Dealing with uncertainty in detailed calibration of traffic simulation models for safety assessment

*Carlos **Lima Azevedo**
Singapore MIT Alliance for Research and Technology
1 Create way, 138602 Singapore
Phone: +65 66011547
Fax: +65 66842118
email: cami@smart.mit.edu

Biagio **Ciuffo**
Institute for Energy and Transport, Joint Research Centre
2749 Via Enrico Fermi, 21027 Ispra, Italy
Phone: +39 0332 789732
Fax: +39 0332 786627
email: biagio.ciuffo@jrc.ec.europa.eu

João Lourenço **Cardoso**
National Laboratory for Civil Engineering
101 Av. Do Brasil, 1700-066 Lisbon, Portugal
Phone: +351 218443661
Fax: +351 218443029
email: jpcardoso@lnec.pt

Moshe E. **Ben-Akiva**
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, United States of America
Phone: +1 6172535324
Fax: +1 6172531130
email: mba@mit.edu

* Corresponding Author

Abstract

With the increasing level of detail of traffic simulation models, the need for a consistent understanding of simulators' performance and the adequate calibration and validation procedures to control uncertainty is crucial, particularly in applications focusing on complex driving behaviour and detailed outputs, such as road safety analysis.

In this work the calibration of traffic microscopic simulation models for safety analysis is analysed considering four different key uncertainty sources: the input data, the calibration methodology, the model structure and its parameters, and the output data. The use of a multi-step sensitivity analysis (SA) framework is proposed and applied to the simulation of an urban motorway scenario, using a complex traffic simulation model with more than one hundred parameters. A three-level analysis is presented: (1) different advanced SA and calibration methods are described, compared and integrated in a multi-step global SA framework; (2) the proposed method is tested using both vehicle trajectory and aggregated traffic data to assess the impact of model parameters uncertainty and different types of input data on relevant outputs; and (3) accident and non-accident scenario-specific calibrations are performed to test the capacity of the simulator in replicating changes in detailed traffic and safety related measurements. Different techniques are adopted in each phase of the global SA and calibration method, attending to the problem complexity, the dimensionality of the experiment, and minimizing the necessary number of model evaluations.

The proposed method successfully identified the role played by all parameters and by the model stochasticity on different safety outputs. The final model calibration, carried out by explicitly considering the presence of uncertainty at different levels, confirmed the potential of advanced microscopic traffic models to adequately replicate detailed traffic and safety measurements, shedding light on different aspects of the interaction between road safety and traffic dynamics.

Keywords

Road safety, traffic simulation, uncertainty management, sensitivity analysis, calibration

1 Introduction

Traffic micro-simulation tools have been developed and enhanced based on an increasing level of modelling complexity. It is becoming recognized the crucial importance of analysing these models, understanding how they work and, in particular, what influences their capability to reproduce the socio-physical phenomena they are intended to simulate. Global sensitivity analysis (SA) is the family of tools to be used with this aim. Together with uncertainty analysis, SA studies how the uncertainties in model inputs affect the model response (Saltelli et al., 2008). These analyses are of high importance in reducing the complexity of the calibration task and minimising the burden of non-influential parameters in such optimization process.

Generally, previous SA on micro-simulation models refers to applications to a sub-model with few parameters, with a clear focus in car-following (CF) behaviour. In fact, when dealing with complex traffic simulation models, it is common practice to make a selection of the parameters to involve in the sensitivity analysis and calibration. On top of this, simplified approaches such as the one-at-time (OAT) approach remain the most adopted method when dealing with microscopic simulation models (see for example Mathew et al., 2010 or Kesting et al., 2008). OAT approaches are based on the estimation of partial derivatives, and assess how uncertainty in one factor affects the model output keeping the other factors fixed at a nominal value. The main drawback of this approach is that interactions among factors cannot be assessed, since they require inputs to be changed simultaneously for several variables. Furthermore, this method restricts the analysis of the model response to the proximity of a certain point, rather than allowing for exploring its full input space (Daamen et al., 2014). Multi-factor analysis of variance (ANOVA) has also been used in the SA of traffic simulation models (see for example Park and Qi, 2005). It allows analysing the effect of two or more parameters on a response variable and it is used to determine both the first-order and the interaction effects between parameters and a response variable. Further to using the standard definition of ANOVA, a more efficient method based on variance decomposition can be used in model SA. This method consists in evaluating two types of sensitivity indices and represents the most advanced and conceptually sound way of performing model SA (Saltelli et al., 2008). It can accommodate parameters interactions and a comprehensive analysis of their input space. However, this method may still require a large number ($N \cdot [k + 2]$, being k the number of parameters and N the dimension of the Monte Carlo experiment) of model evaluations when dealing with complex traffic models. This approach was successfully applied by for the SA of two CF models (Punzo et al., 2009).

More recently, Ciuffo et al. (2014) proposed a multi-step approach to the use of variance-decomposition SA on computationally expensive and high-dimensional traffic simulation models. At each step, a variance-decomposition-based analysis is applied to groups of parameters, selected on the basis of their possible common features, allowing for the reduction of the number of evaluations and focusing at each step on more sensitive parameters. When applying it to a complex driving behaviour model with 101 parameters (Toledo et al., 2007) using aggregated data from loop sensors, the multi-step approach required 80% less model evaluations, when compared with a full SA technique.

While the large majority of these calibration studies focused on aggregated traffic variables as measures of performance (MoP), simulation studies focusing on road safety depend on detailed calibrated outputs such as accelerations, headways and lane changing decisions. Indeed, the

importance of using trajectories when analysing detailed driving behaviour and interaction outputs has been pointed out in recent studies (Cunto et al., 2008, Jie et al., 2013 and Ciuffo et al., 2014). Focusing in rear-end crashes at signalized intersections, Cunto et al. (2008) used a multi-step approach to identify the most sensible parameters of the CF, lane-changing and stochasticity models of VISSIM (PTV, 2009) using real trajectory data. The measure of performance was the crash potential index (CPI) a surrogate safety assessment measure based on the acceleration differential and exogenous conditions. From 13 initial parameters, a first ANOVA reduced the number of sensitive parameters to 6 and a subsequent fractional factorial analysis reduced it to 3. A final genetic algorithm (GA) for calibration estimated the final values of the parameters. Duong et al. (2009) then extended this framework to a multi-criteria optimization, where the CPI was coupled with traffic volume and speed during the optimization of the GA.

In Jie et al. (2013) trajectories were used to calibrate a subset of VISSIM parameters (PTV, 2009) regarding MoP based on speeds and accelerations. After fine tuning a few driver heterogeneity parameters individually, a local (individual and group) OAT SA was carried out on the parameters related to the car-following behaviour. Although focusing in emissions modelling, this study also showed the different detailed kinematic outputs, when using trajectories or aggregated data in the calibration process. Yet, the scope of its SA was limited to 8 parameters and based on a naive method.

Ge et al. (2014) presented a comparison between the variance-decomposition method and a Kriging-based approach (Ciuffo et al., 2013) coupled with the quasi-Optimized Trajectories Elementary Effects (quasi-OTEE) screening technique (Ge and Menendez, 2013) regarding the identification of sensitive parameters of the Wiedemann-74 CF model (PTV, 2009) using trajectory data. The quasi-OTEE SA was used to identify the whole sub-set of influential parameters from the initial 25 set, and the Kriging-based SA was then used to refine the analysis and correctly rank the most influential parameters in a more reliable way, using 40 times less model evaluations. Along with Ciuffo et al. (2014), this study revealed the potential of coupling and replacing variance-decomposition methods with less demanding ones. Furthermore, it has been demonstrated in several studies (Toledo et al. 2004; Cunto et al., 2008; Ciuffo et al., 2014) that coupling advanced SA techniques with metamodels may significantly reduce the computational burden of the calibration and validation tasks. By definition, a metamodel is an approximation of the input/output function defined by the simulation model, where per each MoP and goodness-of-fit (GoF) combination a surrogate of the simulation model could be computed and used for parameter calibration.

In summary, when detailed trajectory data is available for model calibration, four different approaches may be considered:

- **Re-estimation**, where the model is re-estimated using either traditional maximum likelihood or Bayesian approaches based on the most recent set of trajectory data. This approach has been used in the initial estimation of new driving behaviour sub-models (see for example, Toledo et al., 2007).
- **Conditional re-estimation**, in which the model is estimated with a traditional Bayesian approach using the new trajectory data set as main data, but introducing prior knowledge on the parameters values based on the previous estimations. Hogendoorn et al. (2010) used a

maximum likelihood estimator accounting for prior information on the parameter values and serial correlation in trajectory data to estimate a CF model.

- **Disaggregate calibration method**, where each real trajectory observation point is compared with simulated values and the parameters are calibrated using a numerical optimization algorithm. The difficulty in such method is the set-up of starting conditions when using complex driving behaviour models. This method was used in Brockfeld et al. (2004) and Kesting et al. (2008), for example, for the specific calibration of the CF model.
- **Aggregate calibration method**, where a set of aggregate statistics of the real trajectories are pre-defined and compared against the simulated statistics, also using an optimization method.

The difficulty in the first two methods lies in complex likelihood functions and estimation procedures that account for simultaneous estimation of the multiple sub-models that form the driving behaviour framework of traffic simulators. The calibration-based methods have to deal with a large number of model simulations to correctly deal with input, output, model and calibration uncertainties during the calibration algorithm optimization procedure. Yet, the above mentioned recent developments point to the potential benefit of their application in the calibration using detailed data, and therefore solving several limitations of previous methods.

In the next section a generalization of the multi-step framework proposed in Ciuffo et al. (2014), where different SA methods can be selected depending on the sensitivity of the model parameters, is proposed and formulated for the calibration of traffic simulators. In Section 2, the general mathematical details of the various methods used within the multi-step framework are described. The case-study application is then described in Section 3 and its results presented in Section 4. Finally, the last section sketches the main conclusions of the presented study.

