables.

- We explored the impact of measurement errors in vehicle trajectory data on results of car-following and lane-changing model calibrations. For the scope, we proposed a general procedure for vehicle trajectory reconstruction, aimed at eliminating the main inconsistencies and noise from *raw* measurements while preserving *i)* the actual driving dynamics (vehicle stoppages, shifting gears, etc.), *ii)* the *internal* consistency of trajectories (i.e. the consistency among space travelled, speed and acceleration) and *iii)* the platoon consistency (i.e. the actual inter-vehicle spacing). The procedure was applied to the NGSIM I80-1 dataset and successfully restored data consistency. *Reconstructed* data are publicly available for downloading on the MULTITUDE website (MULTITUDE, 2014).

- We proposed a robust methodology to simplify models based on the identification of the parameters of microscopic traffic flow simulation models that have greater influence on the variability of model output. To this aim, variance-based techniques for global sensitivity analysis are formulated in a factor fixing setting. Among the main contributions are: *i)* a novel formulation for the factor fixing setting, where the "model performance", instead of the "model output", is adopted as quantity of interest (i.e. a measure of the distance between simulation and reality); *ii)* a robust design of the Monte Carlo framework for the sensitivity analysis that also includes, as an analysis factor, the main non-parametric input of car-following models that is the leader's trajectory; *iii)* a set of general criteria for "data assimilation" in car-following models, i.e. to set the parameter bounds for the model sensitivity analysis and calibration.

- We developed a microscopic traffic flow simulation tool to perform a trace-driven simulation studies. This allowed us to evaluate the impact of measurement errors in vehicle trajectory data, of correlation structure among estimated model parameters, and of different model simplifications, on the performances of the "aggregate" simulation model.

- We reviewed the existing literature formulation of the Gipps car-following model (Gipps, 1981), and proposed an enhanced model version in order to generalize the "acceleration component" of the model.

7.3   Final Considerations

We divided this section into five parts. The first part contains general considerations that summarize the most relevant findings from all chapters. The remaining four parts pertain to the findings of Chapter 3 (uncertainty in the estimation procedure), Chapter 4 (uncertainty in vehicle trajectory data), Chapter 5 (sensitivity analysis of driver behavioral models), and Chapter 6 (impacts on "aggregate" simulation performances).

7.3.1   General considerations

- A methodological framework for the quantitative management of the different sources of uncertainty entailed in traffic flow simulation models is deemed to be necessary to enhance credibility of such models and the reliability of their predictions.

- Global Sensitivity Analysis can support model development at different levels. In particular, it can be adopted to guide *i)* model simplifications, based on the identification of the importance ranking of parametric inputs, *ii)* model verification, to ascertain whether the model is overly dependent on fragile assumptions/structures, and *iii)* parameter identification, based on the recognition of critical regions in the space of parameters that lead to undesirable model behaviors (e.g. unrealistic stop-and-go wave speed).

- There is a non-trivial relationship between results of the analyses performed on driver behavioral models – here referred as "disaggregate" level of investigation – and the results of a traffic micro-simulation – here referred as "aggregate" level of investigation – where different sub-models (e.g. car-following and lane-changing models) constantly interact with each other as to emulate traffic flows. This reflects in the fact that the adoption of model parameters estimated by "disaggregate" calibration does not necessarily imply "optimal" simulation performances at the "aggregate" level.

7.3.2   Uncertainty in model calibration procedure

- The compensation of the modeling errors and of the system irreducible uncertainty is the basic theoretical motivation for the indirect estimation of model parameters in traffic simulation. Conversely, it generally advises against the direct estimation of the observable parameters, namely, of those parameters which have a physical equivalent in the reality and can be directly measured, like for example the reaction time or the maximum acceleration in car-following models.

- In car-following model calibration, the choice of the measure of performances, goodness of fit function and optimization algorithm have a great impact on the quality of the solution, i.e. the estimated model parameters. This confirmed the complexity of the problem of calibrating car-following models against real trajectory data. More specific conclusions follow:

  o GOF functions based on the GEH statistics are highly affected by the setting of the threshold value. When used in calibration, a wrong setting of this value lead to the loss of uniqueness of the global solution, even in the case of optimization problems on synthetic data, where the global minimizer is unique and well-defined.

  o The Downhill Simplex proved to be not suitable for model calibration. Further, the heuristics was very sensible to the initial starting condition, providing very different sets of optimal parameters depending on the starting point.

  o The Genetic Algorithm and the OptQuest Multistart robustly and repeatedly found the global solution in a synthetic experiment, and therefore are very likely to be adopted in calibration studies.

  o Calibrating the model against the spacing between the leader and the follower gives acceptable results also in terms of the vehicle speed, while the opposite is not equally true.

  o The use of mixed GOF functions that combine both the MoPs (speed and spacing), such as the sum of Theil's Inequality coefficients, performed worse than other functions evaluated on speed or on spacing. Further, the use of absolute measures of the distance between observed and (model) simulated outputs, such as the MAE, entails very low efficiency in the optimization, as

they require a high number of evaluations of the objective functions to satisfy the same stopping rules adopted with the other GOF functions. Moreover, the improvements in finding the global minimizer are negligible.

7.3.3   Uncertainty in vehicle trajectory data

- Vehicle trajectory data from the NGSIM Program (2005) are affected by a large amount of measurement errors, which limit their applicability for any studies in the traffic flow theory. In particular, two types of measurement errors can be recognized: the measurement errors in the "LocalYs" that produce the greatest bias in the accelerations – here called *outliers* – and the residual errors (*noise*). These errors considerably alter both *internal* and *platoon* consistency of vehicle trajectory data.

- The impact of measurement errors in vehicle trajectory data on results of "disaggregate" calibration of car-following and lane-changing model parameters is very limited. In facts, results showed that the model operates like a "filter". These findings are not in line with the results of Ossen and Hoogendoorn (2008a, 2009) where calibration experiments were performed using synthetic data with *uncorrelated Gaussian* error structures added ex-post. A possible explanation of such difference could be due to the substantially different distribution model of the real error structure, which is here obtained empirically by comparing *raw* and *reconstructed* data.

7.3.4   Global Sensitivity Analysis of driver behavioral models

- Results of the global sensitivity analysis carried out on car-following models showed that the leader's trajectory is considerably more important than the parameters in affecting the variability of model performances. Results also unveiled that such variability is a function of the trajectory duration. In particular, as long as duration increases – and so does the exposition to car-following dynamics – the variability of model performances over the parameters' space diminishes. This confirms that in order to encompass heterogeneity of driver behaviours, model

parameters need to be calibrated, and that long trajectories are required for robust estimation.

- A strong asymmetry in the sensitivity of model parameters translated into a very small number of parameters accounting for most of the variability of model output and, consequently, influencing model performances. In the case of the Intelligent Driver Model (IDM; Treiber, 2000), among the six model parameters, the "minimum time headway" explained most of the variance of the error measure, followed by "alpha", "maximum acceleration" and "maximum speed". Further, the importance ranking of IDM parameters was the same when using speed or spacing as measure of performance, though the magnitude of parameter sensitivity resulted different between the two.

- Performances of the simplified models based on the importance ranking, i.e. by fixing non-influential parameters to common values adopted in the literature, are very similar to those of the original model version, where all parameters were considered uncertain. On the other hand, model simplifications turned into a remarkable benefit in the computational effort required for model calibration.

7.3.5   *Impacts on aggregate simulation performances*

- Measurement errors in individual vehicle trajectory data play a little role on aggregate simulation performances, and timidly suggest that the simulation outputs from the experiment with parameters estimated in presence of measurement errors present higher level of congestion.

- Results from the analysis of the impact of parameter correlation structures on aggregate simulation performances suggested that the "disaggregate" calibration of driver behavioral models may not capture important aspects that, in turn, can be revealed only when focusing directly on the "aggregate" simulation model. More specifically, correlation structure among model parameters substantially influence "aggregate" simulation performances. Further, "disaggregate" calibration of model parameters does not necessarily imply "best" simulation performances of the "aggregate" model.

- Moving to the analysis of the impact of model simplifications on "aggregate" performances, the comparative study suggest that the *reduced* model (i.e. where only most influential model parameters were estimated) is likely to outperform the *full* model at the "aggregate" level. This finding is quite surprising, as the *reduced* model had a greater estimation error, compared to the *full* one, at the "disaggregate" level. Although not a proof, this finding may be a symptom of model overfitting in the *full* model estimation, which, conversely, can be interpreted as a consequence of *over-parameterization*.

7.4   Future research

Future research is crucial to exploit the full potential of microscopic traffic flow simulation models in traffic forecasting. In particular, the following research lines may be addressed:

- To investigate the relationship between the variability of real-world observations (day-to-day variability of supply/demand) and the stochasticity predicted by the model.

- To perform uncertainty management of microscopic traffic flow simulation software where both the demand and supply are considered uncertain, whereas current simulation practice does consider only variability in the supply.

- To seek for "global" Goodness of Fit (GOF) functions able to capture the inner structure/driving behaviour/driving style contained in the trajectory data, as expressed/interpreted by the specific model in use. This is also in the course of the recent studies performed by Chiabaut et al. (2010). Local GOF indeed are sensitive to errors in the data, and especially least square ones tend to compensate errors over the whole length of the trajectory;

- To develop methodologies to appropriately bound the space of the admissible inputs in order to preserve the well established macroscopic characteristics of the traffic flow (e.g. maximum stop-and-go wave speed at population level). This can be envisaged by performing Regionalized Sensitivity Analysis of model outputs following a Monte Carlo Filtering approach (e.g. Young et al., 1978; Hornberger and Spear, 1981; Spear et al., 1994; Young et al. 1996; Young 1999).

Bibliography

Ahmed, K. I. (1999). Modeling Drivers' Acceleration and Lane Changing Behavior. *PhD Thesis*. Massachusetts Institute of Technology.

Ahn, S., Z. Zheng, D. Chen, J. Laval (2013). The effects of lane-changing on the immediate follower: Anticipation, relaxation and change in driver characteristics. *Transportation Research Part C*, 26, 367–379.

AIMSUN (2012). *Aimsun Dynamic Simulators Users Manual*, v7. TSS, Barcelona.

Antoniadis, A., J. Berruyer, and R. Carmona (1992). Regression Non Lineaire et Applications. *Collection Economie et Statistiques Avanc´ees*, Economica.

Antoniou, C., J. Barcelo, M. Brackstone, H. B. Celikoglu, B. Ciuffo, V. Punzo, P. Sykes, T. Toledo, P. Vortisch, and P. Wagner (2014). Traffic Simulation: Case for guidelines. M. Brackstone and V. Punzo editors, *JRC Scientific Report*.

Archer, G., A. Saltelli, I. M. Sobol' (1997). Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58, 99–120.

AUSTROADS (2006). *The use and application of microsimulation traffic models*. Austroads Research Report, AP-R286/06, Austroads, Inc.

AUSTROADS (2010). *Guidelines for Selecting Techniques for the Modelling of Network Operations*. Austroads Research Report, AP-R350/10, Austroads, Inc.

Bando, M., K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama (1995). Dynamical model of traffic congestion and numerical simulation. *Physical Review E*, 51(2), 1035–1042.

Bartin, B., K. Ozbay, O. Yanmaz-Tuzel, O., and G. List (2006). Modeling and Simulation of Unconventional Traffic Circles. *Transportation Research Record*, 1965, 201–209.

Bayarri, M. J., J. O. Berger, G. Molina, N. M. Rouphail and J. Sacks (2004). Assessing Uncertainties in Traffic Simulation: A Key Component in Model Calibration and Validation. *Transportation Research Record*, 1876, 32–40.

Beegala, A., J. Hourdakis, and P. G. Michalopoulos, (2005). Methodology for Performance Optimization of Ramp Control Strategies Through Microsimulation. *Transportation Research Record*, 1925, 87–98.

Beven, K. J. and A. M. Binley (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6, 279–298.

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16, 41–51.

Beven, K. (2001). *Rainfall-Runoff Modelling: The Primer*. John Wiley & Sons, Ltd.

Bexelius, S. (1968). An Extended Model for Car-Following. *Transportation Research*, 2(1), 13–21.

Bloomberg, L., and J. Dale (2000). A Comparison of the VISSIM and CORSIM Traffic Simulation Models On A Congested Network. *Transportation Research Record*, 1727, 87–98.

Bonsall, P., R. Liu, and W. Yang (2005). Modelling safety-related driving behaviour— impact of parameter values. *Transportation Research Part A*, 39, 425–444.

Box, G. E., W. G. Hunter, and J. S. Hunter (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd edition. Wiley-Interscience.

Brackstone, M., and M. McDonald (1999). Car-following: a historical review. *Transportation Research Part F*, 2, 181–196.

Brackstone, M., M. Montanino, W. Daamen, C. Buisson and V. Punzo (2012). Use, Calibration and Validation of Traffic Simulation Models in Practice: Results of a

Web based Survey. *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, Washington, D.C.

Bratley, P. and Fox, B. L. (1988). Algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Trans. Math. Software*, 14, 88–100.

Breiman, L. (2001). Random Forest. *Machine Learning*, 45 (1), 5–32.

Brockfeld, E., R. D. Kühne, A. Skabardonis, and P. Wagner (2003). Toward benchmarking of microscopic traffic flow models. *Transportation Research Record*, 1852, 124–129.

Brockfeld, E., R. D. Kühne, and P. Wagner (2004). Calibration and Validation of Microscopic Traffic Flow Models. *Transportation Research Record*, 1876, 62–70.

Brockfeld, E., R. D. Kuhne, and P. Wagner (2005). Calibration and Validation of Microscopic Models of Traffic Flow. *Transportation Research Record*, 1934, 179–187.

Broersen, P. (2006). *Automatic Autocorrelation and Spectral Analysis*. Springer-Verlag.

Butterworth, S. (1930). On the Theory of Filter Amplifiers . *Experimental Wireless and the Wireless Engineer*, 7, 536–541.

Campolongo, F., J. Cariboni and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22, 1509–1518.

Campolongo, F., J. Cariboni and A. Saltelli (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communication*, 22, 182, 978–988.

Cao, B., and Z. Yang (2009). Car-Following Models Study Progress. *Proceedings of the Second International Symposium on Knowledge Acquisition and Modeling*, 3, 190-193.

Cascetta, E. (2009). *Transportation Systems Analysis: Models and Applications*, 2nd edition. Springer.

Chen, A., H. Yang, H. K. Lo, W. H. Tang (2002). Capacity reliability of a road network: an assessment methodology and numerical results. *Transportation Research Part B*, 36, 225–252.

Chiabaut, N., L. Leclercq, and C. Buisson (2010). From heterogeneous drivers to macroscopic patterns in congestion. *Transportation Research Part B*, 44, 299–308.

Chien, S. I., L. N. Spasovic, S. S. Elefsiniotis, R. S. Chhonkar (2001). Evaluation of Feeder Bus Systems with Probabilistic Time-Varying Demands and Nonadditive Time Costs. *Transportation Research Record*, 1760, 47–55.

Chilès, J-P. and P. Delfiner (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, Ltd.

Choudhury, C. F. (2007). Modeling driving decisions with latent plans. *PhD. Thesis*, Massachusetts Institute of Technology.

Ciuffo B., V. Punzo, and V. Torrieri (2007). A framework for calibrating microscopic traffic simulation models, *Proceedings of the 86th TRB Annual Meeting, Washington D.C.*

Ciuffo, B., V. Punzo and V. Torrieri (2008). Comparison between simulation-based and model-based calibrations of traffic flow micro-simulation models. *Transportation Research Record*, 2088, 36–44.

