Do We Really Need to Calibrate All the Parameters? Variance-Based Sensitivity Analysis to Simplify Microscopic Traffic Flow Models

Vincenzo Punzo, Marcello Montanino, and Biagio Ciuffo

***Abstract*—Automated calibration of microscopic traffic flow models is all but simple for a number of reasons, including the computational complexity of black-box optimization and the asymmetric importance of parameters in influencing model performances. The main objective of this paper is therefore to provide a robust methodology to simplify car-following models, that is, to reduce the number of parameters (to calibrate) without sensibly affecting the capability of reproducing reality. To this aim, variance-based sensitivity analysis is proposed and formulated in a "factor fixing" setting. Among the novel contributions are a robust design of the Monte Carlo framework that also includes, as an analysis factor, the main nonparametric input of car-following models, i.e., the leader's trajectory, and a set of criteria for "data assimilation" in car-following models. The methodology was applied to the intelligent driver model (IDM) and to all the trajectories in the "reconstructed" Next Generation SIMulation (NGSIM) I80-1 data set. The analysis unveiled that the leader's trajectory is considerably more important than the parameters in affecting the variability of model performances. Sensitivity analysis also returned the importance ranking of the IDM parameters. Basing on this, a simplified model version with three (out of six) parameters is proposed. After calibrations, the full model and the simplified model show comparable performances, in face of a sensibly faster convergence of the simplified version.**

***Index Terms*—Calibration, car-following, intelligent driver model (IDM), Next Generation SIMulation (NGSIM), sensitivity analysis, traffic microsimulation, uncertainty management, vehicle trajectory.**

I. Introduction

THE Variability present in the real system, the errors in the measurements of system observables, and the model approximations are acknowledged as some of the most common sources of uncertainty in the modeling of real systems. In the field of traffic (simulation), possible examples of such uncertainties are, respectively, the heterogeneity of driver characteristics, the errors in traffic measurements, and the unmodeled time variability of driving behaviors [1], [2].

Despite their different nature, it is common practice to reduce the impact of such uncertainties on the outputs all at once, that is, by calibrating model parameters. Model calibration—which generally consists of seeking the values of parameters that minimize a distance between system and model outputs—allows reducing the uncertainty in the outputs by incorporating all the sources of uncertainty alongside the parametric inputs. In this view, all the commercial software for traffic simulation allows parameter values to be customized by the users in order "to fit" the traffic model to the system at hand.

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, however, make the automated search for optimal parameter values impracticable for most of the practitioners.

Further elements that hinder calibration are as follows: 1) the scarceness, incompleteness, or inconsistency of data as to the model complexity; 2) the improper setup of the calibration problem (in terms of the chosen fit function, the measure of performance, or the optimization algorithm [3]); and 3) the asymmetry in the importance of model parameters.

The last issue, in particular, represents both an obstacle to the calibration and a way for its solution. In fact, often, (law-driven overparameterized) 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 central processing unit time is exponential in the number of parameters) and solve the issue of flat response surfaces [4].

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.

Manuscript received July 18, 2013; revised November 19, 2013, March 5, 2014, and April 14, 2014; accepted May 29, 2014. Date of publication July 8, 2014; date of current version January 30, 2015. This work was supported by the EU COST Action TU0903 MULTITUDE—Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation models. The Associate Editor for this paper was F.-Y. Wang.

V. Punzo is with the Institute for Energy and Transport, European Commission Joint Research Centre, 21027 Ispra, Italy, and also with the Department of Civil, Environmental and Architectural Engineering, University of Naples "Federico II," 80125 Napoli, Italy (e-mail: vincenzo.punzo@unina.it).

M. Montanino is with the Department of Civil, Environmental and Architectural Engineering, University of Naples "Federico II," 80125 Napoli, Italy (e-mail: m_montanino@msn.com).