2 Calibration Method

2.1 The proposed framework

When focusing on safety, the uncertainty in the driving behaviour and the traffic demand parameters is considerable, due to the high number of outputs of interest. In fact, multiple interactions between vehicles and several detailed vehicle motion variables are of interest. All these variables may vary significantly with different traffic scenarios contexts, such as congestion, road characteristics or weather conditions. Unfortunately, large sets of real trajectories have not yet been collected for all these different scenarios, making the direct calibration of driving behaviour under such conditions impossible. To overcome this limitation, a three level process is proposed: (1) a global SA and calibration using trajectories on a generic day; (2) a global SA to identify the most sensitive parameters regarding aggregated traffic data; and (3) scenario-specific calibrations of the parameters identified in (2) using the calibrated values from (1) as initial values (see Figure 1).

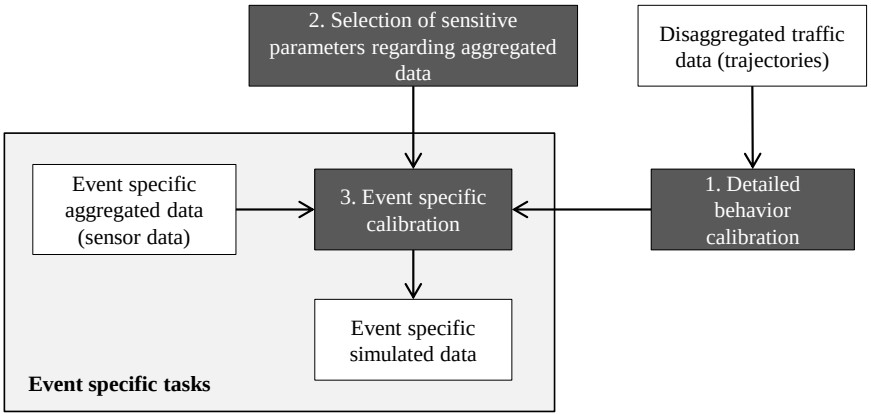


Figure 1 – Proposed framework for road safety oriented calibration.

As traffic simulation models were estimated with data sets from different environments, the use of trajectories collected on the specific site under analysis will improve the replication of the generic driving behaviour (Zhang et al., 2008). Following the proposed framework, existing aggregated data is then used to re-calibrate some relevant driving behaviour parameters and estimate the demand regarding temporal and local specific conditions. Both the calibration of the generic behaviour and the selection of parameters to calibrate using aggregated data are based in global SA, leveraging the control of uncertainty. It is important to point out that any simulated safety variability is therefore linked to changes in detailed traffic conditions and not to any individual risky manoeuvres. This assumption follows the concept of the *safety continuum* that underlies the traffic conflicts technique (Hýden, 1987).

2.2 The multi-step global SA

For the identification of the most sensitive parameters and for their global calibration using trajectories (tasks 1 and 2 in Figure 1), a generalization of multi-step global SA proposed in Ciuffo and Lima Azevedo (2014) was applied. In this method, parameters are grouped with respect to the simulation sub-models they belong to, and a SA is carried out considering the different groups rather than the different parameters. Parameters related to the car-following, the gap acceptance or the lane utility models, for example, are grouped accordingly. Then, iteratively, the most influential parameter groups (sub-models) are singled out, and a subsequent SA on their parameters is carried out. These iterative steps are carried out until a variance-based analysis on individual parameters is feasible. On each of these intermediate steps Ciuffo and Lima Azevedo (2014) used group variance-based analysis. In the present study, we extend this method to the use of less complex SA approaches at intermediate steps, reducing even further the number of model evaluations needed. However, variance-based methods are always preferred and simpler methods are only used when the sensitivity of the parameters is lower. Further, we use the results of each SA step to calibrate the parameters that are dropped from the analysis (less sensitive). By adopting such procedure, several issues of previous SA and calibration of traffic microscopic models were overcome:

1. To consider a large number of input parameters.
2. To account for parameter interaction.

3. To correctly analyse the input space.
4. To accommodate different SA methods.

For the present case study, two different SA methods and a screening procedure were considered: the variance-based SA, the group variance-based SA and the quasi-OTEE screening method. Furthermore, a metamodelling technique was also tested in the evaluation by coupling a Kriging metamodelling with the three mentioned SA techniques. Finally, the Weighted Simultaneous Perturbation Stochastic Approximation, an advanced aggregated calibration technique, was used to optimize the subset of driving behaviour and demand parameters for different traffic scenarios and locations. The mathematical details of all these methods are briefly presented in the next section.

2.3 Methods description

2.3.1 Variance-based SA

The variance-based method based on the Sobol decomposition of variance is one of the most recent and effective global SA techniques (Saltelli et al., 2008 and Sobol, 1976). Given a model in the form $Y = f(Z_1, Z_2, \dots, Z_i, \dots, Z_k)$ two parameters are said to interact when their effect on Y cannot be expressed as a sum of their single effects. With Y being a scalar, a variance-based first order effect for a generic parameter Z_i can be written as:

$$V_{Z_i}\left[E_{Z_{\sim i}}(Y|Z_i)\right] \tag{1}$$

where Z_i is the i^{th} parameter and $Z_{\sim i}$ is the matrix of all parameters but Z_i. Furthermore it is known that the unconditional variance can be decomposed into main effect and residual

$$V[Y] = E_{Z_i}\left(V_{Z_{\sim i}}[Y|Z_i]\right) + V_{Z_i}\left[E_{Z_{\sim i}}(Y|Z_i)\right] \tag{2}$$

Equation (2) shows that for Z_i to be an important parameter we need $E_{Z_i}\left(V_{Z_{\sim i}}[Y|Z_i]\right)$ to be small, i.e. that the closer $V_{Z_i}\left[E_{Z_{\sim i}}(Y|Z_i)\right]$ to the unconditional variance $V[Y]$ the higher the influence of Z_i. Thus we may define our first order sensitivity index of Z_i with respect to Y as:

$$S_i = \frac{V_{Z_i}\left[E_{Z_{\sim i}}(Y|Z_i)\right]}{V[Y]} \tag{3}$$

Sensitivity indices as in (3) can be calculated per each parameter and per each parameters combination. This, however, would need a huge amount of model evaluations. In order to reduce the required efforts, a synthetic indicator to be coupled with the first order sensitivity index (the total effects index) may be defined as follows:

$$S_{T_i} = 1 - \frac{V_{Z_{\sim i}}\left[E_{Z_i}(Y|Z_{\sim i})\right]}{V[Y]} = \frac{E_{Z_{\sim i}}\left(V_{Z_i}[Y|Z_{\sim i}]\right)}{V[Y]} \tag{4}$$

Total effects index of the input parameter iprovides the sum of first and higher order effects (interactions) of the parameter Z_i. When the total index is S_{T_i}= 0 the i^{th} parameter can be fixed without affecting the outputs' variance. Since the analytical feasibility of traffic flow models limit the use of the formulas for the calculation of the variances in eq. 2, the application of this method can be effectively performed in a Monte Carlo (MC) setting (see Saltelli et al., 2008 for the details on the implementation). In the present paper we applied Sobol sequences of Quasi Random Numbers (Sobol,

1976) that ensure a quicker convergence in the numerical calculation of the sensitivity indices (and therefore we applied a Quasi Monte Carlo setting, QMC).

The total cost of the QMC implementation is $N \cdot (k + 2)$, much lower than the N^2 runs of a brute-force method. Since N is usually not lower than 1000, the number of evaluation required by this efficient approach is not negligible, especially for complex and expensive models. For this reason, variance-based SA is usually used on models whose computation time range from seconds to a few minutes. In any case they represent the most reliable way to evaluate the total sensitivity indices, necessary for deciding whether a specific input/parameter should be included or not in the model calibration.

2.3.2 Quasi-OTEE

The Elementary Effects (EE) method is one of the most common screening approaches when dealing with complex models (Morris, 1991). Consider the same model Y specified in the last section. If only the i^{th} parameter is changed by a certain value Δ, the new output will consequently be $Y_{\Delta i} = f(Z_1, Z_2, \dots, Z_i + \Delta, \dots, Z_k)$. The Elementary Effect of the i^{th} parameter, EE_i, is defined as:

$$EE_i = \frac{Y_{\Delta i} - Y}{\Delta} \tag{5}$$

Through randomly generating a number m of X points from the input space, and changing the i^{th} parameter by Δ each time, $Y_{\Delta i}$ is computed and the m *EE*s for the i^{th} parameter can be derived according to the above equation. The mean μ_{EE_i}, the standard deviation σ_{EE_i}, and the absolute mean $\mu^*_{EE_i}$ of these m *EE*s can accordingly be used to infer on the sensitivity of the i^{th} parameter as follows (Morris, 1991):

- If $\mu^*_{EE_i}$is low, then i is a negligible parameter;
- If $\mu^*_{EE_i}$ is high and $\sigma_{EE_i}$is low, i has linear and additive effects but no interactions with others;
- If $\mu^*_{EE_i}$and $\sigma_{EE_i}$are both high, i has non-linear effects and/or strong interactions with others;
- If μ_{EE_i} is low but $\mu^*_{EE_i}$is high, i will have oscillating effects depending on the value assumed by other parameters.

As the model needs to be evaluated twice for calculating each *EE*, the computational cost of the basic EE method is $2 \cdot m \cdot k$. However, some of these evaluations may be used for the computation of different EE_i. By sampling the parameter input space using trajectories, a lower experiment size of $m \cdot (k + 1)$ can be achieved (Morris, 1991).

Very recently, Ge and Menendez (2013) proposed the quasi-Optimized Trajectories EE (quasi-OTEE) approach, achieving similar results to the regular EE with just $(m - n + 1) \cdot (m + n)/2$ model evaluations. Variable n is the number in a sub-set of sampled trajectories that has to be defined a-priori. The definition of the best sub-set relies in an iterative optimization procedure based on finding the most n "spread" trajectories (having the largest Euclidean distance between them) from the original random trajectories m.

Elementary effect methods provide a reasonably cheap and reliable approximation of the total sensitivity indices. For this reason they can be used instead of variance-based techniques when the

computational cost of one model evaluation is higher than just a few minutes and when the number of parameters tends to be high (over 30-50 depending on the model non-linearity).