Ciuffo, B., and V. Punzo (2010). Verification of traffic micro-simulation model calibration procedures: analysis of Goodness-of-Fit measures. *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Ciuffo, B., V., Punzo, and M. Montanino (2012a). The calibration of traffic simulation models. Report on the assessment of different goodness of fit measures and optimization algorithms. *JRC Scientific Report*.

Ciuffo, B., V. Punzo and M. Montanino (2012b). 30 years of the Gipps' car-following model: applications, developments and new features. *Transportation Research Record*, 2315, 89–99.

Ciuffo, B., J. Casas, M. Montanino, J. Perarnau, and V. Punzo (2013). Gaussian process metamodels for sensitivity analysis of traffic simulation models. *Transportation Research Record*, 2390, pp. 87–98. The TRB Joint Simulation Subcommittee peer-reviewed this paper, which was awarded the 2014 Paper Award by the committee.

Clark, S. D. and D. P. Watling (2000). Probit-based sensitivity analysis for general traffic networks. *Transportation Research Record*, 1733, 88–95.

Conn, A. R., N. I. M. Gould, and P. L. Toint (1991). A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds. *SIAM Journal on Numerical Analysis*, 28(2), 545–572.

Conn, A. R., N. I. M. Gould, and P. L. Toint (1997). A Globally Convergent Augmented Lagrangian Barrier Algorithm for Optimization with General Inequality Constraints and Simple Bounds. *Mathematics of Computation*, 66(217), 261–288.

Cooke, R. M. and J. M. van Noortwijk (2000). Graphical methods. In *Sensitivity Analysis*, A. Saltelli, K. Chan, and E. M. Scott editors. John Wiley & Sons, 245–264.

Cukier, R. I., C. M. Fortuin, K. E. Schuler, A. G. Petsheck, and J. H. Schailbly (1973). Study Of The Sensitivity Of Coupled Reaction Systems To Uncertainties In Rate Coefficients. *Journal of Chemical Physics*, 26, 1–42.

Cukier, R. I., H. B. Levine, and K. E. Shuler (1978) Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26, 1–42.

de Rocquigny, E., N. Devictor and S. Tarantola (2008). *Uncertainty in Industrial Practice. A Guide to Quantitative Uncertainty Management*. John Wiley & Sons, Ltd.

DMRB – Design Manual for Roads and Bridges (2013). Transport Appraisal Manual, Department for Transport.

DRACULA (2007). *DRACULA User Manual v. 2.4*. Institute for Transport Studies, University of Leeds.

Drud, A. S. (1994). CONOPT – A large-Scale GRG Code. *ORSA Journal of Computing*, 6(2), 207–216.

Edie, L. C. (1974). *Flow theories*. In *Traffic Science*. D. C. Gazis editor, Wiley & Sons.

Efron, B., and C. Stein (1981). The jackknife estimate of variance. *Annals of Statistics*, 9, 586–596.

Fan, J. and I. Gijbels, (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall.

FHWA (2004). *Traffic Analysis Toolbox Volumes III: Guidelines for Applying Traffic Microsimulation Modeling Software*. R. Dowling, A. Skabardonis, and V. Alexiadis editors, Federal Highway Administration.

Frey, H. C., and S. R. Patil (2002). Identification and Review of Sensitivity Analysis Methods. *Risk Analysis*, 22(3), 553–578.

Fu, L., Q. Liu (2003). Real-Time Optimization Model For Dynamic Scheduling Of Transit Operations. *Transportation Research Record*, 1857, 48–55.

Gazis, D. C., R. Herman, and R. W. Rothery (1961). Non-Linear Follow -the-Leader Models of Traffic Flow. *Operations Research*, 9, 545–567.

Ge, Q., and M. Menendez (2012). Sensitivity Analysis for Calibrating VISSIM in Modeling the Zurich Network. *Presented at 12th Swiss Transport Research Conference,* Ascona.

Ghanem, R. G. and P. D. Spanos, (1991). *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag.

Gipps, P. G. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B*, 15(2), 105–111.

Gipps, P. G. (1986). A Model for the Structure of Lane-Changing Decisions. *Transportation Research Part B*, 20, 403-414.

Glover, F. (1998). A Template for Scatter Search and Path Relinking. *Proceeding of AE '97: Selected Papers from the Third European Conference on Artificial Evolution*, Springer-Verlag.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley.

Granger Morgan, M. and M. Henrion. (1990). *Uncertainty – A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press.

HA (2007). *Guidelines for the Use of Microsimulation Software*. Highways Agency.

Halton, J. (1964). Algorithm 247: Radical-inverse quasi-random point sequence. *Magazine Communications of the Association for Computing Machinery*, 7(12), 701–702.

Hamdar, S. H., and H. S. Mahmassani (2008). Driver car following behavior: from a discrete event process to a continuous set of episodes. *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. Springer.

HCM (2000). *Highway Capacity Manual*. Transportation Research Board, National Research Council. Washington, D. C.

Helbing, D., and B. Tilch (1998). Generalized force model of traffic dynamics. *Physical Review E*, 58(1), 133–138.

Helbing, D. (2001). Traffic and related self-driven many-particles systems. *Reviews of modern physics*, 73, 1067–1139.

Hellinga, B. (2001). Automated Vehicle Identification Tag-Matching Algorithms For Estimating Vehicle Travel Times: Comparative Assessment. *Transportation Research Record*, 1774, 106–114.

Høeg, P. (1995). *Borderliners*. Seal Books Publisher.

Homma, T., and A. Saltelli (1996). Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety*, 52(1), 1–17.

Hoogendoorn, S. P., and S. Ossen (2005). Parameter Estimation and Analysis of Car-Following Models. *Proceedings of the 16th International Symposium on Traffic and Transportation Theory*, University of Maryland.

Hoogendoorn, S. P., S. Ossen, and M. Schreuder (2006). Empirics of Multianticipative Car-Following Behavior. *Transportation Research Record*, 1965, 112–120.

Hoogendoorn, S. P. and R. Hoogendoorn (2010a). Generic Calibration Framework for Joint Estimation of Car-Following Models by Using Microscopic Data. *Transportation Research Record*, 2188, 37–45.

Hoogendoorn, S. P. and R. Hoogendoorn (2010b). Calibration of microscopic traffic-flow models using multiple data sources. *Philosophical Transactions of the Royal Society A*, 368, 4497–4517.

Hornberger, G. and R. Spear (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, 7, 7–18.

Ishigami, T., and T. Homma (1996). An importance qualification technique in uncertainty analysis for computer models. *Proceedings of the Isuma '90 First International Symposium on Uncertainty Modelling and Analysis*, University of Maryland.

Isukapalli, S. S. (1999). *Uncertainty Analysis of Transport–Transformation Models. PhD Thesis*. State University of New Jersey.

Jacques, J., C. Lavergne, and N. Devictor (2006). Analysis in presence of model uncertainty and correlate inputs. *Reliability Engineering and System Safety*, 91, 1126–1134.

Jansen, M. J. W., W. A. H. Rossing, and R. A. Daamen (1994). *Monte Carlo estimation of uncertainty contributions from several independent multivariate sources*. In *Predictability and Nonlinear Modelling in Natural Sciences and Economics*, J. Gasmanand, G. van Straten editors, Kluwer Academic Publishers, 334–343.

Jansen, M. J. W. (1999). Analysis Of Variance Designs For Model Output. *Computer Physics Communications*, 117, 35–43.

Ji, X., and P. D. Prevedouros (2005a). Comparison of methods for sensitivity and uncertainty analysis of signalized intersections analyzed with the HCM. *Transportation Research Record*, 1920, 56–64.

Ji, X., and P. D. Prevedouros (2005b). Effects of parameter distributions and correlations on uncertainty analysis of HCM delay model. *Transportation Research Record*, 1920, 118–124.

Ji, X., and P. D. Prevedouros (2006). Probabilistic analysis of Highway Capacity Manual delay for signalized intersections. *Transportation Research Record*, 1988, 67–75.

Ji, X., and P. D. Prevedouros (2007). Probabilistic Analysis of HCM Models: Case Study of Signalized Intersections. *Transportation Research Record*, 2027, 58–64.

JSTE (2011). *Standard Verification Process for Traffic Flow Simulation Model*. Traffic Simulation Committee, Japan Society of Traffic Engineers.

Kamarianakis, Y., A. Kanas, and P. Prastacos (2005). Modeling traffic volatility dynamics in an urban network. *Transportation Research Record*, 1923, 18–27.

Kerner, B. S., S. L. Klenov, and A. Hiller (2007). Empirical test of a microscopic three-phase traffic theory. *Nonlinear Dynamics*, 49(4), 525–553.

Kesting, A., M. Treiber, and D. Helbing (2007). General Lane-Changing Model MOBIL for Car-Following Models. *Transportation Research Record*, 1999, 86–94.

Kesting, A., and M. Treiber (2008). Calibrating Car-Following Models by Using Trajectory Data. Methodological Study. *Transportation Research Record*, 2088, 148–156.

Kim, J-W., and H. S. Mahmassani (2011). Correlated parameters in driving behavior models: car-following example and implications for traffic microsimulation. *Transportation Research Record*, 2249, 62–77.

Kim, S-J., W. Kim and L. R. Rilett (2005). Calibration of Microsimulation Models Using Nonparametric Statistical Techniques. *Transportation Research Record*, 1935, 111–119.

Koutsopoulos, H. N., and H. Farah (2012). Latent class model for car-following behavior. *Transportation Research Part B*, 46, 563–578.

Kucherenko, S. S. Tarantola and P. Annoni (2012). Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications,* 183, 937–946.

Kuipers, L., and H. Niederreiter (2005). *Uniform distribution of sequences*. Dover Publications.

Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization*, 9(1), 112–147.

Lam, W. H. K. and J. Zhou (2000). Optimal Fare Structure for Transit Networks with Elastic Demand. *Transportation Research Record*, 1733, 8–14.

Laval, J. A., and L. Leclercq (2008). Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. *Transportation Research Part B*, 42, 511–522.

Laval, J. A., and L. Leclercq (2010). A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Philosophical Transaction of the Royal Society A*, 368, 4519–4541.

Law, A. M., and W. D. Kelton (2000). *Simulation Modeling and Analysis*, 3rd edition. McGraw-Hill.

Lawe, S., Lobb, J., Sadek, A. W., Huang, S., Xie, C. (2009). TRANSIMS Implementation in Chittenden County, Vermont. Development, Calibration, and Preliminary Sensitivity Analysis. *Transportation Research Record*, 2132, 113–121.

Lenz, H., C. K. Wagner, and R. Sollacher (1999). Multi-Anticipative Car-Following Model. *The European Physical Journal B*, 7, 331–335.

Leurent (1998). Sensitivity and error analysis of the dual criteria traffic assignment model. *Transportation Research Part B*, 32(3), 189–204.

Leutzbach, W., and R. Wiedemann (1986). Development and Applications of Traffic Simulation Models at the Karisruhe Institut für Verkehrswesen. *Traffic Engineering and Control*, 27(5), 270–278.

Li, Z., H. Liu, and K. Zhang (2009). Sensitivity analysis of Paramics based on 2K-P fractional factorial design. *Proceedings of the 2nd International Conference on Transportaton Engineering*, 3633–3638.

LINDO (2003). *LINDO API User Manual 2.0*. LINDO Systems, Inc., Chicago.

Lownes, N. E., and R. B. Machemehl (2006). Sensitivity of Simulated Capacity to Modification of VISSIM Driver Behavior Parameters. *Transportation Research Record*, 1988, 102–110.

Ma, J., H. Dong and H. M. Zhang (2007). Calibration of Microsimulation with Heuristic Optimization Methods. *Presented at 86th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Ma, T. and B. Abdlulhai, 2002. Genetic Algorithm-Based Optimization Approach and Generic Tool for Calibrating Traffic Microscopic Simulation Parameters. *Transportation Research Record*, 1800, 6–15.

Mara, T. A. and S. Tarantola (2011). Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering and System Safety*, 107, 115–121.

Marczak, F. M., and C. Buisson (2012). New filtering method for trajectories measurement errors and its comparison with existing methods. *Transportation Research Record*, 2315, 35–46.

Marzano V., A. Papola, and F. Simonelli (2009). Limits and perspectives of effective OD matrix correction using traffic counts. *Transportation Research Part C*, 17, 120–132.

Mathew, T. V., and P. Radhakrishnan (2010). Calibration of microsimulation models for nonlane-based heterogeneous traffic at signalized intersections. *Journal of Urban Planning and Development*, 136(1), 59–66.

MATLAB (2009). *Matlab R2009b*. Mathworks Inc.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*, 2nd edition. Chapman & Hall.

McKay, M. D. (1997) Nonparametric variance-based methods of assessing uncertainty importance. *Reliability Engineering & System Safety*, 57(3), 267–279.

McKinnon, K. I. M. (1999). Convergence of the Nelder–Mead simplex method to a non-stationary point. *SIAM Journal of Optimization*, 9, 148–158.

Melkote, S., and M. S. Daskin (2001). An integrated model of facility location and transportation network design. *Transportation Research Part A*, 35, 515–538.

Mesbah, M., M. Sarvi, G. Currie, and M. Saffarzadeh (2010). Policy-Making Tool for Optimization of Transit Priority Lanes in Urban Network. *Transportation Research Record*, 2197, 54–62.

Montanino, M., B. Ciuffo, and V. Punzo (2012). Calibration of microscopic traffic flow models against time-series data. *IEEE Proceedings of the 15th International Conference on Intelligent Transportation Systems*, Anchorage, 108–114.

Montanino, M., and V. Punzo (2013). Making NGSIM data usable for studies on traffic flow theory: a multistep method for vehicle trajectory reconstruction. *Transportation Research Record*, in printing.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974). *Introduction to the Theory of Statistics*, 3rd edition. McGraw-Hill.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33, 161–174.

MULTITUDE (2014). www.multitude-project.eu. COST Action TU0903 - Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation moDEls. Last accessed March 31st, 2014.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *Computer Journal*, 7, 308–313.

Newell, G. F. (2002). A simplified car-following theory: a lower order model. *Transportation Research Part B*, 36, 195–205.

NGSIM (2005). *NGSIM I-80 Data Analysis (4.00 p.m. to 4.15 p.m.)*. Federal Highway Administration. Cambridge Systematics, Inc.

NGSIM (2014). *NGSIM – Next Generation SIMulation*. US Department of Transportation – FHWA. http://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm. Last accessed on March 31st, 2014.

Ni, D., J. D. Leonard II, A. Guin, and B. M. Williams (2004). A Systematic Approach for Validating Traffic Simulation Models. *Transportation Research Record*, 1876, 20–31.

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics.

Oreskes, N., K. Shrader-Frechette and K. Belitz (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263, 641–646.

Ossen, S., and S. P. Hoogendoorn (2005). Car-following behavior analysis from microscopic trajectory data. *Transportation Research Record*, 1934, 13–21.

Ossen, S. and S. P. Hoogendoorn (2007). Driver Heterogeneity in Car Following and Its Impact on Modeling Traffic Dynamics. *Transportation Research Record*, 1999, 95–103.

Ossen, S., and S. P. Hoogendoorn (2008a). Validity of trajectory-based calibration approach of car-following models in presence of measurement errors. *Transportation Research Record*, 2088, 117–125.

Ossen, S., and S. P. Hoogendoorn (2008b). Calibrating car-following models using microscopic trajectory data. A critical analysis of both microscopic trajectory data collection methods, and calibration studies based on these data. *Report TU Delft*.

Ossen, S. and S. P. Hoogendoorn (2009). Reliability of Parameter Values Estimated Using Trajectory Observations. *Transportation Research Record*, 2124, 36–44.

Ossen S. and S. P. Hoogendoorn (2011). Heterogeneity in car-following behavior: Theory and empirics. *Transportation Research Part C*, 19, 182–195.