2.3.3 Kriging Metamodelling

The Kriging model may be viewed as an estimator based on the value of neighbouring points (Matheron, 1963 and Kleijnen, 2007). It first assumes that the output $Y(X)$ of a simulation model is given by:

$$Y(X) = \mu + \delta(X) \tag{6}$$

where *X* is the vector of model input variables, μ the simulation output averaged over the experimental area, and *δ(X)* a zero mean stationary covariance process. The Kriging model uses the following linear predictor $y(X)$of the output of a simulation model for a variable combination *X*:

$$y(X) = \lambda(X, D)^{\mathrm{T}} Y(D) \tag{7}$$

where *D* is the input variables design matrix of the simulation experiment for which the simulation output is known/simulated; and $\lambda(X, D)$ is a matrix of weights between the new variables specific combination *X* to be used as input in the metamodel and the points in matrix *D*. $\lambda(X, D)$ values are not constant but decrease as the distance between *X* and *D* increases. To select the optimal values λ^* for the weights $\lambda(X, D)$ one may use the Best Linear Unbiased Predictor which minimizes the Mean Squared Error of the predictor (Kleijnen, 2007):

$$\lambda^* = \mathbf{\Gamma}^{-1}\left[\gamma + \mathbf{1}\frac{\mathbf{1} - \mathbf{1}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\gamma}{\mathbf{1}^{\mathrm{T}} - \mathbf{\Gamma}^{-1}\mathbf{1}}\right] \tag{8}$$

where**1** is the *n*-dimensional identical vector (*n* is the number of the experiment variable combinations in *D*); $\mathbf{\Gamma} = \mathrm{cov}(Y_i, Y_j)$ with $i, j = 1, \ldots, n$ is the $n \times n$ symmetric and positive semi-definite matrix with the covariance of the simulated experiment outputs $Y(D)$; and γ the *n*-dimensional vector with the covariances between the *n* simulated outputs and the output for the variables' combination to be predicted by the metamodel. In simulation applications, the elements of γ and $\mathbf{\Gamma}$ are estimated using a correlation function which is the product of *k* one-dimensional functions (being *k* the number of variables or parameters of the simulation model) and assuming that these correlations are determined by the distance between the inputs of the specific outputs considered:

$$\mathrm{cov}(Y_i, Y_j) = \prod_{g=1}^{k} \mathrm{cov}(X_{i,g}, X_{j,g}) \tag{9}$$

where $g = 1, \ldots, k$. Furthermore the Kriging metamodel assumes a stationary covariance process, which implies that the covariance depends only on$|X_{i,g} - X_{j,g}|$. According to Kleijnen (2007) the Gaussian correlation function is a popular such function:

$$\mathrm{cov}(Y_i, Y_j) = \prod_{g=1}^{k} \exp\left[-\left(\frac{|X_{i,g} - X_{j,g}|}{\theta_g}\right)^2\right] \tag{10}$$

in which $\theta_g$is a parameter of the correlation function for the variable g, denoting the importance of the variable itself (the higherθ_g is, the lower is the effect due to the variable g). In order to find the best Kriging metamodel for a simulation model, it is therefore only necessary to estimate the k-dimensional vector of θ_g, using a Maximum Likelihood Estimator.

Ciuffo et al. (2013) have applied a variance-based method for SA to a Kriging approximation of a microscopic traffic simulation model. When the number of inputs is lower than 10-20 (depending on the model non-linearity) the accuracy of the Kriging emulator in reproducing the real model can be very high even with a relatively low number of model evaluations. In Ciuffo et al. (2013), 512 model evaluations were sufficient to estimate a reliable Kriging approximation of the Aimsun mesoscopic model (TSS,2012) with 7 input parameters.

2.3.4 Weighted Simultaneous Perturbation Stochastic Approximation (WSPSA)

The WSPSA follows the classical aggregated calibration philosophy, where the differences between observed measurements and the equivalent simulated outputs are minimized using an optimization function:

$$minimize\ z(\theta) = k_1\|y(X) - Y(Z,X)\| + k_2\|Z - Z^a\| \tag{11}$$

where $y(X)$ and $Y(Z,X)$ are vectors of observed measurements and corresponding simulated measurements, Z^a are prior values of the parameters to be calibrated, Y is the model that generates simulated measurements, and k_1 and k_2 are weights depending on the relative confidence on observed measurements and different sets of prior parameter values.

The original simultaneous perturbation stochastic approximation (SPSA) efficiently approximates the gradient function with only two successive measurements of the objective function (independently of the number of parameters) and therefore significantly saves computational time for large-scale problems over traditional gradient methods such as the finite-differences stochastic approximation (Spall, 1992). The general stochastic approximation algorithm starts from an initial estimation of the parameter vector and iteratively traces a sequence of parameter estimates that converge of the objective function's gradient to zero:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k\hat{g}_k(\hat{\theta}_k) \tag{12}$$

where $\hat{\theta}_k$ is the estimate of the decision vector in the k^{th} iteration of the algorithm, $\hat{g}_k$ is the estimated gradient, and a_k is an algorithm parameter that gets smaller as k becomes larger. In SPSA the approximation of the gradient $\hat{g}_k$ depends on two function evaluations from a simultaneous perturbation of the parameters:

$$\hat{g}_k(\hat{\theta}_k) = \frac{z(\hat{\theta}_k + c_k \otimes \Delta_k) - z(\hat{\theta}_k - c_k \otimes \Delta_k)}{2c_{ki}\Delta_{ki}} \tag{13}$$

where z is the optimizing function, $\hat{g}_k(\hat{\theta}_k)$ is the i^{th} element of the gradient vector, Δ_k is a random perturbation vector, generated through a Bernoulli process with values of +1 and -1 with equal probabilities,$\otimes$ is the component-wise multiplication operator, and $c_k$is an algorithm parameter that determines the amplitude of the perturbation:

$$c_{ki} = \frac{c_i}{(k+1)^\gamma} \tag{14}$$

where c_i is the i^{th} element in an algorithm constant parameter vector and $\gamma$is a constant parameter.

Very recently Lu et al. (2013) proposed the Weighted Simultaneous Perturbation Stochastic Approximation by incorporating known spatial and temporal correlation between parameters and measurements to minimize the noise generated by uncorrelated measurements, improving

significantly the performance of the generic SPSA formulation. By introducing a matrix *W* in the generic optimizing function *z*, both spatial and temporal correlations between each parameter and the model outputs may be considered. Each element of *W* (with size $n \times m$) is the relative correlation between the n^{th} model parameter in the m^{th} measurement. Instead of calculating the i^{th} element in the gradient vector $\hat{g}_k(\hat{\theta}_k)$ using the objective function, a weighted sum of the measurement error changes related to the i^{th} parameter is used:

$$\hat{g}_k(\hat{\theta}_k) = \frac{z(\hat{\theta}_k + c_k \otimes \Delta_k) - z(\hat{\theta}_k - c_k \otimes \Delta_k)}{2c_{ki}\Delta_{ki}} W_i \tag{15}$$

where W_i is the ith line in the weight matrix *W*. The output of the evaluation function *z* results now in a vector with length equal to the number of parameters to be calibrated, rather than a scalar:

$$z(\theta) = \begin{bmatrix} k_1[y(X) - Y(Z,X)][y(X) - Y(Z,X)]^{\mathrm{T}} \\ k_2[Z - Z^a][Z - Z^a]^{\mathrm{T}} \end{bmatrix} \tag{16}$$

The way to calculate weight matrices for WSPSA depends on the configuration of the case study, the parameters considered and the available measurements. For example, assignment weights regarding each sensor can be used as weights, for the calibration of origin-destination measurements using loop sensor counts.

The WSPSA method is an extremely efficient optimization method, reaching fast convergence even when dealing with large parameter set. As it outputs a single solution, its use is especially suited for case-studies with a large number of needed calibrations. For further details on the WSPSA and SPSA approaches, the reader may refer to Lima Azevedo (2014) and Spall (1992).

3 The case-study application

The proposed methodology was applied to the calibration of an advanced microscopic simulation model for safety analysis. An urban motorway (A44) near Porto, Portugal, was selected as case study due to several safety related issues: dense traffic, unusually high number of lane changes due to frequent route-choice decision making, short spacing between interchanges and high percentage of heavy goods vehicles. A44 is a 3,940m long dual carriageway urban motorway with 5 major interchanges, two 3.50m wide lanes and 2.00m wide shoulders in each direction (see Interactive Map for the simulated network). Acceleration and deceleration lanes are added to the main carriageway section at all interchanges, although often as short as 150m. On and off-ramps connect to local roads, which generally have tight horizontal curves, intersections or pedestrian crossings, features that tend to impose significant reductions in vehicle speeds.

To test the proposed calibration framework, we focused our attention in the replication of detailed traffic characteristics in accident scenarios, and test whether the simulation outputs would differ much from the non-accident ones. If such variability is replicated by the simulation model its use in specific safety studies is therefore validated. Data on 144 accidents occurring between 2007 and 2009 were collected; the simulator was calibrated separately for each of these events (see Figure 1). A total of 6,400 non-accident scenarios were also randomly selected from the same time period, and used for comparison. Aggregated (loop sensor) data was collected for each of the sampled events and detailed vehicle trajectories were collected on-site for the entire length of the motorway during a generic day.

Finally, we also show the importance of trajectory data in driving behaviour model calibration when the replication of detailed traffic variables is at stake. It is common design (and even research) practice to use simulators calibrated with aggregated data to extract detailed simulated information from the transportation system under analysis. An example of such practice is the analysis of (acceleration-based) fuel consumption outputs or surrogate safety measures with a simulator calibrated with just loop sensor counts or average speeds. Different combinations of accelerations and headway parameters may be compatible with a good calibration performance under loop speed or/and count based optimization. In this paper, the proposed detailed calibration method is also applied for comparing the use of different input data sets: loop sensor data and trajectory data.

<Interactive Map: kml map file attached>

<https://mapsengine.google.com/map/edit?mid=zUbJxE1-0Ukg.kHTC2zNSqI0U>

3.1 The micro-simulation model

The integrated driver behaviour model presented in (Toledo et al., 2007) is of particular interest due to the high interaction of all advanced models describing driving behaviour and the high number of parameters (101). It integrates four levels of tactical decision-making: target lane, gap acceptance, target gap and acceleration, in a latent decision framework based on the concepts of short-term goal and short-term plan. Furthermore, probabilistic frameworks are used to capture drivers' route choice decisions and to assign driving behaviour parameters and vehicle characteristics. This model has been integrated in MITSIMLab, an open-source traffic simulation platform developed in C++ (Yang et al., 1999). It can output simulated vehicles' location, speed and acceleration at 1 second resolution allowing the computation of further detailed traffic variables. Higher resolutions can be reached by changing the simulator open-source code. As a large number of simulations are required for the present case-study, this resolution was kept as such, but the output format was changed to a lighter structure (Lima Azevedo, 2014). MITSIMLabs' driving behaviour model formulation has been enhanced over the years by introducing several complex structures associated with detailed behaviour and behavioural interactions. For the description of the 101 parameters and the model formulation itself, please refer to Ciuffo and Lima Azevedo (2014), Toledo (2007) and Yang et al. (1999).

3.2 The data

Three different traffic data sets were specifically collected for the present study:

- A dynamic seed origin-destination (OD) matrix estimated using the generalized least squares (GLS) simultaneous estimator (Cascetta et al., 1993) and a sample of license plate matching records and vehicle counts (Lima Azevedo, 2014);
- 5 min loop sensor speeds and counts for the existing eight traffic stations, during the three years analysis period direction (see Interactive Map). Erroneous records were filtered out using the daily statistics algorithm (DSA) proposed by Chen et al. (2003). The details on the application of the DSA to the present case study can be found in Lima Azevedo (2014);
- 1855 partial vehicle trajectories collected for a generic morning by aerial remote sensing for the entire length and access links of the A44 motorway and for congested and non-congested periods. A Cessna T210L Centurion II aircraft equipped with high-resolution photographic

equipment fixed to a gyro-stabilizing platform overflew the A44, between 8:45 and 10:45 AM on the 11th of October 2011. Images were recorded at an average rate of 0.5Hz and were then orthorrectified using a 3D terrain model, the camera and lens characteristics, and the precise flight positioning data recorded through differential GPS. The vehicle detection was carried out using coloured background subtraction and its trajectory was then reconstructed using an extension of the k-shortest disjoint path algorithm using motion-based optimization. For further details on the trajectory extraction method and its validation the reader is referred to Lima Azevedo et al. (2014).

3.3 The calibration design and set up

For the SA regarding aggregated loop sensor data, eleven different measures of performance (MoP) were considered to assess the impact of using different combinations of traffic stations and output data (speed vs. counts). MoP computed on each single detector (8), on all the detectors of each road direction (2) and on all the detectors of the network (1) were considered for both count and speed data. Furthermore, to assess the dependence from the selected GoF, three different statistics were used for comparison: the root mean squared error (RMSE), the root mean squared percentage error (RMSPE) and the Theil's inequality coefficient U, (please refer to Hollander and Liu, 2008, for their formulation and description). Therefore, a total of 66 different GoF-MoP combinations were considered in the identification of the most relevant parameters regarding the loop sensor data.

Similarly, for the trajectory-based SA and calibration, a set of statistics for the simulated trajectories were extracted and compared against its real counterpart collected on-site. A set of 9 MoP were selected for describing the trajectory data: speed, acceleration, deceleration, headway, time-to-collision (TTC), deceleration rate to avoid crash (DRAC), number of lane-changes (NLC), and lead and lag gaps during a lane-change. For each of these variables (except for NLC) 11 statistics were considered in the computation of its GoF, characterising their distribution within the A44 and its access links: the minimum value, nine percentiles (10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th) and the maximum value of the distribution. Again, in order to assess the dependence from the GoF measure selected, the RMSE, RMSPE and U were computed, resulting in a total of 27 GoF-MoP combinations to be analysed during the SA.

The sizes for the variance-based SA QMC experiment, the trajectory sampling of the quasi-OTEE and the Kriging experiment size are defined in the next section.

For the calibration of each specific event using the WSPSA algorithm, the average speed and sensor counts for the 30 min before the accident occurrence were used as measurements in the optimizing function. The full trajectory-based calibrated parameters and the GLS-estimated dynamic seed OD were used as initial parameters. The assumed weights of the optimizing function (eq. 11) were $k_1^{counts} = 0.3$, $k_1^{speeds} = 0.5$ and $k_2 = 0.2$. These values were defined previously, assuming the contribution of each data source to the calibration process. As we focus on detailed traffic statistics a higher contribution was given to speed related data. A sensitivity analysis on these weight values may, however, enhance the calibration final results. Finally, the constant parameters of the WSPSA algorithm were set to previously estimated values under a generic SPSA application to MITSIMLab calibration (Vaze et al., 2009).

For carrying out this computationally demanding task, MITSIMLab was installed under Scientific Linux in a cluster with 80 cores with 1GB of RAM memory each. This resource allowed for an extended parallelization of the required number of simulations.

4 Results

4.1 The effect of calibration data on simulated safety outputs

4.1.1 Calibration using aggregated data

In early research (Ciuffo and Lima Azevedo, 2014) the proposed methodology was applied to the identification of sensitive parameters regarding loop sensor data, using the current case study and for the specific day of trajectory extraction. The multi-step SA carried out was a two step procedure.

A first group variance-based SA (N = 2048) was performed on 15 groups accounting for the entire 101 parameter set: the reaction time distribution parameters (group *1*), the car-following acceleration model (*2*), the free-flow acceleration model (*3*), the merging model (*4*), the mandatory lane change model (*5*), the yielding model (*6*), the nosing model (*7*), the nosing control parameters (*8*), the courtesy yielding model (*9*), the driver heterogeneity parameters (*10*), the target gap acceleration model (*11*), the gap acceptance model (*12*), the lane utility functions (*13*), the target gap model (*14*), and the origin-destination variability parameters (*15*).

A second full variance-based SA (N = 512) was then carried out with the 41 most sensitive parameters. This second step resulted in the identification of the final 11 most sensitive parameters, which accounted for almost the 90% of the output's variance:

- μ_{RT} and μ_{dv}^{h} are the mean of the reaction time and headway threshold distributions respectively. μ_{DS} is the distribution mean of the desired speed factor. These are known to be important parameters, especially when analysing individual models separately;
- α_{cf}^{acc} and α_{cf}^{dec} are the constant parameters of the CF acceleration and deceleration models;
- β_{cf}^{acc} is the speed parameter in the CF acceleration model, and it emerged as sensitive mostly when analysing non-congested situations;
- γ_{cf}^{dec} and ρ_{cf}^{dec} are the gap and speed differences between the subject and the leader vehicles of the CF deceleration model. Although γ_{cf}^{dec} was already found as significant in previous SA of MITSIMLab (Kurian, 2000), it is clear that it is closely linked to the speed difference and both parameters should be taken jointly into account;
- α^{CL}, δ_{lum}^{1} and θ_{MLC} are parameters of the lane selection (utility) model: the constant parameter of the current lane utility, the one-lane-change-required to exit parameter, and the distance-to-exit parameter. The latter two parameters are used together (combined) in the lane utility function formulation. The two lane carriageway layout of the A44 motorway clearly conditioned this outcome, as the network configuration almost only requires for one mandatory lane-change throughout its entire extension.

The reader is referred to Toledo et al. (2007) and Lima Azevedo (2014) for a description of the mathematical formulations of the models where these parameters are used.

The methodology required a total of 55,808 model evaluations instead of the 421,888 (-80%) otherwise required for applying variance based techniques to the whole set of parameters.

The sizes of the QMC experiments were found to be sufficient as all GoF converged. For stochasticity control, three replications of each combination were considered during this analysis. This number was conditioned by computational resources. The results were, in general, the same for all GoF. However, a clearly slower convergence of the QMC results was obtained for the RMSE. At each step, the less sensitive parameters were set to the best speed-based U combination accounting for all sensors.

Finally, the most sensitive 11 parameters found were integrated into a final Kriging calibration, in which a set of 13,312 combinations ($1024 \times (11 + 2)$) were computed, each with 10 replications for stochasticity control (see Ciuffo and Lima Azevedo, 2014 for details).

4.1.2 Calibration using disaggregated data

Similarly to the previous section, the first step of the application of the proposed method to the trajectory data was a group variance-based SA performed on the same 15 groups (101 parameters). A total of 34,816 non-replicated simulations (assuming $N = 2048$) were carried out to compute the group sensitivity indices based of the 27 GoF mentioned earlier.

The resulting number of parameters in the most sensitive groups (56) is substantially higher than in the previous aggregated data based SA (41). Furthermore, the size of the QMC experiment may still reach a very high number, due to the multiple nature of the MoP considered for the trajectory-based SA. Also, the initial grouping has a consistent structure based on the sub-models of MITSIMLab, and a different grouping might be counter-intuitive. Therefore, the quasi-OTEE method was used as a second step to further identify most influential inputs. This also generalises the proposed multi-step SA to other techniques.

For this second step, a set of $m = 500$ trajectories and $n = 100$ quasi-OTEE were selected. 56,000 simulations would be necessary for the basic EE method (total number of parameters $k = 56$), whereas 5,700 are needed for the quasi-OTEE method. The less sensitive 46 parameters values were set to the combination with the best combined U from the previous step (combination of all 11 MoP considered for trajectories). The quasi-OTEE allowed to quickly identify the 15 most sensitive parameters, although without providing a quantification of the related uncertainty. With these 15 parameters a full final variance-based SA is now computationally feasible.

We started with a smaller size for the QMC experiment of $N = 256$, thus with 4,352 model evaluations. If convergence had not been achieved, additional simulations would have been necessary. Again, the other 41 parameters values were set to the values of the best overall GoF combination of the quasi-OTEE analysis. As we are collecting individual vehicle observations and the stochasticity was assumed to be captured by the large number of vehicle position observations computed, no replications of each combination were performed at this point. Convergence was reached and this final variance-based SA allowed for the identification of nine model parameters (out of 101) that accounted for, at least, 50% of the output's variance of each computed trajectory-based MoP:

- μ_{RT} and σ_{RT} are the mean and standard deviation of the reaction time distribution;
- μ_{DS} and σ_{DS} are the mean and standard deviation of the alternative desired speed distribution;

- α_{cf}^{dec} and ρ_{cf}^{dec} are the constant parameter and speed difference between the subject and the leader vehicles parameter of the car-following deceleration model;
- β_{tail}^{TL}, δ_{lum}^{1} and θ_{MLC} are the tailgating, one lane-change required to stay in path, and the distance to exit parameters of the lane selection model.