Ossen, S., S. P. Hoogendoorn, and B. G. H. Gorte (2006). Interdriver Differences in Car-Following. A Vehicle Trajectory-Based Study. *Transportation Research Record*, 1965, 121–129.

Papageorgiou, M. (1998). Some Remarks On Macroscopic Traffic Flow Modelling. *Transportation Research Part A*, 32(5), 323-329.

PARAMICS (2003). *Quadstone Paramics V4.2 Analyser Reference Manual.* Quadstone Paramics, Edinburgh.

Patè-Cornell, M. E. (1996). Uncertainties in risk analysis: six levels of treatment. *Reliability Engineering & System Safety*, 54(2–3), 95–111.

Patel, I., A. Kumar, and G. Manne (2003). Sensitivity Analysis of CAL3QHC Roadway Intersection Model. *Transportation Research Record*, 1842, 109–117.

Pel, A. J., S. P. Hoogendoorn, and M. C. J. Bliemer (2010). Impact of Variations in Travel Demand and Network Supply Factors for Evacuation Studies. *Transportation Research Record*, 2196, 45–55.

Peng, Z-R. and E. Beimborn (2001). A breakeven analysis for statewide its project identification and assessment. *Transportation Research Record*, 1777, 105–115.

Powell, M. J. D., 1973. On Search Directions for Minimization Algorithms. *Mathematical Programming*, 4, 193–201.

Punzo, V. (2014). Future Directions for Managing Uncertainty in Stochastic Traffic Models. *Presented at the 93rd TRB Annual Meeting*, Washington, D. C.

Punzo, V., and F. Simonelli (2005). Analysis and comparison of microscopic traffic flow models with real traffic microscopic data. *Transportation Research Record*, 1934, 53–63.

Punzo, V., D. J. Formisano, and V. Torrieri (2005). Non-Stationary Kalman Filter for the Estimation of Accurate and Consistent Car-Following Data. *Transportation Research Record*, 1934, 3–12.

Punzo, V., and A. Tripodi (2007) Steady-State Solutions and Multiclass Calibration of Gipps Microscopic Traffic Flow Model. *Transportation Research Record*, 1999, 104–114.

Punzo, V., and B. Ciuffo (2009). How Parameters of Microscopic Traffic Flow Models Relate to Traffic Dynamics in Simulation Implications for Model Calibration. *Transportation Research Record*, 2124, 249–256.

Punzo, V., and B. Ciuffo (2011). Sensitivity Analysis Of Car-Following Models: methodology and application. *Presented at the 90th TRB Annual Meeting*, Washington, D. C.

Punzo, V., B. Ciuffo, and E. Quaglietta (2011a). Kriging meta-modelling to verify traffic micro-simulation calibration methods. *Proceedings of the 90th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Punzo, V., Borzacchiello, M. T., and B. Ciuffo (2011b). On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C*, 19(6), 1243–1262.

Punzo, V., B. Ciuffo, and M. Montanino (2012). Can results of car-following models calibration based on trajectory data be trusted?. *Transportation Research Record*, 2315, 11–24. The TRB Traffic Flow Theory and Characteristics Committee peer-reviewed this paper, which was awarded the 2012 Greenshields Prize by the committee, honoring the contributions of Bruce Greenshields to the field of traffic flow theory.

Punzo, V., M. Montanino and B. Ciuffo (2014a). Do we really need to calibrate all the parameters of microscopic traffic flow models? A variance-based sensitivity analysis framework. *IEEE Transactions on Intelligent Transportation Systems*, in printing.

Punzo, V., B. Ciuffo and M. Montanino (2014b). *Sensitivity Analysis*. In *Traffic Simulation and Data: Validation Methods and Applications*. W. Daamen, C. Buisson and S. P. Hoogendoorn editors, CRC Press Taylor & Francis Group.

Rabitz, H., Ö.F. Alis, J. Shorter, and K. Shim (1999). Efficient input–output model representations. *Computer Physics Communications*, 117, 11–20.

Ranjitkar, P., T. Nakatsuji, and M. Asano (2004). Performance Evaluation of Microscopic Flow Models with Test Track Data. *Transportation Research Record*, 1876, 90–100.

Rodier, C. J., and R. A. Johnston (2005). Uncertain socioeconomic projections used in travel demand and emissions models: could plausible errors result in air quality nonconformity?. *Transportation Research Part A*, 36, 613–631.

Rosen, R. (1991). *Life Itself – a Comprehensive Inquiry into Nature, Origin, and Fabrication of Life*. Columbia University Press.

Sacks, J., W. Welch, T. Mitchell, and H. Wynn (1989). Design and analysis of computer experiments. *Statistical Science*, 4, 409–435.

Sadek, A. W., and B. Baah (2003). Cost-Effectiveness Of Intelligent Transportation System (Its) Deployment In A Medium-Sized Area: Case Study With Its Deployment Analysis System. *Transportation Research Record*, 1826, 16–24.

Saltelli, A. (2002). Making best use of model valuations to compute sensitivity indices. *Computer Physics Communications*, 145, 280–297.

Saltelli, A., and R. Bolado, (1998). An Alternative Way to Compute Fourier Amplitude Sensitivity Test. *Computational Statistics & Data Analysis*, 26, 445–460.

Saltelli A., S. Tarantola and K. P. S. Chan (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41, 39–56.

Saltelli, A. and S. Tarantola (2002). On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of the American Statistical Association*, 97, 702–709.

Saltelli, A., S. Tarantola, F. Campolongo and M. Ratto (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd.

Saltelli, A., M. Ratto, S. Tarantola and F. Campolongo (2005). Sensitivity analysis for chemical models. *Chemical Reviews* 105, 2811–2828.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety*, 91 (10-11), 1109–1125.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Gariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). *Global Sensitivity Analysis. The Primer*. John Wiley & Sons.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181, 259–270.

Santner, T., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer.

Schultz, G. G. and L.R. Rilett (2004). Analysis of distribution and calibration of car-following sensitivity parameters in microscopic traffic simulation models. *Transportation Research Record*, 1876, 41–51.

Schuster, T. D., J. Byrne, J. Corbett, Y. Schreuder (2005). Assessing the Potential Extent of Carsharing. A New Method and Its Implications. *Transportation Research Record*, 1927, 174–181.

Shannon, C. (1949). Communication in the Presence of Noise. *Proceeding of the Institute of Radio Engineers*, 37(1), 10–21.

Sinha, K. C., and S. Labi, (2007). *Transportation Decision Making. Principles of Project Evaluation and Programming*. Wiley & Sons, Ltd.

Smith, S., and L. Lasdon (1992). Solving Large Sparse Nonlinear Programs Using GRG. *ORSA Journal on Computing*, 4(1), 2–15.

Smola, A. J., and B. Scholkopf (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199–222.

Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7, 86–112.

Sobol, I. M. and Y. L. Levitan (1976). The production of points uniformly distributed in a multidimensional cube. *Tech. Rep. 40*, Institute of Applied Mathematics, USSR Academy of Sciences.

Sobol, I. M., Y. L. Levitan, B. V. Shukman, and V. I. Turchaninov (1992). Quasirandom sequence generators. *IPM Zak.*, 30.

Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1, 407–414.

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55, 271–280.

Sobol', I. M. (2007). Global sensitivity analysis indices for the investigation of nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 19(11), 23–24.

Sobol', I. M., S. Tarantola, D. Gatelli, S. Kucherenko, and W. Mauntz (2007). Estimating The Approximation Error When Fixing Unessential Factors In Global Sensitivity Analysis. *Reliability Engineering and System Safety*, 92, 957–960.

Song, G., L. Yu, and Y. Zhang (2012). Applicability of traffic micro-simulation models in vehicle emission estimations: a case study of VISSIM. *Transportation Research Record*, 2270, 132–141.

Spear, R., T. Grieb and N. Shang (1994). Factor uncertainty and interaction in complex environmental models. *Water Resources Research*, 30, 3159–3169.

Spyropoulou, I. (2007). Simulation using Gipps' car-following model. An in-depth analysis. *Transportmetrica*, 3(3), 231–245.

Sun, G., G. Arr, and P. Ramachandran (2007). Vehicle Reidentification as Method for Deriving Travel Time and Travel Time Distributions: Investigation. *Transportation Research Record*, 1826, 25–30.

TAC (2008). *Best Practices for the Technical Delivery of Long-Term Planning Studies in Canada. Final Report*. Transportation Association of Canada.

Tam, M-L. and W. H. K. Lam (1999). Analysis of Demand for Road-Based Transport Facilities A Bi-Level Programming Approach. *Transportation Research Record*, 1685, 73–80.

TfL (2010). *Traffic Modelling Guidelines. TfL Traffic Manager and Network Performance Best Practice*, version 3.0. J. Smith and R. Blewitt editors, Transport for London.

Thiemann, C., M. Treiber, and A. Kesting. Estimating acceleration and lane-changing dynamics based on NGSIM trajectory data. *Transportation Research Record*, 2088, 90–101.

Tobin, R., and T. Friesz (1988). Sensitivity analysis for equilibrium network flow. *Transportation Science*, 22, 242–250.

Toledo, T. (2003). Integrated driving behavior modeling. *PhD. Thesis*, Massachusetts Institute of Technology.

Toledo, T., H. Koutsopoulos, and M. Ben-Akiva (2009). Estimation of an integrated driving behavior. *Transportation Research Part C*, 17, 365–380.

Treiber, M., A. Hennecke, and D. Helbing (2000). Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, 62(2), 1805–1824.

Treiber, M., A. Kesting, and D. Helbing (2006). Delays, Inaccuracies And Anticipation In Microscopic Traffic Models. *Physica A*, 360, 71–88.

Treiber, M., and A. Kesting (2013). *Traffic Flow Dynamics. Data, Models and Simulation*. Springer.

Ugray, Z., L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martì (2005). A Multistart Scatter Search Heuristic for Smooth NLP and MINLP Problems. *Operations Research/Computer Science Interfaces Series*, 30(1), 25–57.

van Hinsbergen, C. I. J., H. W. C. Van Lint, S. P. Hoogendoorn, and H. Van Zuylen (2009). Bayesian calibration of car-following models. *12th IFAC Symposium on Transportation Systems*, 91–97.

Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley & Sons, Ltd.

VISSIM (2011). *VISSIM 5.40 User Manual*. PTV, Karlsruhe.

VTRC (2006). *Microscopic Simulation Model Calibration And Validation Handbook*. B. Park, J. Won editors. Virginia Transportation Research Council.

Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.

Wan, B., and N. M. Rouphail (2004). Using Arena for Simulation of Pedestrian Crossing in Roundabout Areas. *Transportation Research Record*, 1878, 58–65.

Wang, H., W. Wang, J. Chen, and M. Jing (2010). Car-Following Model Calibration and Analysis of Intra-Driver Heterogeneity. *Advanced Materials Research*, 108(111), 805–810.

Ward, J. A., and R. E. Wilson (2011). Criteria for convective versus absolute string instability in car-following models. *Philosophical Transactions of the Royal Society A*, 467, 2185-2208.

Wegman, E. J. (1990). Hyperdimensional Data Analysis Using Parallel Coordinates. *Journal of the American Statistical Association*, 85 (411), 664–675.

Willumsen, L. G., and J. d. D. Ortuzar (2011). *Modelling Transport*, 4th edition. Wiley & Sons, Ltd.

Wilson, R. E. (2001). An analysis of Gipps's car-following model of highway traffic. *IMA Journal on Applied Mathematics*, 66 (5), 509–537.

Wilson, R. E. (2008). Mechanisms for spatio-temporal pattern formation in highway traffic models. *Philosophical Transactions of the Royal Society A*, 366, 2017–2032.

Yan, H., and W. H. K. Lam (1996). Optimal road tolls under conditions of queueing and congestion. *Transportation Research Part A*, 30(5), 319–332.

Yang, H. (1998). Sensitivity Analysis for Queuing Equilibrium Network Flow and Its Application to Traffic Control. *Mathematical Computing Modeling*, 22 (4-7), 247–258.

Yang, Q., and H. N. Koutsopoulos (1996). A microscopic traffic simulation for evaluation of dynamic traffic management systems. *Transportation Research Part C*, 4, 113–129.

Yang, Q., H. N. Koutsopoulos, and M. Ben-Akiva (2000). A simulation laboratory for evaluating dynamic traffic management systems. *Transportation Research Record,* 1710, 122–130.

Young, P. C., R. C. Spear and G. M. Hornberger (1978). Modeling badly defined systems: some further thoughts. In *Proceedings SIMSIG Conference*, Canberra, 24–32.

Young, P. C. and D. J. Pedregal (1999). Recursive and en-bloc approaches to signal extraction. *Journal of Applied Statistics*, 26, 103–128.

Appendix A

Global Sensitivity Analysis Techniques Based on Sobol' Decomposition of Variance

A.1 Introduction

Variance-based methods have assessed themselves as versatile and effective among the various available techniques for sensitivity analysis of model output.

Variance based methods have a long history in sensitivity analysis. They start with a Fourier implementation in the seventies (Cukier et al., 1973), and have a milestone in the work of Sobol' (1993). The total sensitivity indices have been introduced by Homma and Saltelli (1996), although the concept was proposed in Jansen et al. (1994). For reviews, see Helton et al. (2006), Saltelli et al. (2008).

The Appendix is organized as follows. Section A.2 presents the mathematical formulation based on the Sobol' decomposition of variance. Possible settings for numerical calculation of the sensitivity indices are provided in Section A.3, while some remarks on the application of these techniques are reported in Section A.4.

A.2 Mathematical Formulation

Given a model of the form $Y = f(X_1, X_2, \ldots X_k)$, with Y a scalar, a variance based first-order effect for a generic factor X_i can be written as (see notations in Table A.1):

Table A.1: Notations.

N	Sample size
k	Number of factors
X_i	Generic factor
$x_{j,i}$	Generic value for factor X_i taken from row j of X_i
Y	Generic scalar model output equal to $Y = f(X_1, X_2, \ldots, X_k)$
$\mathbf{X}$	$N \times k$ matrix of input factors
$\mathbf{A}, \mathbf{B}$	$N \times k$ sample matrices of input factors
$\mathbf{X}_{\sim i}$	$N \times (k-1)$ matrix of all factors but X_i
$\mathbf{A}^{(i)}_{\mathbf{B}}$	Matrix, where column i comes from matrix $\mathbf{B}$ and all other $k-1$ columns come from matrix $\mathbf{A}$
$\mathbf{B}^{(i)}_{\mathbf{A}}$	Matrix, where column i comes from matrix $\mathbf{A}$ and all other $k-1$ columns come from matrix $\mathbf{B}$
N^T	Total cost of a sensitivity analysis in terms of model evaluations
$V_{Xi}(\cdot)$, $E_{Xi}(\cdot)$	Variance or mean of argument ($\cdot$) taken over X_i
$V_{\mathbf{X}\sim i}(\cdot)$, $E_{\mathbf{X}\sim i}(\cdot)$	Variance or mean of argument ($\cdot$) taken over all factors but Xi

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right) \tag{A.1}$$

where X_i is the i-th factor and $\mathbf{X}_{\sim i}$ denotes the matrix of all factors but X_i. The meaning of the inner expectation operator is that the mean of Y is taken over all possible values of $\mathbf{X}_{\sim i}$ while keeping X_i fixed. The outer variance is taken over all possible values of X_i. The associated sensitivity measure (first order sensitivity coefficient) is written as:

$$S_i = \frac{V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right)}{V(Y)} \tag{A.2}$$

Due to the known identity (Mood et al., 1974):

$$V(Y) = V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right) + E_{X_i}\left(V_{\mathbf{X}_{\sim i}}(Y / X_i)\right) \tag{A.3}$$

S_i is a normalized index, as $V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right)$ varies between zero and $V(Y)$. $V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right)$ measures the first-order (e.g. additive) effect of X_i on the model output, while $E_{X_i}\left(V_{\mathbf{X}_{\sim i}}(Y / X_i)\right)$ is customarily called the residual.