The interaction component is still very important for all MoP and is responsible for the increased share in the output variance for the selected nine parameters. With this methodology 44,868 model evaluations were performed, instead of 212,992 (-79%) required with the full variance based method.

Finally, these 9 most sensitive parameters were calibrated using the Kriging approach. For the experiment design, a set of 11,264 combinations (1024 × (9 + 2)) with 10 replications each were computed.

4.1.3 Comparison and Discussion

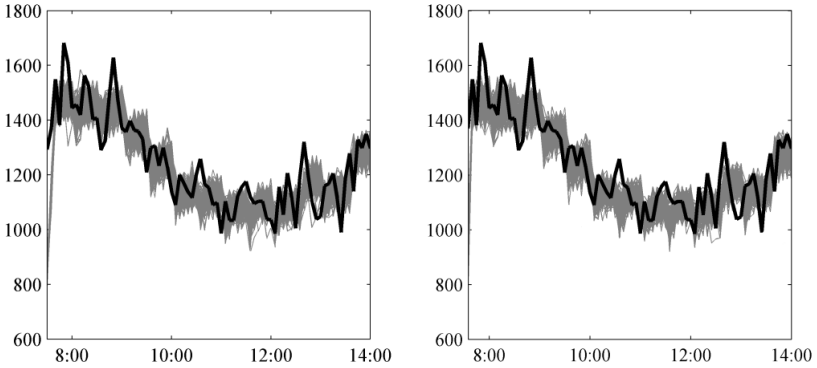
As expected, the proposed global SA framework using the distributions of trajectory statistics resulted in a more complex procedure than using data from a small set of loop sensors.

The majority of the relevant parameters from the SA with loop-based data was also detected as sensitive in the last steps of the trajectory-based SA. This vouches the consistency of both the proposed global SA and the MITSIMLab driving behaviour model itself. The reaction time, desired speed and car-following (CF) deceleration constant parameters, and the lane selection (utility) model parameters were revealed as fundamental simulation models. However, some other parameters are also important in the replication of trajectory statistics: the standard deviations of relevant driving behaviour heterogeneity distribution parameters, namely those for reaction time and the desired speed; interaction parameters (speed difference and density) of the CF acceleration model; and even parameters of the nosing and courtesy yielding models. The identification of calibration parameters is very sensitive to each case study configuration and observed traffic conditions; in fact, these two models were expected to be important models in the busy A44 case study.

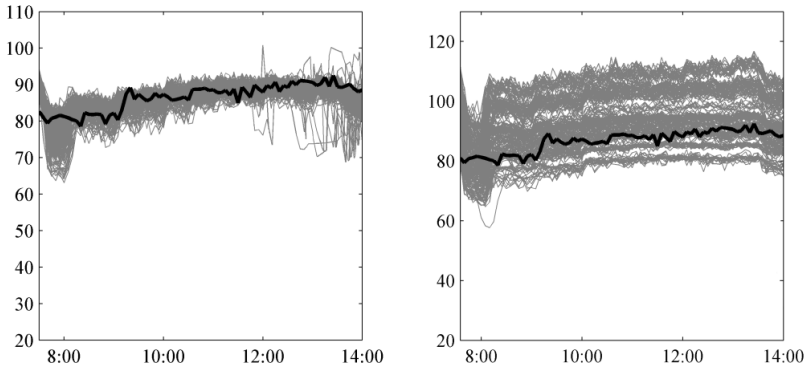
After the two final calibrations, setting a single best solution for replication is not advisable for several reasons: the Kriging approximation might not capture small changes existing in the true model; a single best option may easily change, depending on the daily traffic data; and the best combination for a specific MoP is most likely different from the best one for another measure. For these reasons the thirty best combinations with comparable performances were kept for the validation testing. This number was selected by rounding the number of combinations with a (speed-based / combined) GoF measure (for aggregated / trajectory data) with U <0.085.

From Figure 2 it is clear that both trajectory and loop based calibrations allowed for good replication of sensor outputs. However, a clear improvement in the overall detailed variables distributions is observed for the trajectory-based calibration (see Figure 3). This improvement is even more significant when analysing road section types that do not have loop sensors anywhere in the A44 layout, and therefore no sensor-based speed measurements (Figure 4). A perfect fit is not reached due to the limited calibration iterations, the intrinsic modelling errors and the choice of a combined GoF accounting for an overall (and not variable specific) MoP as optimizing function.

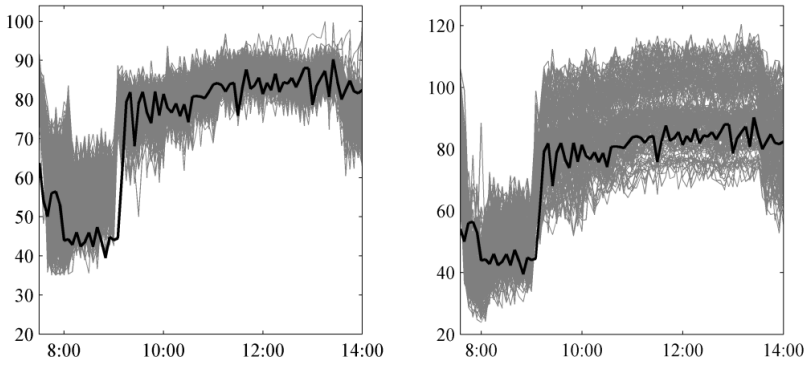
Besides the clear dependence between the calibrated parameter values and the chosen MoP, these results show the importance of using trajectory data in the calibration of driving behaviour models when the replication of detail variables is at stake. Yet, when the calibration process aims at reaching a less comprehensive model (i.e., only replicating generic aggregated network efficiency measurements) trajectory data might not bring significant improvements, especially when the driving behaviour model is robust and the sensor coverage is comprehensive and well distributed (covering merging, weaving areas, etc).



a) Full network average traffic flow (veh./h/lane) by time of day

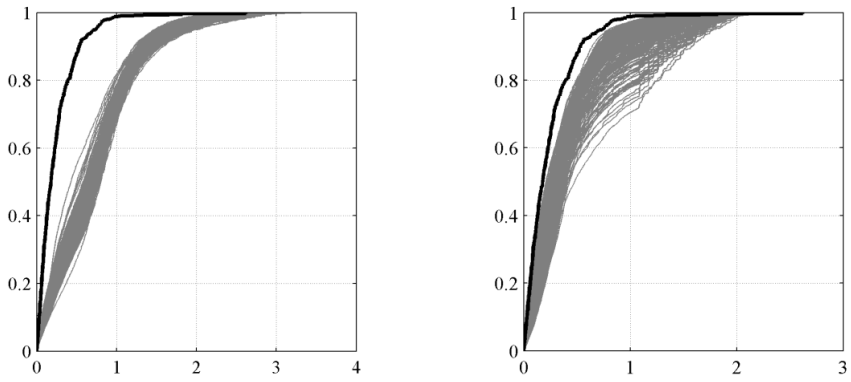


b) Full network average speed (km/h) by time of day

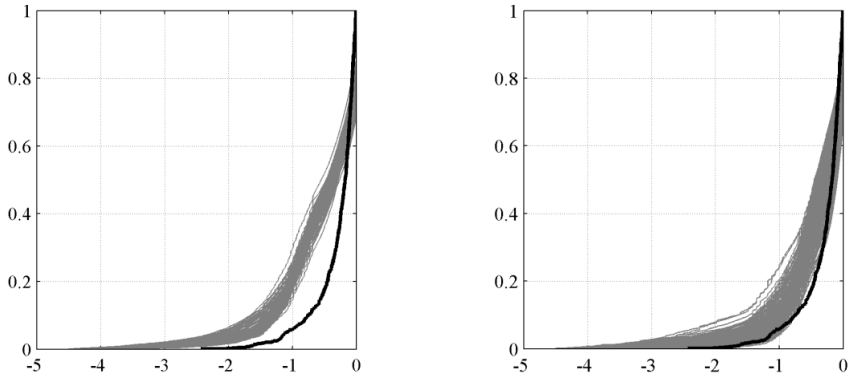


c) Loop sensor #401 average speed (km/h) by time of day

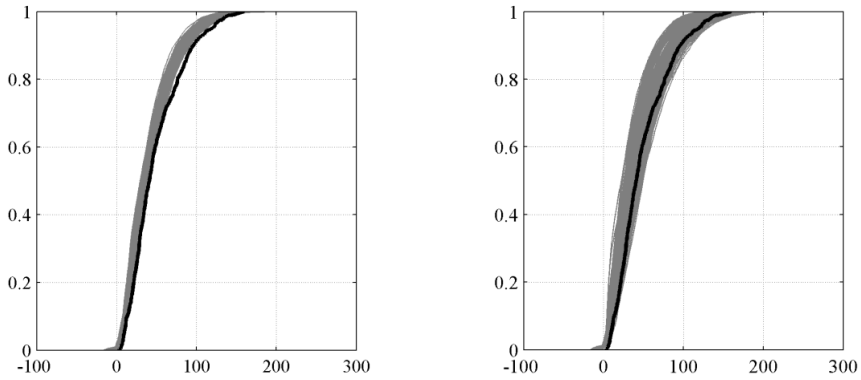
Figure 2 – Loop sensor output from sensor-based (left) and trajectory-based (right) calibrations (grey) and its real counterparts collected on-site (black).



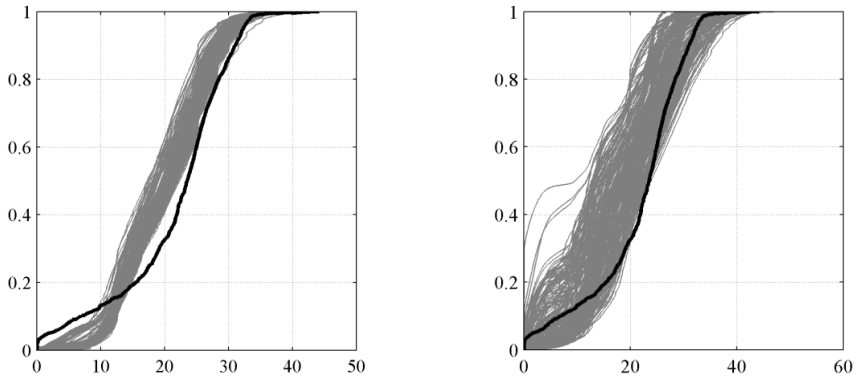
a) Acceleration CDF for the entire network (m/s^2)



b) Deceleration CDF for the entire network (m/s^2)

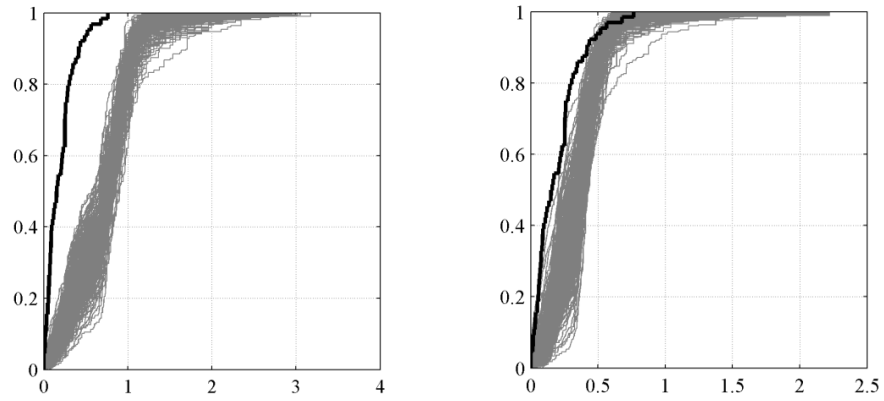


c) Headway CDF for the entire network (m)

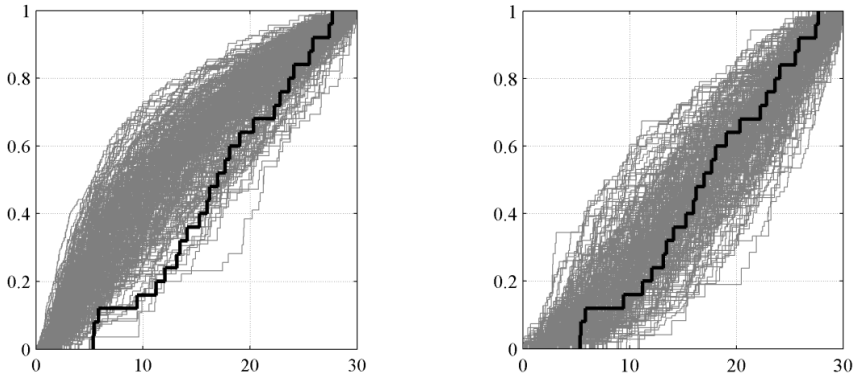


d) Speed CDF for the entire network (m/s)

Figure 3 – Trajectory output from sensor-based (left) and trajectory-based (right) calibrations (grey) and its real counterparts collected on-site (black).



a) Acceleration CDF in the right lane of two-lane sections, grade between 0 and 2%, and low traffic conditions (m/s^2)



b) TTC CDF for two lane sections in the left lane of two-lane sections, grade lower than -2%, and low traffic conditions (s)

Figure 4 – Trajectory output from sensor (left) and trajectory-based (right) calibrations (grey) and its real counterparts (black).

4.2 The effect of accident data-based calibration on driving behaviour parameters

In this section the focus is on the calibration of MITSIMLab to replicate detailed variables for a large set of different scenarios (time periods and network locations). As it was concluded from the previous sections, the selection of a set of "best" combinations and a high number of replications is always preferable when dealing with stochastic simulation applications. This however, may cause a significant increase in the size of the simulation scenarios set. In fact, for each simulation scenario, the simplest Kriging metamodel may need thousands of replications for the selection of its best set of combinations. To overcome this computational burden, the WSPSA described above was used for the event-specific calibration. This type of simultaneous demand-supply method views the calibration process as an optimization problem reaching a unique solution, rather than controlling data variability using multiple combinations.

When applying the above method to this case study, the parameter set to be calibrated is composed of the dynamic OD pairs of interest and the selected 11 most sensitive driving behaviour parameters regarding aggregated measurements (from section 5.1.1), as it is the only available data for event-specific calibration.

The dynamic OD has a total of 100 OD paths per each 30 min period. As no significant intra-variability was found in almost all intervals, a total of 100 demand parameters may be considered for each 30 min period. The rest of the parameters were set to their best values for the trajectory-based calibration (see section 4.1.2). Regarding the specification of the weight matrix *W* of the WSPSA, each matrix entry was defined as the relative correlation between the flow of a specific OD pair in a specific period p_1 (parameter to be calibrated) and the count of sensor (measurement) in period p_2. Due to the small size of our case study, these correlations were calculated using simple static assignment proportions directly computed from the network configuration. As all vehicles departing at time *p* reach their destination at *p + 1* at most, the static simplification is acceptable. Driving behaviour parameter weights in *W* were set to 1 as no distinction was made between individual effects on different loop sensor output. Using just 30 iterations of this WSPSA set up, the RMSNE was reduced on average by 38%.

4.2.1 Comparison of parameters values

In Figure 5 the probability density estimates of the calibrated 11 parameters for accident (dashed area) and non-accident (light grey area) events are presented. These estimates are based on a normal kernel function, using a window parameter based on 100 equally spaced points that cover the range of each parameter. In solid and dashed thin vertical bars are the default parameters values estimated in previous research efforts by Ahmed (1999) and Toledo et al. (2007), respectively. It is worth remembering that Ahmed (1999) estimated an independent formulation of the lane-changing and acceleration models, and thus, not estimating some of the parameters considered in the analysis. Also, it is crucial to understand that there are interactions between these parameters and the analysis of the variability of a single parameter should be carefully done.

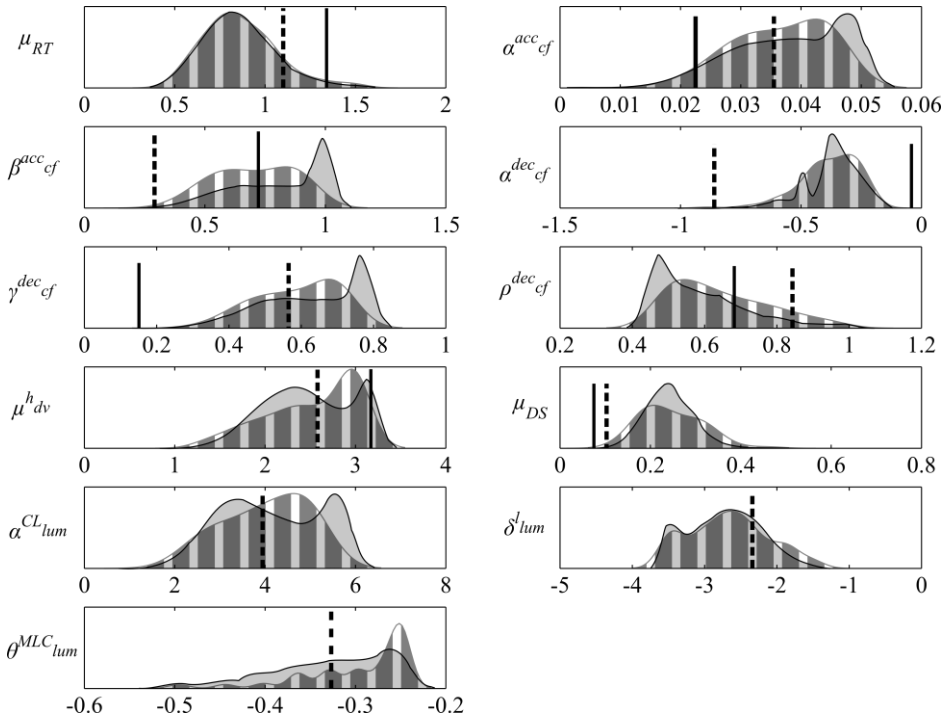


Figure 5 – Distribution of calibrated parameters for accident (dashed area) and non-accident (light grey area) events.

It is clear from Figure 5 that the distribution of μ_{RT}, μ_{DS},ρ_{cf}^{dec} and θ_{MLC}show substantial differences from past estimates. Yet, the estimated mean reaction time (μ_{RT}) for both accident and non-accident events remained close to the value estimated using the real trajectory data set and to typical low safety thresholds values found in the literature. Also, the estimated desired speed add-on (regarding the speed limit) parameter, μ_{DS}, will obviously result in higher free flow speed values.

No significant differences in the distribution of μ_{RT} and μ_{DS} were observed between accident and non-accident calibrations. Lower values for α_{cf}^{acc} and β_{cf}^{acc} were observed for the calibrated accident events, generally resulting in lower CF acceleration rates under the same conditions. The lower γ_{cf}^{dec} and higher ρ_{cf}^{dec} for calibrated accident events, result in higher deceleration rates for these conditions: a lower γ_{cf}^{dec} parameter, the headway parameter for the CF model, results in deceleration rates more sensitive to the headway distance to the front vehicle; ρ_{cf}^{dec} is the speed difference deceleration parameter for the CF model and its higher value results in deceleration rates more sensitive to the speed difference between the follower and the leader vehicles. The higher headway threshold mean μ_{dv}^{h} for accident events represents a broader control of CF model over the free flow acceleration model, i.e. a vehicle is under the influence of the front vehicle stimulus for larger headways. Finally, the θ_{MLC} distribution shows the importance of the distance to the desired exit in the lane change decision (the greater is θ_{MLC} the more significant is the effect of the distance to exit to the lane utility value). For calibrated accident events, higher parameter values result in an expected higher number of lane changes for shorter distances to exit.