Another popular variance-based measure is the total effect index (Homma and Saltelli, 1996; Saltelli and Tarantola, 2002):

$$ST_i = \frac{E_{\mathbf{X}_{\sim i}}\left(V_{X_i}\left(Y / \mathbf{X}_{\sim i}\right)\right)}{V(Y)} = 1 - \frac{V_{\mathbf{X}_{\sim i}}\left(E_{X_i}\left(Y / \mathbf{X}_{\sim i}\right)\right)}{V(Y)} \qquad (A.4)$$

ST_i measures the total effect, i.e. first- and higher-order effects (interactions) of factor X_i. One way to visualize this is by considering that $V_{\mathbf{X}_{\sim i}}\left(E_{X_i}\left(Y / \mathbf{X}_{\sim i}\right)\right)$ is the first-order effect of $\mathbf{X}_{\sim i}$, so that $V(Y) - V_{\mathbf{X}_{\sim i}}\left(E_{X_i}\left(Y / \mathbf{X}_{\sim i}\right)\right)$ must give the contribution of all terms in the variance decomposition which do include X_i.

The decomposition equations describing the variance-based framework is given in the following. These apply to a square integrable function $Y = f(X_1, X_2, \ldots X_k)$ defined over Ω, the k-dimensional unit hypercube:

$$\Omega = \left\{X \middle| 0 \le x_i \le 1; x = 1,...,k\right\} \qquad (A.5)$$

We further suppose that the factors are uniformly distributed in [0, 1]. The steps of a variance-based framework are as follows:

- Functional decomposition scheme:

$$z = z_0 + \sum_{i=1}^{k} z_i + \sum_{i=1}^{k}\sum_{j>i}^{k} z_{i,j} + ... + z_{1,2,...,k} \qquad (A.6)$$

where $z_i = z_i(X_i)$, $z_{i,j} = z_{i,j}(X_i, X_j)$ and so on for a total of $2k$ terms, including z_0. Each term is square integrable over Ω. Eq. (A.6) is known as Hoeffding decomposition. See Archer et al. (1997), Rabitz et al. (1997) for reviews, and Efron and Stein (1981), Sacks et al. (1989), Sobol' (1993) for useful reading. The unicity condition for Eq. (A.6) is granted by Sobol' (1993):

$$\int_0^1 z_{i_1, i_2, ..., i_S}\left(x_{i_1}, x_{i_2}, ..., x_{i_S}\right) dx_{i_W} = 0 \qquad (A.7)$$

where $1 \le i_1 < i_2 < \ldots < i_S \le k$ and $i_W \in \{i_1, i_2, \ldots, i_s\}$. The functions $z_{i1,i2,...,is}$ are obtained from:

$$z_0 = E(Y), \; z_i = E_{\mathbf{X}_{\sim i}}(Y \mid X_i) - E(Y), \; z_{i,j} = E_{\mathbf{X}_{\sim ij}}(Y \mid X_i, X_j) - z_i - z_j - E(Y) \qquad (A.8)$$

and similarly for higher-orders.

- Relation between functions $z_{i1,i2,...,is}$ and partial variances in Eq. A.9 (Sobol', 1993):

$$V_i = V[z_i(X_i)] = V_{X_i}[E_{\mathbf{X}_{\sim i}}(Y|X_i)]$$

$$V_{i,j} = V[z_{i,j}(X_i, X_j)] = V_{X_i, X_j}[E_{\mathbf{X}_{\sim i,j}}(Y|X_i, X_j)] - V_{X_i}[E_{\mathbf{X}_{\sim i}}(Y|X_i)] - V_{X_j}[E_{\mathbf{X}_{\sim j}}(Y|X_j)]$$

and so on for higher-order terms. All terms are linked by:

$$V(Y) = \sum_{i=1}^{k} V_i + \sum_{i=1}^{k}\sum_{j>i}^{k} V_{i,j} + ... + V_{1,2,...,k} \tag{A.10}$$

Dividing both sides of the equation by $V(Y)$, we obtain:

$$\sum_{i=1}^{k} S_i + \sum_{i=1}^{k}\sum_{j>i}^{k} S_{i,j} + ... + S_{1,2,...,k} = 1 \tag{A.11}$$

Relations for the second and higher-order terms in Eq. (A.9) as well as formula Eq. (A.10) hold if the factors are independent, which is the setting adopted throughout the present work[1].

Note that given the assumption of independence of input factors we may avoid to explicitly include the probability distribution function p_i of factor X_i in the integrals for the estimates of functions in Eq. (A.8). This implies that notation $\int z_i(x_i) dx_i$ can be used in place of the more verbose $\int z_i(x_i) p(x_i) dx_i$ as the factors *pdf* may be embedded in the function $z_i(x_i)$. Without loss of generality all factors can be conceived as defined in Ω and the mapping from Ω to the actual distribution of X_i is intended to be part of the definition of z.

---

[1] Independency assumption implies that model parameters are uncorrelated. This is a very strong hypothesis, as parameters of microscopic traffic flow simulation models were shown to show a certain degree of correlation among each other (for details, please refer to Chapter 4). However, an enhanced methodological framework to deal with dependent or correlated input factors is currently under investigation by the field research (see Jacques et al., 2006; Mara and Tarantola, 2011; Kucherenko et al., 2012). Notwithstanding, we applied such established framework throughout the study, being conscious of its limitation.

Indices S_i, ST_i can also be interpreted in terms of expected reduction of variance. This interpretation also holds when the input factors are not independent (Saltelli and Tarantola, 2002):

- $V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right)$ is the expected reduction in variance that would be obtained if X_i could be fixed.

- $E_{\mathbf{X}_{\sim i}}\left(V_{X_i}(Y / \mathbf{X}_{\sim i})\right)$ is the expected variance that would be left if all factors but X_i could be fixed. This holds since $V_{\mathbf{X}_{\sim i}}\left(E_{X_i}(Y / \mathbf{X}_{\sim i})\right)$ is the expected reduction in variance that would be obtained if all factors but X_i could be fixed.

For this reason, Jansen (1999) calls $V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right)$ and $E_{\mathbf{X}_{\sim i}}\left(V_{X_i}(Y / \mathbf{X}_{\sim i})\right)$ top and bottom marginal variances, respectively. For additive models the two terms coincide, as ST_i may be simply viewed as S_i plus all (null) interaction terms including factor X_i.

A.3 Numerical Calculation Settings

We discuss here existing estimators to compute in a single set of simulations both sets of indices S_i and ST_i. A more exhaustive review on topic can be found in Saltelli el al. (2008).

By "simulation" we mean here the computation of an individual value for Y corresponding to a sampled set of k factors $X_1, X_2, \ldots, X_k$.

We imagine to have two independent sampling matrices $\mathbf{A}$ and $\mathbf{B}$, with $a_{j,i}$ and $b_{j,i}$ as generic elements. The index i runs from 1 to k, the number of factors, while the index j runs from 1 to N, the number of simulations. We now introduce matrix $\mathbf{A}^{(i)}{}_{\mathbf{B}}$ ($\mathbf{B}^{(i)}{}_{\mathbf{A}}$) where all columns are from $\mathbf{A}$ ($\mathbf{B}$) except the i-th column which is from $\mathbf{B}$ ($\mathbf{A}$). S_i can be computed from either the couple of matrices $\mathbf{A}$, $\mathbf{B}^{(i)}{}_{\mathbf{A}}$ or $\mathbf{B}$, $\mathbf{A}^{(i)}{}_{\mathbf{B}}$, e.g.:

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y / X_i)\right) = \frac{1}{N} \sum_{j=1}^{N} \left[f(\mathbf{A})_j \cdot f\left(\mathbf{B}_{\mathbf{A}}^{(i)}\right)_j \right] - f_0^2 \qquad (A.12)$$

where $(\mathbf{B})_j$ denotes the j-th row of matrix $\mathbf{B}$ (Sobol', 1993).

The computation of ST_i proceeds from definition Eq. (A.4), where $V_{\mathbf{X}_{\sim i}}\left(E_{X_i}(Y / \mathbf{X}_{\sim i})\right)$ is obtained from Homma and Saltelli (1996):

$$V_{\mathbf{X}\sim i}\left(E_{\mathbf{X}\sim i}\left(Y / \mathbf{X}_{\sim i}\right)\right) = \frac{1}{N} \sum_{j=1}^{N} \left[f(\mathbf{A})_j \cdot f\left(\mathbf{A}_{\mathbf{B}}^{(i)}\right)_j\right] - f_0^2 \tag{A.13}$$

Eqs. (A.12) and (A.13) were derived following the approach outlined in Ishigami and Homma (1996) and Saltelli (2002). A review is found in Saltelli et al. (2010).

According to Eqs. (A.12) and (A.13) all that is needed to compute both sets of S_i and ST_i for the k factors is the triplet of matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{B}^{(i)}_{\mathbf{A}}$, or alternatively (equivalently) $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}^{(i)}_{\mathbf{B}}$. As shown in Saltelli et al. (2010), the latter setting is proved to perform better when the adopted sampling scheme to build the matrices A, B is based on the Sobol' quasi-random sequences (Sobol', 1992). For the above reason, this setting was adopted in the present work.

2N simulations are needed for computing *Y* corresponding to matrices $\mathbf{A}$, $\mathbf{B}$ while *kN* simulations are needed to compute *Y* from matrices $\mathbf{A}^{(i)}_{\mathbf{B}}$ for all factors. As a result the cost of the analysis is $N \cdot (k + 2)$ with *N* a sufficiently large number to guarantee numerical stability for calculated indices.

As shown in Saltelli et al. (2010), the estimator for S_i has been improved by Saltelli (2002) and Sobol' et al. (2007), who proposed:

$$V_{X_i}\left(E_{\mathbf{X}\sim i}\left(Y / X_i\right)\right) = V(Y) - \frac{1}{N} \sum_{j=1}^{N} f(\mathbf{A})_j \cdot \left[f(\mathbf{A})_j - f\left(\mathbf{A}_{\mathbf{B}}^{(i)}\right)_j\right] \tag{A.14}$$

Finally, in Saltelli et al. (2010), the authors suggested a further improvement which uses the triplet $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}^{(i)}_{\mathbf{B}}$, instead of the original formulation which uses $\mathbf{A}$, $\mathbf{B}$, $\mathbf{B}^{(i)}_{\mathbf{A}}$, as for the performance benefit related to the use of Sobol' quasi-random sequences.

A numerical improvement of estimator for ST_i has been proposed in Sobol' (2007):

$$V_{\mathbf{X}\sim i}\left(E_{\mathbf{X}\sim i}\left(Y / \mathbf{X}_{\sim i}\right)\right) = \frac{1}{N} \sum_{j=1}^{N} \left[f(\mathbf{A})_j \cdot f\left(\mathbf{A}_{\mathbf{B}}^{(i)}\right)_j\right] - f_0^2 \tag{A.15}$$

An alternative formulation of the estimator for ST_i has been proposed by Jansen (1999), which improves the computational convergence of the calculation. Jansen's formula proceeds via $E_{\mathbf{X}\sim i}\left(V_{X_i}\left(Y / \mathbf{X}_{\sim i}\right)\right)$ rather than via $V_{\mathbf{X}\sim i}\left(E_{X_i}\left(Y / \mathbf{X}_{\sim i}\right)\right)$:

$$E_{\mathbf{X}\sim i}\left(V_{X_i}\left(Y / \mathbf{X}_{\sim i}\right)\right) = \frac{1}{2N} \sum_{j=1}^{N} \left[f(\mathbf{A})_j \cdot f\left(\mathbf{A}_{\mathbf{B}}^{(i)}\right)_j\right]^2 \tag{A.16}$$

Based on the best-practices for the calculation of sensitivity indices reported in Saltelli et al. (2010), in the present study we applied Eq. (A.14) for the calculation of S_i, and Eq. (A.16) for the calculation of ST_i.

For a discussion on the available computation schemes for the design of the triplet $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}^{(i)}_{\mathbf{B}}$, please refer to Saltelli et al. (2010).

A.4 Caveats

Variance-based methods are powerful in quantifying the relative importance of input factors or groups. The main drawback of variance-based methods is the cost of the analysis, which, in the case of computationally intensive models, may become prohibitive even when using the approach described above.

In the framework reviewed in this Appendix, $N \cdot (k + 2)$ model runs for a full set of S_i and ST_i require the adoptions of at least $N = 1000$.

In terms of computational time, thousands or tens of thousands of model executions can be either trivial or unfeasible, depending on the model at hand. A viable alternative for computationally expensive models is the adoption of a screening technique, such as the Elementary Effect test which is a good proxy for the total sensitivity indices.

A double step framework could be set up in case of expensive and high-parameterized models, where the Elementary Effect test can be used to reduce the number of factors, and a successive variance-based analysis can be run only on the reduced set of factors.

Appendix B

Gipps' Car-Following Model: An Enhanced Version

B.1   Introduction

The contents of this Appendix are reported in Ciuffo et al. (2012b).

In 1981, in the vol. 15B of the Transportation Research Part B journal, an article by P.G. Gipps appeared with the title "A behavioural car-following model for computer simulation" (Gipps, 1981). This paper was bound to have a considerable impact in the traffic flow theory and practice, and the model described therein to be widely known as the Gipps' car-following model.

Car-following models try to explicitly reproduce the complex dynamics governing the actions of the driver/vehicle system, while following another vehicle. Dozens of car-following models have been presented hitherto and several new ones are continuously proposed (the reader can refer to various sources presenting a clustered review of the topic, such as in Brackstone and McDonald, 1999; Helbing, 2001; Cao and Yang, 2009), based on different assumptions regarding the driving strategy adopted by a vehicle to adapt its speed to the presence of an immediate vehicle downstream in the same lane.

Car-following models have two main applications: *i)* modeling the "aggregate" traffic propagation and evolution, and *ii)* modeling the microscopic behaviour of the vehicle

during a "follow-the-leader" activity. In the first case, car-following models are usually included within a broader modeling framework of micro-simulation software, as depicted in Chapter 2. In the second case, car-following models are mainly used in the design of on-board devices to assist drivers keeping safe and comfortable driving conditions (e.g. intelligent speed adaptation systems, collision avoidance systems, etc.).

One reason for a high number of car-following models proposed could be motivated by their overall incapability to reproduce both traffic propagation and driving interactions without relying upon the over-fitting produced by their parameters, with some of them usually unnecessary and/or without a clear physical interpretation (this certainly poses serious concerns about their capability to reproduce unpredictable conditions). As a result, most of the applications using car-following models usually adopt the less recent "classical" models. The Gipps' car-following model is one of them.

Reasons for the fascination of the Gipps' model primarily resides in the clear physical context adopted in its derivation: a driver adapts its speed in order to *i)* smoothly reach the desired speed or *ii)* safely proceed behind its leader. In addition, Wilson (2001) demonstrated that, similarly to other "reductionist models" like that of Bando et al. (1995), the Gipps' model may allow a uniform traffic flow to lose stability for certain ranges of its parameters. Stability loss is an important feature as it allows for typical traffic mechanisms to be reproduced (such as flow breakdown and spontaneous traffic jam formation).

However, as noticed in Ranjitkar et al. (2005) and Spyropoulou (2007), some properties of the Gipps' model have been hidden by the positions assumed by Gipps himself and, thus, the scope of the model might even be enlarged.

For all these reasons, in this Appendix we aim to summarize the main features of the Gipps' car-following model as they have been derived in different studies and applications. Furthermore, we presented some analyses on the "acceleration component" of the Gipps' model, providing insights on the effect that the relaxation of three parameters, usually considered as fixed, may have on the model performances.