Even if the estimates of straightforward safety influencing parameters such as the reaction time or the desired speed do not have significant differences for both the accident and the non-accident samples, their combination with other parameters may still be related with unsafe events. The desired speed parameter (as it is specified in MITSIMLab, i.e. only for free flow conditions) for example is not, in fact, a primal explanatory factor in the occurrence of the rear-end collisions or side collisions under dense traffic scenarios as observed in the A44. The complexity of the underlying mechanisms of the relationship between the chosen driving behaviour model and unsafe events is thus exposed.

4.2.2 Detailed traffic data

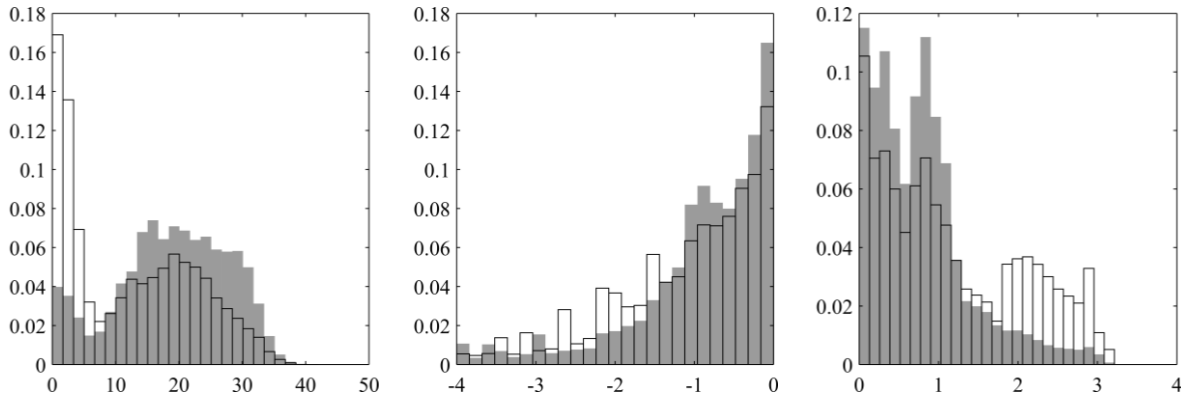
After the calibration of each of the considered accident events and sampled non-accident events, we are now able to generate artificial data for each of them, and check of the simulator ability to replicate more dangerous conditions. Artificial vehicle trajectories simulated for the location and time of each event were used to generate the detailed traffic variables of interest regarding safety (see section 4.3). For the accident occurrences, the 144 simulations resulted in an average of about 1.5×10^5observations of vehicle motion variables at a frequency of 1Hz. These observations were recorded for the 50 m section upstream the accident location, and within the 5 min period before its occurrence. The 6,400 sampled non-accident events resulted in about 4.5×10^6 observations for the same spatial and temporal units. A high number of replications is always desirable when working with simulated data. However, when dealing with the above mentioned number of observations per simulated scenario, the total number of trajectory records for multiple replications quickly becomes unmanageable. Thus, due to computational limitations, only three replications of each event were

performed. Table 1 contains summary statistics of several variables for both accident and non-accident events. The distributions of speed, acceleration and headway are shown in Figure 6.

Variable		Mean	Std. dev.	Median
Speed (m/s)	Accidents	12.51	10.00	12.19
	Non-accidents	18.97	8.78	19.50
	Real trajectories	21.73	7.38	22.84
Acceleration (m/s^2)	Accident	1.17	0.89	0.93
	Non-accident	0.79	0.61	0.71
	Real trajectories	0.66	0.87	0.30
Deceleration (m/s^2)	Accident	-1.10	0.92	-0.87
	Non-accident	-0.92	0.86	-0.72
	Real trajectories	-0.85	0.90	-0.44
Headway (m)	Accident	21.83	29.87	6.80
	Non-accident	38.23	34.57	29.50
	Real trajectories	45.84	33.90	35.86
Lead side gap before a lane change (m)	Accident	4.49	6.95	1.90
	Non-accident	9.68	10.91	4.5
	Real trajectories	11.90	8.87	10.53
Lag side gap before a lane change (m)	Accident	3.68	5.37	1.87
	Non-accident	10.19	8.71	8.56
	Real trajectories	12.46	8.99	11.77

Table 1 –Statistics of variables related to artificial trajectories for the 5 min before accident and non-accident events and to real trajectories collected on-site.

The average speed is lower for the accident events than for the non-accident events. This suggests that some accidents took place at lower speed sections (such as entry and exit ramps) or under more dense traffic scenarios. However, this lower average speed does not mean that the drivers have made adequate speed choices. In fact, the speed standard deviation, often used as a surrogate indicator for multiple vehicle crashes, is higher. Both acceleration mean and standard deviation are significantly higher for simulated accidents than for non-accident events. Similarly, deceleration values are more conservative for non-accident events. A possible explanation is the presence of denser traffic conditions for the simulated accident events. Front relative speeds are defined as the speed of the front vehicle minus the speed of the subject vehicle, under car-following situations. Their distributions do not differ much for both samples. However, the headway values for the accident events are much smaller than for the non-accident.



a) speed (m/s) b) deceleration (m/s^2) c) acceleration (m/s^2)

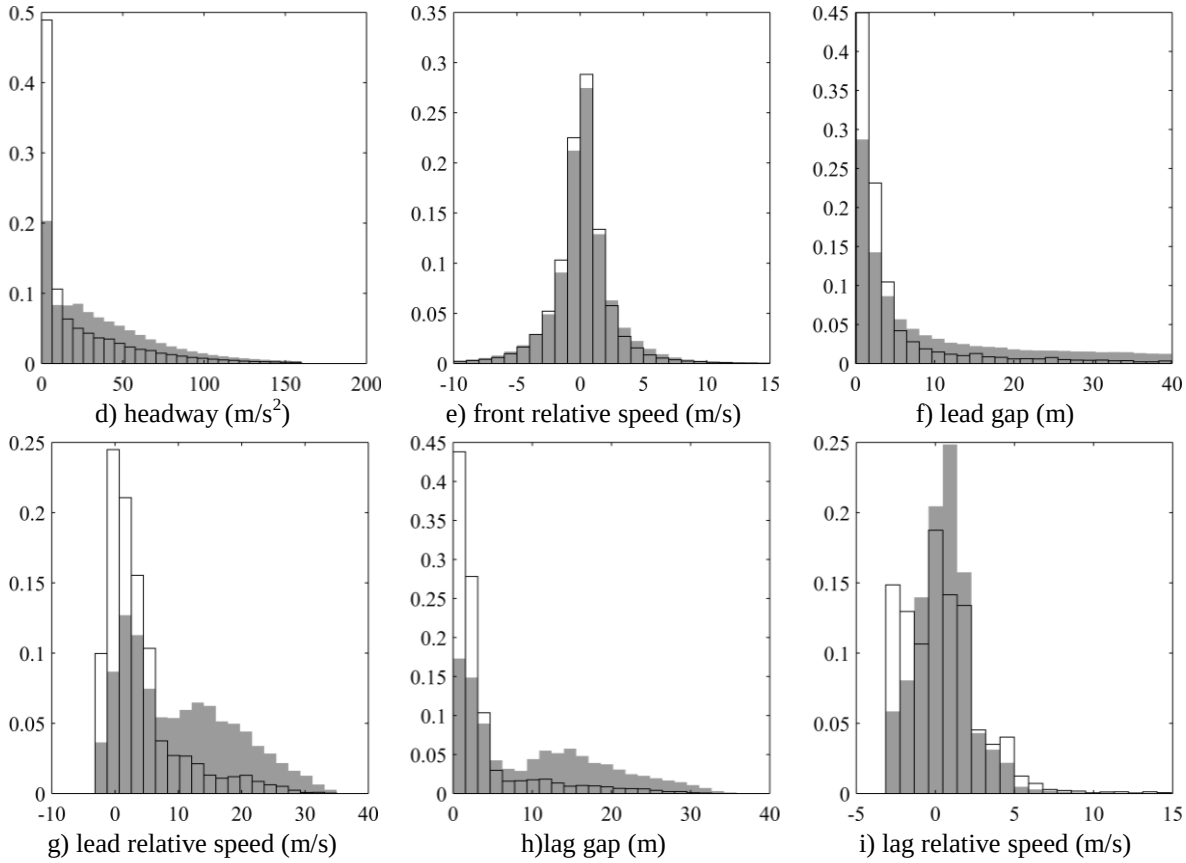


Figure 6 –Distribution of simulated variables for accident (white) and non-accident (grey) events

The relations between the subject and the lead and lag vehicles in the target lane affect the gap acceptance and gap choice behaviours and, therefore, lane change conflicts. The statistics of the lead and lag gaps (for both left and right lanes) and relative speeds were only computed when a driver wished to switch lanes. The average values for both lead and lag gaps for accident events are much smaller than the non-accident ones and than those found in previous studies (Toledo et al. 2007, Choudhury, 2007).

5 Conclusions and Future Research

In this paper, the entire problem of calibration is treated under the light of uncertainty management. An application-ready global SA based calibration is proposed, estimating a large set of parameters and using a multi-step approach. Different SA techniques can be chosen depending on the complexity of each step, allowing for a significant reduction of the number of simulations and for the use of detailed analysis to more sensitive parameters set. The proposed framework clearly brings advanced methods to the global calibration of complex traffic microscopic simulation tools and was successfully applied to a specific motorway safety study using a driver behaviour model with 101 parameters.

The proposed method was coupled with a metamodel based calibration to test the importance of trajectory data in the replication of detailed traffic variables, typically used in safety analysis. Poorly adjusted trajectory outputs deteriorate the usefulness of simulated traffic conflict surrogates, which as a result tend to reflect just the influence of traffic on crash frequencies. In fact, similar phenomena

were detected in early traffic conflict definitions (Hauer and Gårder, 1986). Despite the unquestionable value of trajectory data, its collection is still burdensome. The proposed method can also be used to test the simulators performance after calibration using other improved data sets, such as section-specific distributions of speeds, accelerations or headways at the lane level (more frequently available through video-based traffic monitoring systems). Yet, if detailed traffic and safety analysis are at stake, practitioners and researchers should always seek for such improved data to control uncertainty. The present document clearly show the importance of this uncertainty control in the calibration of simulators for the replication of safety outputs, both at the input data and at the methodological framework levels.

The method was also coupled with a powerful simultaneous demand-supply calibration method and successfully applied in the calibration of a large set of accident and non-accident scenarios. The different values of the calibrated parameters allowed identifying diverse simulated behaviours in accident and non-accident scenarios, relying also on the interpretation of the intrinsic nature of each parameter itself. Furthermore, the artificial data generated by the differently calibrated models showed a clear divergence in the simulated outputs typically used for safety assessment, corroborating the usefulness of advanced traffic microscopic simulation tools in the replication of detailed interactions and driving behaviour.

It is worth mentioning that the calibration tests were carried out separately considering accident and non-accident events. An accident type specific calibration should be tested in future research as the identification of different types of detailed behaviour (and, therefore, model parameters) might be of interest. Furthermore, to fully assess the efficiency of the aggregated calibration logic using disaggregated data presented in this paper, a comparison with other estimation frameworks needs to be carried out, namely: the direct and the conditioned estimations methods, both using maximum likelihood or Bayesian techniques directly on space-time observations, and outside of the simulation tool; and with the disaggregated calibration, where an optimizing function is specified in terms of space-time observations. The simulation tool must then be configured to match each real initial trajectory and the simulation positions be compared with the real ones. All these enhancements should be carried out alongside the development of more advanced behavioural models, which will allow the replication of near-accident driving behaviour.

Acknowledgements

The research contained within this paper benefited from the participation in the European Union European Cooperation in Science and Technology Action TU0903 “Methods and tools for supporting the use, calibration and validation of traffic simulation models” and the support of the Portuguese National Grid Initiative (https://wiki.ncg.ingrid.pt). A special thank goes to Qiao Ge and Monica Menendez, from ETH, and to Lu Lu and Constantinos Antoniou, from MIT and NTUA, who supported the authors in the application of the quasi-OTEE and WSPSA methods.

References

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008), *Global Sensitivity Analysis: The Primer*, 1st Edition. Wiley-Interscience. February 2008.

Mathew, T. V. and Radhakrishnan, P. (2010), "Calibration of Microsimulation Models for Nonlane-Based Heterogeneous Traffic at Signalized Intersections". *Journal of Urban Planning and Development*, Volume 136, Issue 1, pp. 59–66.

Kesting, A., Treiber, M. (2008), "Calibrating Car-Following Models using Trajectory Data: Methodological Study". *Transportation Research Record: Journal of the Transportation Research Board*, No. 2088, pp. 148–156, National Research Council, Washington, D.C., USA.

Daamen, W., Buisson, C., Hoogendoorn S. P. (Ed., 2014), *Traffic Simulation and Data: Validation Methods and Applications*, CRC Press.

Park, B. and Qi, H. (2005), "Development and Evaluation of a Procedure for the Calibration of Simulation Models". *Transportation Research Record: Journal of the Transportation Research Board*, No. 1934, pp. 208–217, National Research Council, Washington, D.C., USA.

Punzo, V. and Ciuffo, B. (2009), "How Parameters of Microscopic Traffic Flow Models Relate to Traffic Dynamics in Simulation". *Transportation Research Record: Journal of the Transportation Research Board*, No. 2124, pp. 249–256, National Research Council, Washington, D.C., 2009.

Toledo, T., Koutsopoulos, H. N. and Ben-Akiva, M. (2007), "Integrated driving behaviour modeling". *Transportation Research Part C: Emerging Technologies*, Volume 15, pp. 96-112, 2007.

Ciuffo, B. and Lima Azevedo, C. (2014), "A Sensitivity-Analysis-Based Approach for the Calibration of Traffic Simulation Models" *IEEE Transactions on Intelligent Transportation Systems*, vol.PP, no.99, pp.1,12. doi: 10.1109/TITS.2014.2302674, 2014.

Ciuffo, B., Casas, J., Montanino, M., Perarnau, J., and V. Punzo (2013), "Gaussian process metamodels for sensitivity analysis of traffic simulation models". Transportation Research Records, Journal of the Transportation Research Board, no. 2390, pp. 87-98.

Cunto, F. and Saccomanno, F. F. (2008), "Calibration and validation of simulated vehicle safety performance at signalized intersections". *Accident Analysis and Prevention*. Vol. 40, pp 1171-1179, 2008.

Jie, L., Zuylen, H. V., Chen, Y., Viti, F. and Wilmink, I., (2013), "Calibration of a microscopic simulation model for emission calculation". *Transportation Research Part C: Emerging Technologies*, Volume 31, pp. 172–184, 2013.

PTV (2009), *VISSIM 5.20 User Manual*, Karlsruhe, Germany.

Duong, D., Saccomanno, F. F. and Hellinga, B. R. (2009), "Calibration of microscopic traffic model for simulating safety performance". *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, Washington D.C., United States of America, 2009.

Ge, Q., Ciuffo, B. and Menendez, M. (2014), "An efficient sensitivity analysis approach for high dimensional and computationally expensive traffic simulation models". *Reliability Engineering & System Safety*, forthcoming.

Toledo, T. and Koutsopoulos, H. N. (2004), "Statistical Validation of Traffic Simulation Models". *Transportation Research Record: Journal of the Transportation Research Board*, No. 1876, pp. 142–150, National Research Council, Washington, D.C., 2004.

Hoogendoorn, S. P. and Hoogendoorn, R. (2010), "Generic Calibration Framework for Joint Estimation of Car-Following Models by Using Microscopic Data", *Transportation Research Record: Journal of the Transportation Research Board*, No. 2188, pp. 37-45, National Research Council, Washington, D.C., 2010.

Brockfeld, E., Kühne, R. D. and Wagner, P. (2004), "Calibration and Validation of Microscopic Traffic Flow Models", *Transportation Research Record: Journal of the Transportation Research Board*, No. 1876, pp. 45-56, National Research Council, Washington, D.C., 2004.

Zhang, M., Ma, J. and Dong, H. (2008), *Developing Calibration Tools for Microscopic Traffic Simulation Final Report Part II: Calibration Framework and Calibration of Local/Global Driving Behavior and Departure/Route Choice Model Parameters*. California PATH Research ReportUCB-ITS-PRR-2008-7. University of California, Davis, USA.

Hydén, C., (1987), *The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique*. Tech. rep., Lund University, Lund, Sweden.

Sobol, I.M. (1976), "Uniformly Distributed Sequences with an Additional Uniform Property". *Computational Mathematics and Mathematical Physics*, Volume 16, Issue 5, pp. 236-242.

Morris, M. D.(1991), "Factorial sampling plans for preliminary computational experiments". *Technometrics*, Volume 33, Issue 2, pp. 161–174.

Ge, Q. and Menendez, M. (2013), "An improved approach for the sensitivity analysis of computationally expensive microscopic traffic models: a case study of the Zurich network in VISSIM". *Proceedings of the 92nd Annual Meeting of the Transportation Research Board*, Washington D.C., United States of America, 2013.

Matheron, G., (1963), "Principles of geostatistics". *Economic Geology*, Volume 58, pp. 1246–66.

Kleijnen, J. P. C. (2007), "Kriging Metamodeling in Simulation: A Review". *European Journal of Operational Research*, Volume 192, Issue 3, pp. 707–716.

Spall, J. C. (1992), "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation". *IEEE Transactions on Automatic Control*, Volume 37, pp. 332–341.

Lu, L., Yan, X., Antoniou, C. and Ben-Akiva, M. E., (2014). "W-SPSA: An Enhanced SPSA Algorithm for the Calibration of Dynamic Traffic Assignment Models". *Working paper*.

Lima Azevedo, C. (2014), "Probabilistic safety analysis using traffic microscopic simulation". PhD Thesis, Instituto Supeior Técnico, University of Lisbon, Portugal, 2014.

Qi, Y., Koutsopoulos, H. N. and Ben-Akiva, M. E. (2000), "Simulation Laboratory for Evaluating Dynamic Traffic Management Systems". *Transportation Research Record: Journal of the*

Transportation Research Board, No. 1710, pp. 122-130, National Research Council, Washington, D.C., 2000.

Lima Azevedo, C., Cardoso, J. L. and Ben-Akiva, M. E. (2014), "Applying Graph Theory to Automatic Vehicle Tracking by Remote Sensing". *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, Washington D.C., United States of America, 2014.

Hollander, Y. and Liu, R. (2008), "The principles of calibrating traffic micro-simulation models". *Transportation*, Volume 35, pp. 347–362, 2008.

Cascetta, E., Inaudi, D., Marquis, G. (1993), "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts". *Transportation Science*, Volume 27, Issue 4, p. 363–373

Chen, C., Kwon, J., Rice, J., Skabardonis, A., Varaiya, P. (2003), "Detecting Errors and Imputing Missing Data for Single-Loop Surveillance Systems". *Transportation Research Record* 1855, pp. 160–167.

Vaze, V., Antoniou, C., Wen, Y., Ben-Akiva, M. E. (2009), "Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance". *Transportation Research Record: Journal of the Transportation Research Board*, No. 2090, TRB, National Research Council, Washington, D.C., 2009, pp. 1–9.

Kurian M. (2000), *Calibration of a Microscopic Traffic Simulator*. MSc. Thesis, Massachusetts Institute of Technology, Cambridge, United States of America, 2000.

Ahmed, K. (1999), *Modeling Drivers' Acceleration and Lane Changing Behavior*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, United States of America, 1999.

Choudhury, C. F.(2007), *Modeling Driving Decisions with Latent Plans*. PhD Thesis, Massachusetts Institute of Technology. Cambridge, MA, USA.

TSS (2012), *Aimsun Dynamic Simulators Users Manual v7*. TSS, Barcelona, Spain.

Hauer, E., Gårder, P. (1986), "Research into the validity of the traffic conflicts technique". *Accident Analysis and Prevention*, Volume 18, Issue 6, pp 471-481.