The Appendix is organized as follows. In Section B.2, the standard Gipps' car-following model is presented. In Section B.3, a literature review of the analyses carried out on the model is provided, together with a description of the main innovative features they were able to introduce. Finally, in Section B.4, we presented our analysis on the acceleration

component of the model, proposing an enhanced model formulations which allows for more realistic representation of the *speed-acceleration* function.

B.2   Original Formulation

The Gipps' car-following model is the most commonly used model pertaining to the class of "safety distance" or "collision avoidance" models. Models of this class aim to specify a safe following distance, and to adapt the driver's behaviour in order to always keep it. The basic idea behind the model is that each driver plans his or her speed for the following instant (i.e. after a delay τ) such that he/she can safely stop even in the event of the leading vehicle suddenly braking. In case the driver has no vehicles in front, instead, speed planned for time (t + τ) is obtained from an inequality equation, derived experimentally, that combines two conditions: *i)* that the speed never exceeds the driver's desired speed, and *ii)* that acceleration decreases with increasing speed untill it becomes null when the desired speed has been reached.

According to the Gipps' model, then, the speed attained by a vehicle at a given time instant (t+τ) (in which the delay τ is the "apparent" driver's reaction time; Gipps, 1981), is given by:

$$v_n(t+\tau) = Max\{0, min\{v_{a,n}(t+\tau), v_{b,n}(t+\tau)\}\} \tag{B.1}$$

with:

$$v_{a,n}(t+\tau) = v_n(t) + 2.5 \cdot a_n^{Max} \cdot \tau \cdot \left(1 - \frac{v_n(t)}{V_n^{Max}}\right) \cdot \sqrt{0.025 + \frac{v_n(t)}{V_n^{Max}}} \tag{B.2}$$

$$v_{b,n}(t+\tau) = -b_n \cdot \left(\frac{\tau}{2}+\theta\right) + \sqrt{b_n^2 \cdot \left(\frac{\tau}{2}+\theta\right)^2 + b_n \cdot \left[2 \cdot \left(x_{n-1}(t) - x_n(t) - S_{n-1}\right) - \tau \cdot v_n(t) + \frac{v_{n-1}(t)^2}{\hat{b}_{n-1}}\right]} \tag{B.3}$$

where:

-       $v_n(t)$ and $v_{n-1}(t)$ are, respectively, the follower's and leader's speed at time *t* [m/s];

-       a_n^{Max} is the follower's maximum acceleration rate [m/s^2];

-       τ is "the apparent reaction time, a constant for all vehicles" (Gipps, 1981) [s];

- V_n^{Max} is the follower's maximum desired speed, that is "the speed at which the driver of vehicle n wishes to travel" (Gipps, 1981) [m/s];

- b_n is "the most severe braking that the driver of vehicle n (i.e. the follower) wishes to undertake" (Gipps, 1981) [m/s^2];

- $\theta = \tau/2$ is an additional "comfort" time lag that allows the follower not to brake always at his or her maximum desired rate [s];

- $x_n(t)$ and $x_{n-1}(t)$ are, respectively, the follower's and leader's position at time t, measured at the front bumper [m];

- $S_{n-1} = L_{n-1} + Safety$ is the effective size of the leader's vehicle, that is "the physical length plus a margin into which the following vehicle is not willing to intrude, even when at rest" (Gipps, 1981) [m];

- L_{n-1} is the physical length of the leader's vehicle of the leader [m];

- *Safety* is the safety margin "into which the following vehicle is not willing to intrude, even at rest" (Gipps, 1981) [m];

- $\hat{b}_{n-1}$ is the follower's estimate of the leader's maximum deceleration rate [m/s^2];

Please note that the deceleration rates, b_n and $\hat{b}_{n-1}$, are intended as absolute values.

In practice, the driver chooses the minimum speed between two possible alternatives, where the first ($v_{a,n}$) accounts for the driver's willingness to reach his desired speed, while the second ($v_{b,n}$) aims to preserve a safe distance behind the leader.

It is worth mentioning that in the model derivation, Gipps considered also an additional term, θ (an additional "comfort" delay in the braking component of the model), added in the analytical derivation of the model to allow the follower not to brake always at his or her maximum desired rate. Gipps then assumed for θ to be equal to τ/2. In fact, Gipps proved that in the case θ = τ/2 and $b_n > \hat{b}_{n-1}$ (i.e. "the willingness of the preceding driver to brake hard had not been underestimated"; Gipps, 1981), a vehicle travelling at a safe speed would be able to maintain a safe speed and distance indefinitely. Indeed, the relative magnitude of braking rates is the cornerstone for model stability. As shown successively by Wilson (2001), $b_n > \hat{b}_{n-1}$ is a sufficient condition for the linear stability of the model.

B.3 Analyses and Applications

As already pointed out, the Gipps' car-following model is one of the most widely used models in both research applications and practice. In particular, it represents the building block for different micro-simulation software, such as AIMSUN (2012) and DRACULA (2007).

The objective of this section is to summarize the way the Gipps model is implemented in the different software or research applications, in order to understand potential benefits derived from the experiences carried out in their development.

B.3.1 Equilibrium solutions for uniform flow

The work of Wilson (2001) represents the most complete analysis of the Gipps' model. He found equilibrium solutions of the model under the hypothesis of uniform flows, in the form of steady-state.

In such conditions, all vehicles travel at the same speed (v_{eq}) and thus their spacing (h_{eq}) is constant and time-independent. The relationship which arises between speed and spacing in steady-state is the so-called *speed-headway* function $v_{eq}=V(h_{eq})$.

Steady-state solutions allowed the author to derive a monotonically increasing *speed-headway* function for the general case of $b_n \neq \hat{b}_{n-1}$, which describes the equilibrium distance of the follower h_e:

$$v_{eq}\left(h_{eq}\right)= mid\left[0,\left(\frac{\tau+\theta}{\dfrac{1}{b_n}-\dfrac{1}{\hat{b}_{n-1}}}\right)\cdot\left(-1+\sqrt{1+\frac{2\cdot\left(h_{eq}-S_{n-1}\right)\cdot\left(\dfrac{1}{b_n}-\dfrac{1}{\hat{b}_{n-1}}\right)}{\left(\tau+\theta\right)^2}}\right),v_n\right] \qquad (B.4)$$

Since vehicles will drive more slowly (for safety reasons) as the spacing decreases, *V* is expected to be an increasing function. Following the mathematical derivations, the speed-headway function is the root of the following quadratic equation in v_{eq}:

$$\left(\frac{b_n}{\hat{b}_{n-1}}-1\right)\cdot v_{eq}{}^2 -2\cdot b_n\cdot\left(\tau+\theta\right)\cdot v_{eq}+2\cdot b_n\cdot\left(h_{eq}-S_{n-1}\right)=0 \qquad (B.5)$$

Graphically, it results in the parabolic curve shown in Figure B.1 (taken from Wilson, 2001).
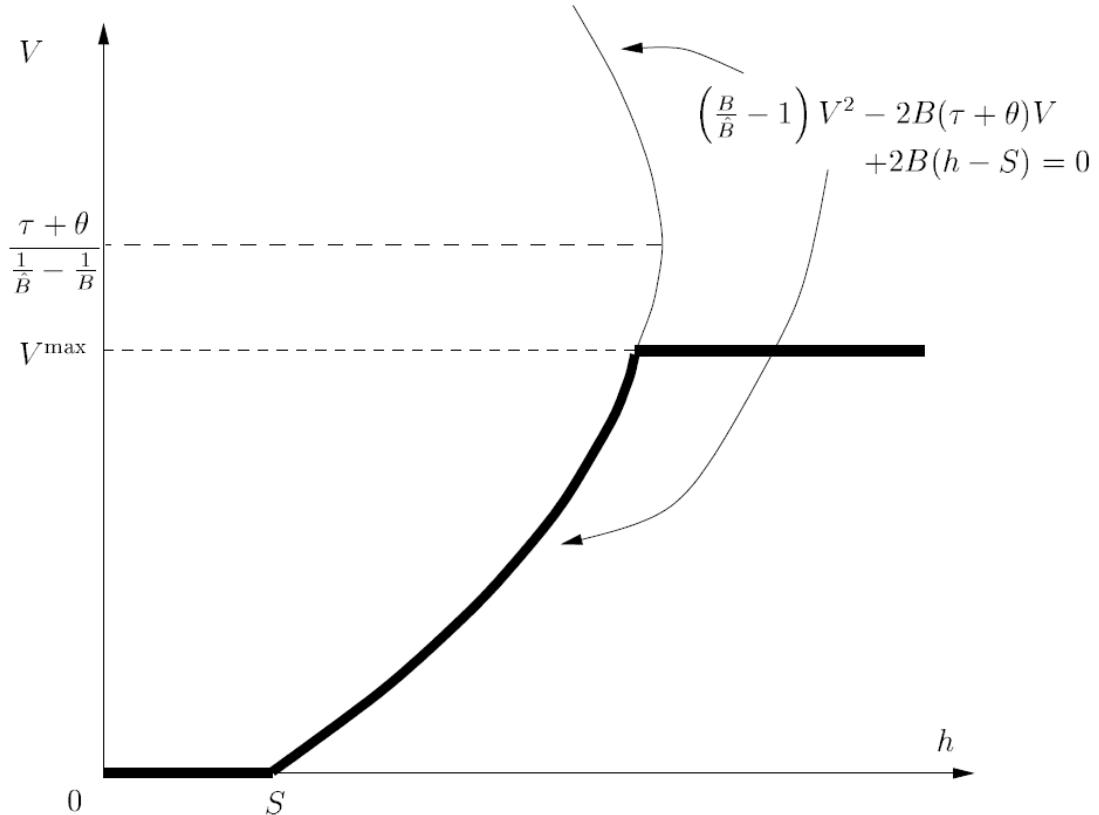
Figure B.1: Sketch of the *speed-headway* function under uniform flow condition (taken from Wilson, 2001).

Figure B.1 shows that the *speed-headway* function could be multi-valued at some points, for specific sets of parameters. Since it is widely accepted in traffic engineering community that this function should be a single-valued non-decreasing function, a constraint needs to be set.

The rightmost point of the parabolic curve (see Figure 3) is obtained at:

$$V^* = \frac{\tau + \theta}{\dfrac{1}{\hat{b}_{n-1}} - \dfrac{1}{b_n}} \tag{B.6}$$

Hence problems occur if:

$$V_n^{Max} > \frac{\tau + \theta}{\frac{1}{\hat{b}_{n-1}} - \frac{1}{b_n}} \tag{B.7}$$

Thus, the *speed-headway* function, obtained in uniform conditions, is well-defined if:

$$V_n^{Max} \leq \frac{\tau + \theta}{\frac{1}{\hat{b}_{n-1}} - \frac{1}{b_n}} \tag{B.8}$$

It is worth noting that condition in Eq. (B.8) is valid under the assumption of uniform flows, which can be never reached in real traffic conditions. Nevertheless, if we assume that the follower's speed and the headway at each simulation step are representative of a (possible) steady-state solution in uniform flow, the non-linear constraint in Eq. (B.8) still holds. Moreover, such constraint does not prevent the model parameters from generating a global instable car-following regime, in uniform flow. Indeed, the relation which must hold at the onset of linear instability is the following (for details, please refer to Wilson, 2001):

$$v_{eq} > \frac{\theta}{\frac{1}{\hat{b}_{n-1}} - \frac{1}{b_n}} \tag{B.9}$$

Since $v_{eq} \leq V_n^{Max}$, the region of the instable parameters, for a well defined *speed-headway* function, can be derived from the following inequalities:

$$\frac{\theta}{\frac{1}{\hat{b}_{n-1}} - \frac{1}{b_n}} < v_{eq} \leq V_n^{Max} \leq \frac{\tau + \theta}{\frac{1}{\hat{b}_{n-1}} - \frac{1}{b_n}} \tag{B.10}$$

This identifies the range of useful parameter values, where a uniform flow is unstable (i.e. there is at least one unstable value), while preserving a single-value *speed-headway* function.

According to Wilson, this should be the condition for the Gipps' model to reproduce typical traffic mechanisms, such as flow breakdown and spontaneous traffic jam formation.

Following the work of Wilson, the derivation of the *headway-speed* function (dual of the *speed-headway* provided by Wilson, 2001) allowed Punzo and Tripodi (2007) to derive

the macroscopic traffic flow models (i.e. the fundamental diagram) corresponding to the Gipps' microscopic equation, including an explicit formula for the flow at capacity. They extended the stationary models to the case of multi-class flows, also providing a framework for their calibration, that is for the calibration of the Gipps' microscopic parameters against average speeds and counts at detectors. This procedure was also applied in Ciuffo et al. (2008) where the authors showed the potential benefits of using the calibrated parameters of the stationary *speed-flow* relationship as starting point for the calibration of the AIMSUN traffic micro-simulation software.

B.3.2 AIMSUN implementation

In AIMSUN (2012), the original Gipps' model is coupled with different control strategies. The different versions actually refer to different strategy for the selection of $\hat{b}_{n-1}$ (since it is not considered as an additional parameter). Differently from reality, the simulation environment knows the maximum deceleration rate parameter of each vehicle (as it relates to the deceleration parameter of the leader vehicle in the simulation).

For this reason, in the *first* version of the model, the condition $\hat{b}_{n-1} = b_{n-1}$ is assumed. However, this condition does not prevent the model to crash (i.e. follower vehicle intrudes leader car).

This is the case also for the *second* version of the model, where it is assumed that

$$\hat{b}_{n-1} = \frac{b_n + b_{n-1}}{2}.$$

The *third* version of the model is, instead, more in line with the original Gipps' formulation, where the parameter ϑ (the "sensitivity factor") is introduced in order to generalize the value of $\hat{b}_{n-1}$: $\hat{b}_{n-1} = \vartheta \cdot b_n$

More interestingly, this model version also introduces a minimum headway the follower wishes to undertake. This additional parameter introduces the following control strategy:

$$if \quad \left(\left[x_{n-1}(t+\tau) - S_{n-1}\right] - \left[x_n(t) + v_n(t+\tau)\cdot\tau\right] < v_{n-1}(t+\tau)\cdot minTH_n\right) \ \{$$

$$v_{n-1}(t+\tau) = \frac{\left[x_{n-1}(t+\tau) - S_{n-1}\right] - x_n(t)}{minTH_n + \tau} \tag{B.11}$$

$$\}$$

where $minTH_n$ is the new parameter representing the minimum time headway to be kept between the follower and the leader. Though interesting, this solution modifies the dynamics described by the Gipps' model.

Alternatively to this approach, in order to preserve the original model formulation and resulting dynamics, it would be wise having a relationship among the model parameters such as that presented in Eq. (B.10), based on the linear stability analysis of the model. Similar relationships for a general car-following model formulation are proposed in Ward and Wilson (2011).

B.3.3 Numerical integration schemes

The model presented in Eqs (B.1 – B.3) is a delayed differential equation (being τ the delay). In the Gipps' original paper (Gipps, 1981), solution of Eq. (B.1) is made simple by adopting an integration step just equal to the delay τ. A forward Euler method on acceleration (i.e. a trapezoidal integration scheme on speed) is there adopted for calculations. The same approach is usually applied in the common practice. In the following we will refer to this integration scheme as "classic integration scheme".

In AIMSUN (2012), instead, a different approach is adopted. The integration step dt is a sub-multiple of the delay τ (up to a minimum of 0.1s) and the Gipps' model is applied at each simulation step. The space travelled by the vehicle is calculated considering the speed constant over the integration step. This method allows for a more accurate solution of the system of differential equations than the approach adopted by Gipps. At the same time, however, it alters the model dynamics, producing different simulation results. In the following we will refer to this integration scheme as "continuous integration scheme".

Besides the loss of real solution in the Gipps' model due to wrong parameters combinations, we need to ensure real solutions for the model at the beginning of the simulation. In fact, especially in the case of calibrating the model against real trajectory data, the first simulation steps are usually driven by the real boundary conditions (initial speed, initial spacing, etc.), rather than by the model and its parameters. For this reason, the domain of the parameters has to be further constrained, by means of the following non-linear condition:

$$b_n{}^2 \cdot \left(\frac{\tau}{2} + \theta\right)^2 + b_n \cdot \left[2 \cdot \left(x_{n-1}(0) - x_n(0) - S_{n-1}\right) - \tau \cdot v_n(0) + \frac{v_{n-1}(0)^2}{\hat{b}_{n-1}}\right] \geq 0 \qquad (B.12)$$

Eq. (B.12) relates to the initial state of the simulation (i.e. at t=0), and prevents the following vehicle to intrude the effective size of the leader at the first simulation step.

B.4   The Acceleration Component

The acceleration component of the Gipps car-following model tries to resemble the behaviour of a driver when the headway with the vehicle ahead is sufficiently large. As pointed out in Gipps (1981), it was empirically derived.

However, for the sake of generality, it can be rewritten as follows:

$$v_{a,n}(t + \tau) = v_n(t) + \alpha \cdot a_n^{Max} \cdot \tau \cdot \left(1 - \frac{v_n(t)}{V_n^{Max}}\right) \cdot \left(\beta + \frac{v_n(t)}{V_n^{Max}}\right)^\gamma \qquad (B.13)$$

in which α, β and γ are the parameters that, in the original formulation (Gipps, 1981), as well as in the field literature, are assumed equal to 2.5, 0.025 and 0.5, respectively.

However, from a modeling point of view, they are model parameters (like the reaction time), and therefore their impact on the speed-acceleration relationship should be properly investigated, also to understand if their indirect estimation is necessary to improve model performances.

The original values proposed in Gipps (1981) allow for a_n^{Max} to have the physical meaning of the maximum acceleration attained by the vehicle. However, according to the original parameter values, this maximum acceleration rate is not attained in correspondence of $v_n = 0$ but for $v_n \cong 0.32 \cdot V_n^{Max}$, as shown in Figure B.2.

$$\frac{a_n(t)}{a_n^{Max}}$$

$$\frac{v_n(t)}{V_n^{Max}}$$
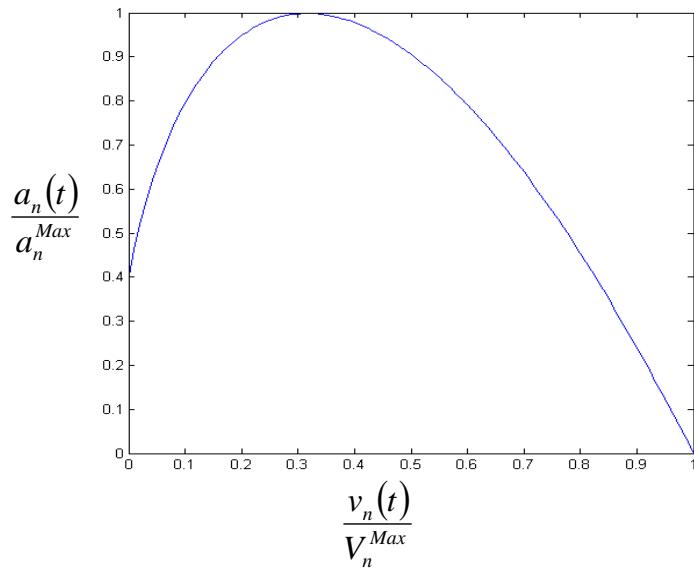
Figure B.2: Normalized speed-acceleration function resulting from the adoption of the original parameter valued for α, β, and γ.

The speed-acceleration profile defined by the Gipps' model is quite different from real speed-acceleration profiles, due, in particular, to the presence of shifting gears (see the blue lines Figure B.3).

$$a_n(t)$$

$$a_n(t)$$
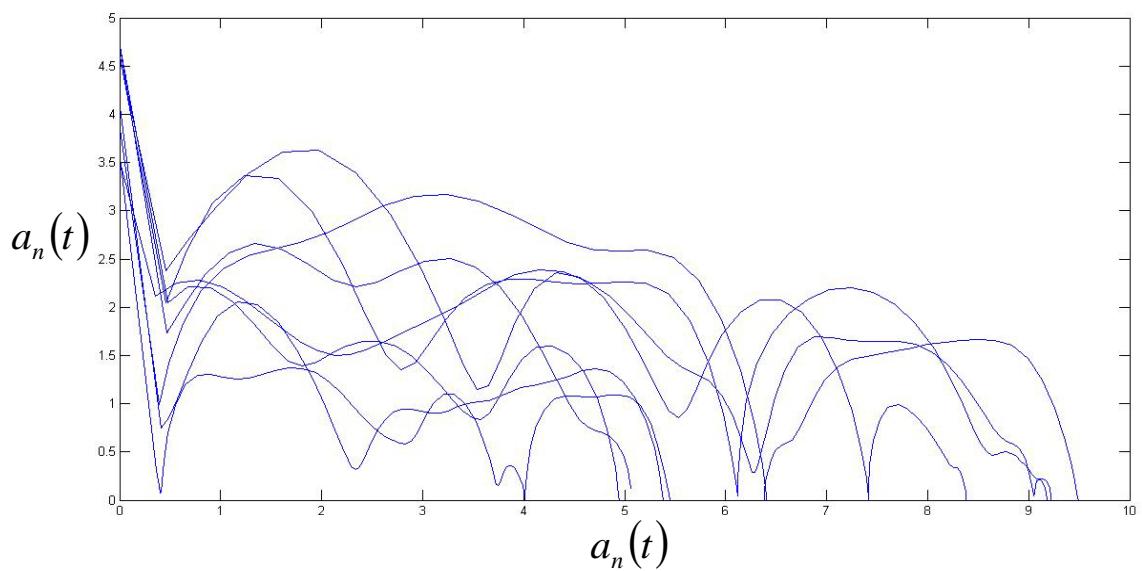
Figure B.3: Speed-acceleration function from real trajectory measurements (each blue line relates to a single trajectory measurement) and from the model (green line). A (possible) desired trend is represented by the red line.

Therefore, it is claimed in this work that some improvements can be achieved by calibrating α, β, and γ parameters (in particular because the maximum acceleration seems to arise for $v_n = 0$ in real driving).

On this basis, we can rewrite Eq. (B.13) as follows:

$$v_{a,n}(t + \tau) = v_n(t) + a_n^{Max} \cdot \tau \cdot k(t) \tag{B.14}$$

with:

$$k(t) = \alpha \cdot \left(1 - \frac{v_n(t)}{V_n^{Max}}\right) \cdot \left(\beta + \frac{v_n(t)}{V_n^{Max}}\right)^{\gamma} \tag{B.15}$$

Therefore, to keep a_n^{Max} with its physical meaning of maximum vehicle acceleration, we need to impose the following condition:

$$max(k(t)) = 1 \Rightarrow max\left(\alpha \cdot \left(1 - \frac{v_n(t)}{V_n^{Max}}\right) \cdot \left(\beta + \frac{v_n(t)}{V_n^{Max}}\right)^{\gamma}\right) = 1 \tag{B.16}$$

Let assume:

$$\begin{cases} x = \dfrac{v_n(t)}{V_n^{Max}} \\ y = \alpha \cdot (1 - x) \cdot (\beta + x)^{\gamma} \end{cases} \tag{B.17}$$

We can thus rewrite the condition in Eq. (B.16) in the following form:

$$y\left(x : \frac{dy}{dx} = 0\right) = 1 \tag{B.18}$$

With few algebraic manipulations we obtain:

$$\frac{dy}{dx} = 0 \Rightarrow x = \frac{\gamma - \beta}{1 + \gamma} \tag{B.19}$$

Eq. (B.19) implies that the relationship among α, β, and γ satisfying condition in Eq. (B.16) is the following:

$$\alpha \cdot \gamma^{\gamma} \cdot \left(\frac{1 + \beta}{1 + \gamma}\right)^{\gamma + 1} = 1 \tag{B.20}$$

Different combinations of values for α, β, and γ satisfying Eq. (B.20) are reported in Figure B.4.
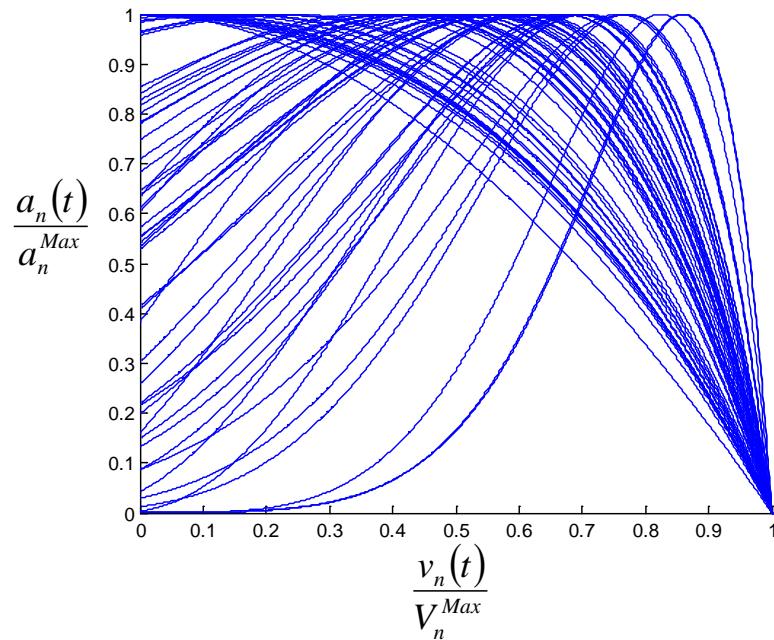
Figure B.4: *Speed-acceleration* function with α, β, and γ parameters satisfying Eq. (B.20).

However, as can be foreseen from the figure, following the proposed approach we cannot assure that the maximum of the speed-acceleration function would occur for positive value of the normalized speed, that is when $\gamma \geq \beta$.

Further, by definition, Eq. (B.20) does not allow to have *speed-acceleration* functions attaining a maximum value greater than 1 for negative values of the normalized speed, still preserving that, in the domain [0,1] it holds:

$$\left(\frac{a_n(t)}{a_n^{Max}} \right)_{v_n=0} = 1 \tag{B.21}$$

Indeed, speed-acceleration profiles derived from this condition are probably the preferred ones.
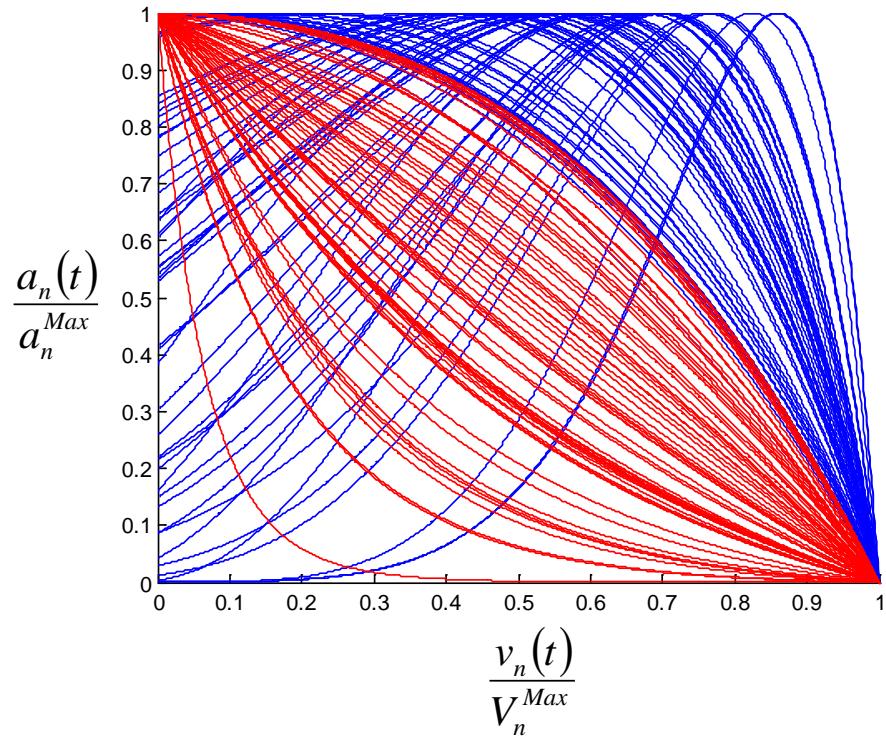
Therefore, imposing that:

$$\left(\alpha \cdot (1 - x) \cdot (\beta + x)^{\gamma} \right)_{x=0} = 1 \tag{B.22}$$

It results:

$$\alpha \cdot \beta^{\gamma} = 1 \tag{B.23}$$

Therefore, the following two conditions fully characterize the speed-acceleration function, giving rise to the profiles reported in Figure B.5:

$$\alpha = \phi(\beta,\gamma) = \begin{cases} \dfrac{1}{\gamma^\gamma \cdot \left(\dfrac{1+\beta}{1+\gamma}\right)^{\gamma+1}} & \text{if } \gamma \geq \beta \quad (a) \\[4ex] \dfrac{1}{\beta^\gamma} & \text{if } \gamma < \beta \quad (b) \end{cases} \tag{B.24}$$

Figure B.5: Full characterization of the *speed-acceleration* function according to conditions in Eq. (B.24). Blue curves relate to Eq. (B.24a), while red ones to Eq. (B.24b).

On this basis, the acceleration component of the Gipps model can be formulated as:

$$v_{a,n}(t+\tau) = v_n(t) + \phi(\beta,\gamma)\cdot a_n^{Max}\cdot\tau\cdot\left(1-\frac{v_n(t)}{V_n^{Max}}\right)\cdot\left(\beta+\frac{v_n(t)}{V_n^{Max}}\right)^\gamma \tag{B.25}$$

In order to reduce the total number of parameters, we also studied the effect of considering $\alpha = 1$. In this case, Eq. (B.24) expresses the relationship between β and γ:

$$\beta = \phi'(\gamma) = \begin{cases} 1 & \alpha = 1, \gamma \leq 1 \\ \dfrac{1+\gamma}{\gamma^{\frac{\gamma}{\gamma+1}}} -1 & \alpha = 1, \gamma > 1 \end{cases} \tag{B.26}$$

Resulting speed-acceleration plots in case of $\alpha = 1$ are plotted in Figure B.6.
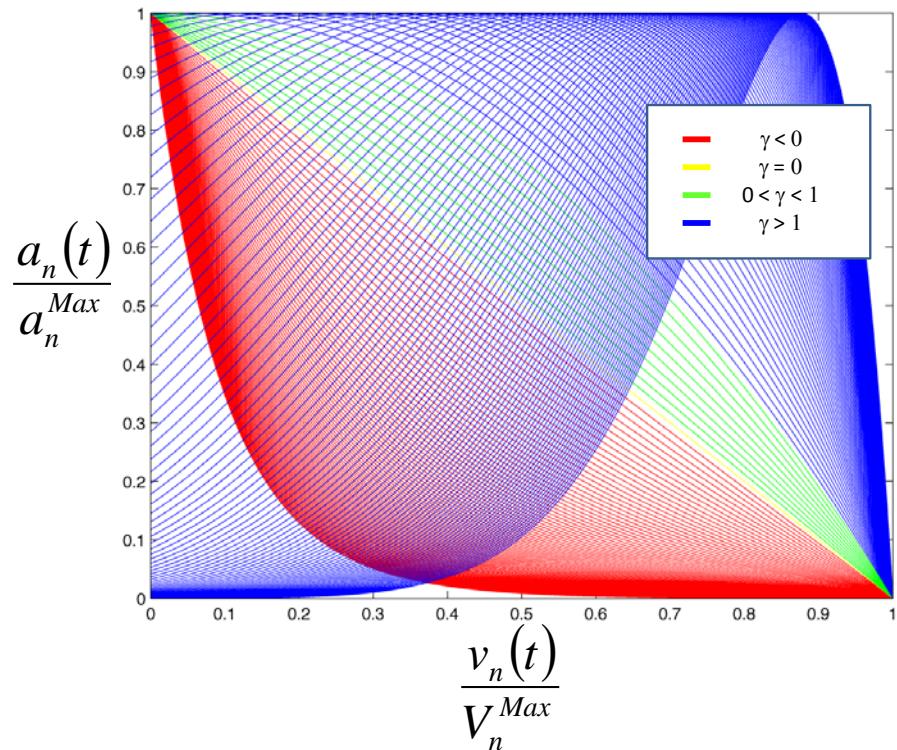
Figure B.6: Full characterization of the *speed-acceleration* function in case of $\alpha = 1$, based on Eq. (B.26).

As shown in Figure B.6, convexity types present in Eq. (B.24) are preserved also in Eq. (B.26), when fixing $\alpha = 1$.

Appendix C

Goodness Of Fit Functions in the Frequency Domain

C.1 Introduction

The contents of this Chapter are reported in Montanino et al. (2012).

When dealing with car-following models, various settings have been specified across the years to solve the problem of indirect estimation of model parameters against vehicle trajectory data. An exhaustive literature review on this topic is given in Chapter 3.

All that said, few efforts have been devoted so far in order to understand the influence of the calibration setting (combination of optimization algorithm, Measure of Performance and Goodness Of Fit function) on the results, despite its expected relevance. This topic was also investigated in Chapter 3. The study, while confirming the complexity of the estimation problem, also showed that none of the tested settings gave completely satisfactory results.

A possible explanation may be found in the fact that the commonly used GOFs are not able to capture the dynamics of the traffic measurements, i.e. the correlation structure of the time-series data which calibration is performed against.

Indeed, this may be a consequence of the integral nature of such GOFs (for example, the error measures), that simply cumulate the residuals between observed and simulated

outputs, but are not able to capture the time correlation in the trajectory data. To clarify this point, we may draw, for instance, two trajectories that have completely different patterns but the same score in the goodness of fit function (e.g. a first trajectory oscillating around the measured trajectory and a second one that, instead, is simply a translation of the measured one). What usually happens, therefore, is that only a limited portion of trajectory data presenting high residuals from the measured trajectory – because reflecting measurement errors or depicting non-normative behaviors that models are unable to capture – drives the whole parameters estimation (i.e. the estimation of parameters that affect the simulation of the whole trajectory). This local compensation effect can yield parameters values which, though minimizing the residuals, provide trajectories that do not reflect the actual driving pattern or, at least, that are not the "best" obtainable with that model.

In this study, therefore, a time-series approach is adopted to solve such troubles, which basically means defining a goodness of fit function to be evaluated in the frequency domain rather than in the time domain. This is argued to provide estimates of parameters values which can better capture the driving pattern (we could say the "driving style") because exploiting the information on the autocovariance in the time series data. In such a way, therefore, the estimation would also not been driven by the local "irregularities" in the data.

The Appendix is organized as follows. Section C.2 will briefly introduce the methodology adopted in this study to evaluate model performances in the frequency domain. Section C.3, instead, mirrors at the description of the application, presenting the case study (model and data), the design of experiment, and the results of model calibration using the proposed GOF in the frequency domain.

C.2   Methodology

According to the findings presented in Chapter 3, in the field of car-following model calibration, the use of error measures, as well as statistical GOF functions, may lead to ill-posed problems. This mainly depends by the integral nature of the traditional objective functions which locally cumulate the errors, but are unaware of the consecutiveness (dynamics) of the observations.

On the other hand, observed time series is a realization of a stochastic process, giving rise to a random signal. From this point of view, spectral analysis is a well suited statistical tool commonly employed in the physical sciences to study the time-dependent nature of physical processes. As observations are generally autocorrelated (i.e. time-wise correlated), an investigator cannot apply the statistical tools commonly used for studying independent observations. Spectral analysis, however, can be used to study the salient properties of such processes and to present them in an easily interpretable fashion for descriptive and comparative purposes.

In the hypothesis of wide-sense stationary stochastic process (i.e. mean and variance are constant over all time indexes and the covariance between two arbitrary time indexes n and m depends only on the difference *(n – m)* and not on the values of *n* and *m* themselves), mathematical models known as covariance stationary stochastic processes are useful representations of autocorrelated time series. The covariance between two observations x_n and x_{n+k} of a stationary stochastic process is defined as:

$$r(k) = \text{cov}(x_n, x_{n+k}) = E\big[(x_n - \mu) \cdot (x_{n+k} - \mu)\big] \tag{C.1}$$

The quantity *r(k)* is defined for all integer values of *k*, and it is called the autocovariance function of *X*. It measures the covariance between pairs at a distance or lag *k*, for all different values of *k*. Therefore, it is a function of lag *k*.

The autocovariance function represents all there is to know about a normally distributed stochastic process because together with the mean, it completely specifies the joint probability distribution function of the data. Other properties may be interesting, but they are limited to the single realization of the stochastic signal or process at hand. If the process is approximately normally distributed, the autocovariance function will describe most of the information that can be gathered about the process. Only if the distribution is far from being normal, it might become interesting to study higher order moments or other characteristics of the process.

Like the covariance between two variables, the autocovariance function *r(k)* also can be normalized to give the autocorrelation function (ACF) *ρ(k)*:

$$\rho(k) = \frac{r(k)}{\sigma_X^2} \tag{C.2}$$

The value for the autocorrelation at lag 0 is 1. This also follows from the definition of stationarity where the correlation should be only a function of the time lag between two observations; the lags -k and k are equal in that respect. Thus, the autocorrelation function is symmetrical about the origin (lag 0) where it attains its maximum value (i.e. 1). An example of the ACF plot is presented in Figure C.1(a), for a couple of leader/follower vehicles.

It can be shown that the Fourier transform of the autocovariance function $r(k)$ is the spectral density function (spectrum) of the signal $h(\omega)$ and, as a consequence, the Fourier transform of $\rho(k)$ equals to the normalized spectrum $\varphi(\omega)$. Therefore, the analysis of the correlogram (i.e. plot of the ACF as function of the lag k) is the analogous of the spectral analysis in the frequency domain. The idea beyond the work is to use the ACF of the time-series to compare the performances of the simulation model in reproducing the autocorrelated observed signal. In the literature, absolute accuracy measures compute the absolute value of the differences between spectra, through the so-called Integrated Mean Square Error (IMSE) measure:

$$IMSE = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\phi^{obs}(\omega) - \phi^{sim}(\omega) \right]^2 d\omega = \sum_{k=-\infty}^{\infty} \left[\rho^{obs}(k) - \rho^{sim}(k) \right] \qquad (C.3)$$

The estimation of the autocovariance function requires some care because a true autocovariance function should be positive-semidefinite. That is a prerequisite for a positive Fourier transform at all frequencies. Because that Fourier transform should represent the power spectral density, it is necessary that it is not negative for any frequency. Nevertheless, the estimator that has been mostly used in computation is based on the definition of the covariance between two stochastic variables, applied to each lag individually. By taking the average covariance between the two stochastic variables x_n and x_{n+k} for different values of index n, an estimate for $r(k)$ is found. This estimator is often called the "sample autocovariance" or the "lagged product autocovariance". Combining all individual estimates for different values of k gives the estimated autocovariance function.
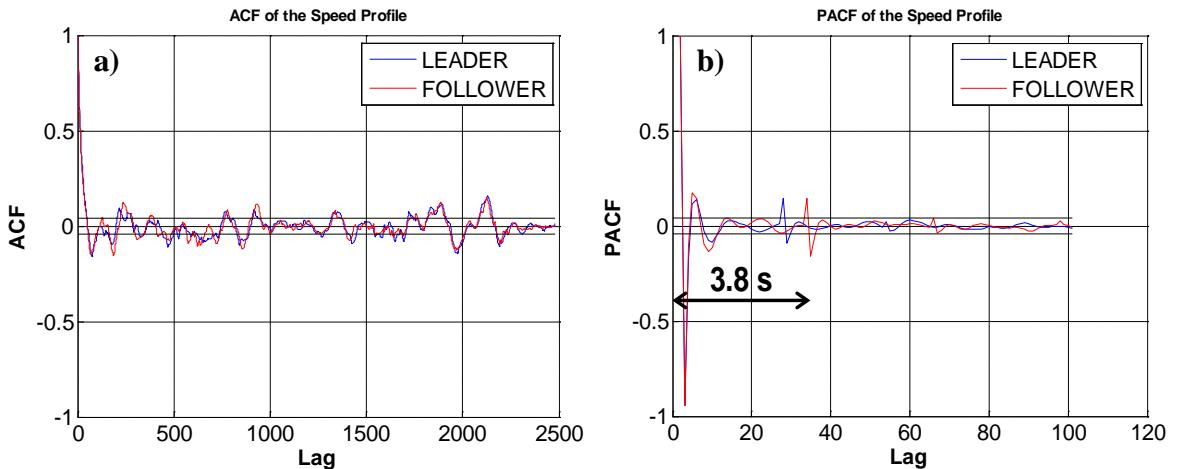
Figure C.1: ACF (a) and PACF (b) plots of the speed profiles of the leader and follower vehicles, from the experiment *30B* used in this study (see Section C.3.3 for data description). The two horizontal black lines represent the upper and lower bounds of the confidence interval with level of significance of 95%.

Figure C.1(a) shows the autocorrelation function plots related to the speed time-series of a couple of leader (blue line) and follower (red line) vehicles. In this case the *X* vector in Eq. (C.1) is made by the $i=1,\dots,N$ observation of the speed data, representing the time-series. For data description, please, refer to Section C.3.3. The two horizontal black lines represent the upper and lower bounds of the confidence interval with level of significance of 95%. If the ACF values oscillated entirely within this range from a given lag value k*, the confidence interval would tell us that such oscillation is not significantly different from being zero, and therefore, observations with lags greater than k* could be considered uncorrelated. However, we can see that uncorrelation among successive observations is never reached. This is due to the approximation in the calculation procedure of the total autocorrelation (Broersen, 2006). A more accurate estimation of the autocovariance function can be expressed through the partial autocorrelation function (PACF). Indeed, the PACF plot presented in Figure C.1(b) shows that the autocorrelation completely vanishes for lags greater than about 4 seconds (data was acquired at 10 Hz; please, refer to Section C.3.3 for details), though we can observe a clear autocorrelation only for lags smaller than 1 second.

However, a more accurate estimation of the autocovariance function through the PACF is performed at the expense of much longer computation time. For this reason, in this

study, we adopted the ACF, rather than the PACF, as a proxy of the autocovariance function to evaluate the IMSE.

It is worth noting that the study objective is to highlight the need of a specific analysis of the autocorrelation properties of time-series in order to effectively compare simulation results with observations. Therefore, this primary attempt is not to be intended conclusive, as different (more reliable) error measures in the frequency domain (i.e. directly obtainable from the analysis of the power density spectrum) might be used (see, for example, the cepstrum measure in Broersen, 2006). Further, the application of the Fourier analysis is strictly conditioned to the hypothesis of wide-sense stationary processes. In this exploratory study, this assumption has been taken for granted, though we are aware that traditional observations of traffic measurements are far from being covariance stationary because of their relative shortness. Further, the ACF unequivocally represents the complete information enclosed in the time series only in normally distributed stochastic process, as it was assumed herein.

Therefore, the analysis of non-linear and non-stationary time-series will be the objective of future research.

C.3 Case Study

In the following paragraphs, the IDM car-following model (Treiber et al., 2000) is reviewed from the literature. Then, motivations and procedure to generate synthetic data are described. Finally, the optimization problem is set up together with the design of the experiments.

C.3.1 The model

The Intelligent Driver Model (IDM) belongs to the class of social force models (Treiber et al., 2000). In this section we will recall the model formulation from the literature, while a more exhaustive review can be found in Chapter 5, where the model was introduced:

$$a_f(t) = a_f^{Max} \cdot \left\{ 1 - \left[\frac{v_f(t)}{V_f^{Max}} \right]^{alpha} - \left[\frac{\Delta S^*(t)}{\Delta s(t) - L_l} \right]^2 \right\}$$

$$\Delta S^*(t) = \Delta S_0 + max \left\{ \Delta S_1 \cdot \sqrt{\frac{v_f(t)}{V_f^{Max}}} + T \cdot v_f(t) + \frac{v_f(t) \cdot [v_f(t) - v_l(t)]}{2 \cdot \sqrt{a_f^{Max} \cdot |b_f|}}, 0 \right\}$$

(C.4)

where:

- $v_f(t)$ and $a_f(t)$ are, respectively, the follower's speed and acceleration at time t;

- V_f^{Max} is the follower's maximum desired speed;

- a_f^{Max} corresponds to the acceleration applied by the follower at a start when the distance from his/her leader is much bigger than the distance Δs_0; it also corresponds to the deceleration of a vehicle which is travelling at its desired speed and whose distance from the leader approximates the desired distance;

- b_f is a sort of deceleration rate between normal and emergency conditions (Treiber et al., 2000);

- $v_l(t)$ is the leader's speed at time t;

- L_l is the physical length of the leader's vehicle;

- $\Delta s(t)$ is the difference between leader's and follower's positions at time t, taken from the front bumper;

- $\Delta S^*(t)$ is the rear end-front follower's desired distance from the leader;

- ΔS_0 is the rear end-front follower's desired distance from the leader at stop;

- ΔS_1 is a non-zero parameter necessary for features requiring an inflection point in the equilibrium flow-density (Treiber et al., 2000);

-       T is the minimum time headway between leader and follower;

-       *alpha* is a model parameter.

It is worth noting that the *max(...)* operator in equation 4 is necessary in order to avoid that the follower's desired distance from the leader becomes lower than ΔS_0, for negative speed differences (i.e. $v_f(t) < v_l(t)$).

C.3.2 Data description

According to verification methodology presented in Chapter 3, preliminary to any performance comparison among different optimization problem specifications (i.e. a combination of MoP, GOF function and optimization algorithm), one should first verify that each "testing" specification is able to find the global optimum solution, i.e. the value of model parameters that correspond to the global minimum of the objective function.

In this view, when calibrating the model parameters against real data, the global minimizer (i.e. the set of optimal parameter values) is unknown (for more details, please, refer to Chapter 3). On the other hand, in the case of synthetic data, i.e. generated from the model itself by fixing the model parameters to a set of known (or "true") values, the global minimum of the optimization problem is known. Therefore, if the calibration procedure fails in "rediscovering" it, it is useless to perform any comparison based upon real data.

Regarding the choice of the real leader's trajectory to feed the IDM car-following model, we felt comfortable to assess, based on recent findings, that car-following models are more likely to better reproduce short vehicles trajectories rather than long ones. Indeed, as a matter of fact, short trajectory data are more likely to contain less information on the variability of the driver's behavior over time – the so-called intra-driver variability – and, thus, they are more likely to be better reproduced by a single set of model parameters that is kept fixed across time, independently from the GOF function adopted in the calibration experiment. On the other hand, vehicles' trajectories that last longer have a higher probability to contain information related to a time-varying driving behavior of the driver, making driving dynamics more complex. In this case, the calibration experiment is more challenging, as we look for a single set of parameters able to reproduce, at the best, (possibly) different driving behaviors over time.

Further, since the aim of the study is to test the performance of a frequency domain-based GOF function against traditional error-based measures in the time domain, we thought that the use of long trajectories was even more challenging, as the hypothesis of wide-sense stationary process may be more unrealistic.

As a consequence, a significant test bed should be designed using a long trajectory with variable driving dynamics (for instance, a mix of long accelerations, heavy brakes and pure car-following).

Therefore, based on the above requirements, the leader's trajectory was taken from the experiment 30B, described in Punzo and Simonelli (2005), carried out on a two-lane rural highway in the area surrounding Naples. Vehicles trajectories are about 5-min long and present a range traffic dynamics (long acceleration, sudden brakes and stops). Trajectory data was acquired through instrumented vehicles, equipped with kinematic differential GPS receivers that recorded the position of the vehicle at 0.1 second interval. Differential GPS data were further processed by means of the procedure described in Punzo et al. (2005), based on a non-stationary Kalman filter. More details on the data can be found instead in Punzo and Simonelli (2005).

In this study, the values of the parameters that were used to generate the synthetic follower trajectory were the following: $alpha = 4$, $T = 0.5$ s, $V_f^{Max} = 22$ m/s, $a_f^{Max} = 4.5$ m/s^2, $b_f = 4$ m/s^2, $\Delta S_0 = 1$ m, $\Delta S_1 = 0$.

Figure C.2 shows the (real) leader's and the (synthetic) follower's speed profiles (respectively, the blue and the red lines).

These values are consistent with the parameter bounds set in Punzo and Ciuffo (2011) where the global sensitivity analysis was performed to draw inference on the sensitivity of model factors with respects to the system output.

Speed Profiles of the leader and of the (synthetic) follower
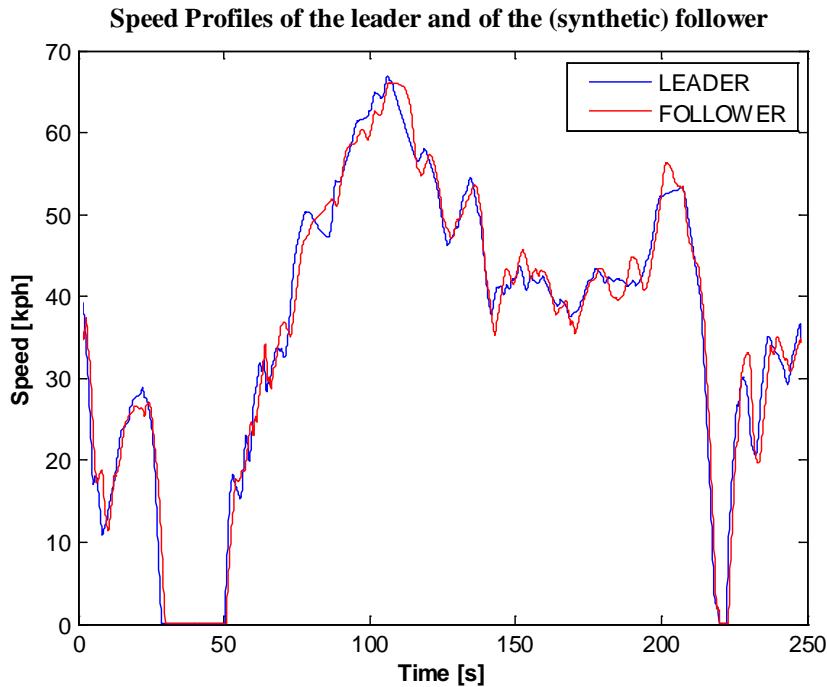
Figure C.2: Leader's (blue line) and (synthetic) follower's (red line) speed profiles.

C.3.3 Design of Experiment

To compare the performances of the GOF approach based on the ACF against traditional settings, we adopted the RMSE as the comparative error measure, according to the findings of Chapter 3, where we showed that such function allows the algorithm to rediscover the "known" values of the parameters with a probability of almost 100%, though at the expense of high computational efforts (i.e. large number of iterations to converge). The MoP here adopted is the speed.

The optimization algorithm used in the study is the OptQuest Multistart (LINDO, 2003) implemented in MATLAB (2009), where it combines the seeking behaviour of a gradient based local NLP (Non Linear Programming) solvers with the global optimization abilities of a Scatter Search. For more details, please refer to Chapter 3.

Since the starting point of the optimization procedure can strongly influence the capability of finding the global minimum (especially in the case of ill-posed problems), the robustness of the new setting has been compared with traditional ones, repeating each calibration experiment 64 times by using different starting conditions sampled from

the Sobol' LPτ low-discrepancy quasi-random sequence coded in MATLAB language (Sobol' et al., 1992).

C.4   Calibration Results

In this section, the analysis of the results of the calibration experiments on synthetic data is presented.

Firstly, we were interested in assessing the ability of each problem setting (Algorithm/GOF function/MoP) in finding the "known" global solution. For a single calibration attempt, this can be measured either in terms of the distance between the optimal solution found by the heuristic and the known global minimum or, by the score of the objective function after the calibration (having in mind that in a synthetic experiment the score of the global minimum is zero).

However, results from a single calibration attempt are not really informative on the uncertainty in the specific calibration process. In facts, calibration attempts differing in the starting point of the optimum search often provide different results (for details, see Chapter 3). For this reason, multiple calibration attempts starting from different initial points are needed. This is even more so in the case of real trajectory data which often give flat and waved response surfaces, with no "well-defined" global minimum but multiple local minima, each one potentially very far from the others.

Therefore, to evaluate performances of a specific problem setting we proposed and applied the two indicators presented in Chapter 3:

-      The "Frequency of the original parameters", which measures the number of times, out of the 64 attempts of a calibration experiment, in which the optimization algorithm was able to rediscover the original parameters (i.e. the values which generated the synthetic global optimum) with an error on the single parameter of $\pm 5\%$. This indicator reveals the ability of the specific calibration setting to find a solution in the close neighborhood of the known global solution, that is to rediscover the original parameters.

-      The "Frequency of the best score" which measures the number of attempts in which the optimization algorithm attained its best score i.e. the lowest score of the objective function over the 64 attempts of a calibration experiment. Such solution,

of course, is the best provided by the specific calibration setting but does not necessarily coincide with the known global minimum. Therefore, the indicator measures the robustness of the specific calibration setting as to the variation of the starting point of the search (but not the ability to rediscover the global minimum).

The results are presented in Table C.1. In addition, the average number of iterations needed to reach convergence was approximately the same in the two cases (about 30'000 for the RMSE and about 29'000 for the IMSE) with a standard deviation of about 10'000 in both the settings.

Table C.1: Comparison of the results of the calibration of the IDM model parameters against *synthetic* speed time-series using the RMSE and the IMSE.

ALGORITHM	GOF/MoP	Frequency of the original parameters ± 5% error (%)	Frequency of the best score (%)	Average number of iteration to converge	STD of the number of iteration to converge
OptQuest Multistart	RMSE(V)	100	100	30'611	9'203
	IMSE(V)	100	100	29'814	12'550

Apparently, the evidence here presented does not show any beneficial effect from the use of the spectral analysis with respect to traditional error measures. However, this step has to be considered only as a verification of the setting in the ability to find the global minimum. The main improvement indeed was not expected to be seen in the experiments with synthetic data (as, in this case, the optimization problem is well posed and a well-defined global minimum does exist) but in those with real data, where the commonly used error measures usually suffer from the problems described in the introduction. From an optimization point of view, this means an ill-posed optimization problem characterized by several local minima and wide flat regions in the response function against the parameters' domain.

To prove this conjecture, we performed the same calibration experiments using the real leader trajectory data from experiment 30B to feed the IDM model. Since the two GOF functions were not homogeneous among each other, a validation function was needed to cross-compare the calibration results. To this aim, the sum of Theil's inequality coefficients on speed and spacing was adopted, accordingly to the methodology

presented in Chapter 3. It is also worth noting that such validation is not fair with the proposed GOF as it is in the time domain.

The results are presented in Table C.2.

Table C.2: Comparison of the results of the calibration of the IDM model parameters against *real* speed time-series using the RMSE and the IMSE.

ALGORITHM	GOF/MoP	Frequency of the best score (%)	Validation Score	Average number of iteration to converge	STD of the number of iteration to converge
OptQuest Multistart	RMSE(V)	100	0.20	36'653	5'055
	IMSE(V)	100	0.11	38'080	232

With both the measures of goodness of fit, the algorithm always gets the same solution (which is therefore expected to represent the actual global minimum of the two optimization problems) even considering always different starting points. However:

- the two solutions found in the two experiments are different in terms of parameters' values.

- the solution found using the new GOF in the frequency domain resulted in a significantly lower value of the validation score.

- looking at the average number of iterations, values are almost the same, but the deviation from these average values are much greater when the RMSE is applied instead of the IMSE.

These results seem confirming that the proposed GOF better specifies the optimization problem as it results *i)* more robust to the choice of the starting point in the optimization, and *ii)* computationally less expensive.

Appendix D

Framework for the Calibration of Not-Stochastic Lane-Changing Model

D.1 Introduction

In the literature, different types of lane-changing models have been proposed across the years. According to a very basic classification, that is quite accepted in the field research, we may distinguish between rule-based and discrete choice-based lane-changing models.

Rule-based models – such as, for example, those proposed by Gipps (1986), Yang and Koutsopoulos (1996), Yang et al. (2000), Kesting et al. (2007), and so on – simulate driver's choice to change or not-change lane as a binary choice dependent on the resulting interactions that a possible lane-change would produce on the surrounding vehicles in the traffic stream (e.g. variation of the deceleration rate of the follower vehicle in the target lane). Provided their nature, these models are not stochastic, and therefore they will be indicated here as "not-stochastic" lane-changing models. Further, given the easiness in the interpretation of model parameters when used by practitioners, these models are frequently adopted in commercial micro-simulation software (e.g. AIMSUN(2012), VISSIM(2011), PARAMICS (2003)).

Conversely, discrete choice-based models – for instance, those proposed by Ahmed (1999), Toledo (2003), Choudhury (2007) – predict driver's behavior through random

utility models (RUM), where the probability of lane-changing depends on both driver's perception/latent attributes and the surrounding traffic conditions (e.g. available gap).

Unlike the second class of models, whose estimation problem was rather consolidated in the framework of RUM and deeply investigated in the last years, the calibration of rule-based lane-changing models is a very under-researched issue.

However, the indirect estimation of rule-based lane-changing model parameters is deemed to be very important for the use of commercial simulation packages and, at the best of our knowledge, no methodological framework to handle this problem was provided in the field literature.

Therefore, the objective of this study is to propose a preliminary methodological approach to rule-based lane-changing model calibration, being aware that further investigation and research on this topic is warmly needed to establish a reliable framework.

For this purpose, the test model adopted here in the MOBIL lane-changing model (Kesting et al., 2007) reviewed in Section D.2. The proposed methodology, applied to the test model, instead, is briefly summarized in Section D.3.

D.2 MOBIL Lane-Changing Model

The MOBIL lane-changing model (Kesting et al., 2007) takes into account the anticipated advantages and disadvantages of a prospective lane change in terms of single-lane accelerations. Compared with the explicit lane-changing model, the formulation in terms of accelerations of a longitudinal model has several advantages. Among the others, the use of the acceleration function of the car-following model as the "potential benefit" function for lane-changing choices allows for a compact formulation with a small number of additional parameters, while ensuring consistency and integration with car-following model. For a more detailed discussion, please refer to Kesting et al. (2007).

Given symmetric lane-changing rules (Kesting et al., 2007), the probability of a vehicle i to change lane from the current lane to a target one is formulated in Eq. (D.1):

$$p_i(t) = \begin{cases} 1 & \text{if } \tilde{a}_i(t+1) - a_i(t+1) + pf \cdot [\tilde{a}_n(t+1) - a_n(t+1) + \tilde{a}_c(t+1) - a_c(t+1)] > \Delta a^{\text{threshold}} \text{ AND } \tilde{a}_n(t+1) \geq -b_{\text{Safe}} \\ 0 & \text{otherwise} \end{cases}$$

where:

- $p_i(t)$ is the probability that vehicle i would change lane to the target lane at time t;

- $a_i(t+1)$, $a_n(t+1)$ and $a_c(t+1)$ are, respectively, the acceleration of the vehicle i in the current lane, and of its follower vehicles n (in the target lane) and c (in the current lane) at time $t+1$, assuming that vehicle i did not change lane at time t;

- $\tilde{a}_i(t+1)$, $\tilde{a}_n(t+1)$ and $\tilde{a}_c(t+1)$ are, respectively, the acceleration of the vehicle i in the target lane, and of its follower vehicles c (in the current lane) and n (in the target lane) at time $t+1$, assuming that vehicle i did change lane at time t;

- pf is the politeness factor, defined in the range [0, 1], that takes into account how much the vehicle i takes into account the disadvantage (acceleration losses) caused to the follower vehicles n (in the target lane) and c (in the initial lane);

- $\Delta a^{\text{threshold}}$ is a parameter that "prevents lane changes if the overall advantage is only marginal compared with a keep lane directive" (Kesting et al., 2007);

- b_{Safe} is the maximum safe deceleration rate that "prevents accidents as long as its value is not greater than the maximum possible deceleration of the underlying longitudinal model" (Kesting et al., 2007).

For further details on the theoretical hypothesis at the basis of model formulation, and for a review of the application contexts, please refer to Treiber and Kesting (2013).

As the "potential benefit" of a lane-change choice for a vehicle i at time t depends on the single-lane accelerations of the vehicle i and of its follower vehicles in the current (vehicle c) and target (vehicle n) lanes, the acceleration functions for each of the involved vehicles needed to be preliminary estimated. This is the reason why MOBIL model parameters estimation could be performed only after the calibration of the acceleration model which is used to simulate longitudinal interaction among vehicles.

In the present work, we applied the IDM model parameters estimated for each individual vehicle.

D.3   MOBIL Model Calibration

As we wanted to disaggregate estimate MOBIL parameters for each vehicle i, at each time t we calculated the accelerations of vehicles i, n and c in Eq. (D.1) at time $t+1$ simulating the IDM model with the *measured* input vehicle positions at time t, i.e. we did not use simulated accelerations to update vehicle positions.

Indeed, we wanted to reduce as much as possible the fact the MOBIL estimation results were conditioned to the calibrated values for the IDM model parameters.

That said, the estimation of model parameters for each individual vehicle in the traffic stream aimed at reproducing the observed time evolution of driver choices, being either of changing lane or not.

On this basis, the proposed estimation framework is based on the following key considerations:

i.   the *measured* number of lane-changes performed by a driver is much smaller than the number of times in which the driver choose not to change lane;

ii.   provided the nature of the *measured* choice (which is actually binary) and its rarity, the capability of the model to reproduce the exact time instant at which the vehicle changes lane should not be considered as the measure of performance in the estimation;

Therefore, the proposed methodology is based on the concept of *scenario*.

For each vehicle i, we divided the observed trajectory in a succession of scenarios, where each scenario is defined as the time interval in which vehicle i's leaders and followers in the current and target lanes (both on the right-hand side and on the left-hand side, for symmetric lane-changing rules; for the MOBIL model, vehicles n and c) did not change.

Indeed, the set of measured states (positions, speeds, accelerations) of the interacting vehicles in a scenario (such as vehicle i, n, c, in the MOBIL model) can be considered as a *single* observation of a panel data in discrete-choice model estimations.

Therefore, in the proposed framework, for a given set of MOBIL parameters, each scenario has a positive realization if the model predicted the observed behavior of vehicle i, that could be either a lane-change or not; otherwise, it is a negative realization.

On this basis, the objective function used in the estimation process is the number of positive scenarios which occurred for a given set of lane-changing model parameters, and the optimization algorithm aims at maximizing its value.