B. Ciuffo is with the Institute for Energy and Transport, European Commission Joint Research Centre, 21027 Ispra, Italy (e-mail: biagio.ciuffo@jrc.ec.europa.eu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2014.2331453

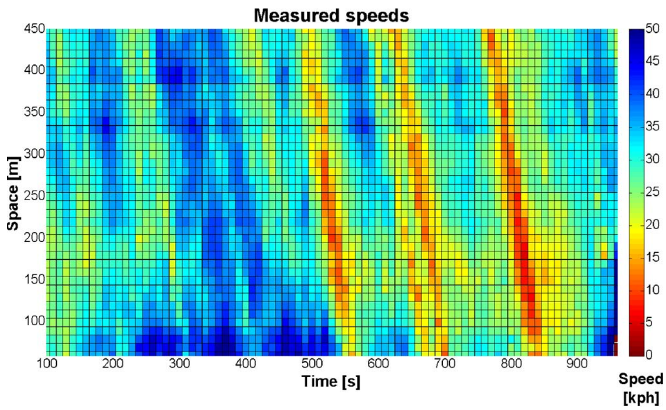

Fig. 1. Edie's space mean speeds from the "reconstructed" I80-1 data set (4:00–4:15 P.M.). Space–time resolution for speed calculation is 10 m × 10 s.

Such methodologies fall in 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" [5]. *Factor fixing*, 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 [6].

The main objective of this paper is therefore to provide a methodology to verify whether it is possible to reduce the number of car-following model 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) [7]; however, the methodology is absolutely general.

Since the main input of a car-following model is the leader's vehicle trajectory and the results of a sensitivity analysis are conditional on the inputs, the input trajectory was included as an additional analysis factor, making results more general.

Therefore, the analysis involved all the vehicle trajectories of one of the "I80" Next Generation SIMulation (NGSIM) data sets [8]: the data set gathered from 4:00 P.M. to 4:15 P.M., called "I80-1" in the following. The data actually used in this paper, however, are those "reconstructed" from the original NGSIM data to solve the many inconsistencies and errors [21]. Reconstructed data are publicly available at [22].

In the reconstructed data set, traffic conditions are moderately congested, and the trajectories comprehend a wide range of dynamics, including stops (the 25% of vehicles—509 out of 2037—experienced speed values below 5 km/h). A comprehensive picture of the traffic flow conditions in the observed time–space domain is given by the contour plot of Edie's mean speeds displayed in Fig. 1.

In addition to the increased generality of the sensitivity analysis results, the inclusion of more than 2000 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. Nonetheless, transferability of results to different contexts, such as interrupted traffic, needs further verification. This paper is organized as follows. In Section II, an introduction to variance-based sensitivity analysis and a brief description of the IDM model are reported. In Section III, the methodology applied throughout the work is described, whereas the results of the application to the IDM model are presented in Section IV. A brief conclusion ends this paper.

II. BACKGROUND

A. Variance-Based Sensitivity Analysis in a Factor-Fixing Setting

The sensitivity analysis technique applied in this paper belongs to the family of the so-called variance-based techniques, which were first employed by Cukier *et al.* [9], generalized by Sobol' [10], [11] with a Monte-Carlo-based implementation of the concept, and then enhanced by Saltelli *et al.* [12] for computation efficiency. For a detailed explanation on the topic the reader can refer to Saltelli *et al.* [12]. As shown in the field literature, those methods were proved to overcome most of the limitations of other common adopted approaches, such as one-at-a-time analysis, differential methods, and regression/correlation analysis. For details, the reader may refer to [6].

The basic idea of the method relies on the assumption that the variance is a good proxy for the uncertainty in the outputs. Given this, the method is based on the well-known variance decomposition formula [13]. Given a model $Y = f(X_1, X_2, \ldots, X_k)$, where $X_i \ \forall i \in [1, k]$ are the input stochastic variables, called *factors*, and Y is the *output* stochastic variable, the variance of the output can be decomposed as

$$V(Y) = V_{X_i}\left(E_{\overline{X}_{\sim i}}(Y|X_i)\right) + E_{X_i}\left(V_{\overline{X}_{\sim i}}(Y|X_i)\right) \quad (1)$$

where X_i is the ith factor, and $\overline{X}_{\sim i}$ denotes the vector of all factors but X_i.

The first component $V_{X_i}(E_{\overline{X}_{\sim i}}(Y|X_i))$ is called the "main (or first-order) effect" of X_i. The associated sensitivity measure, called the "first-order sensitivity index," is equal to the first-order effect normalized over the total (or unconditional) variance, i.e.,

$$S_i = \frac{V_{X_i}\left(E_{\overline{X}_{\sim i}}(Y|X_i)\right)}{V(Y)}. \quad (2)$$

It can be interpreted as the portion of the output variance that is due only to the variation of the input factor X_i. Therefore, the first-order effect captures 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 the interaction (or 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, i.e.,

$$\sum_{i=1}^{k} S_i + \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k} S_{i,j} + \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{l=1 \\ l \neq \{i,j\}}}^{k} S_{i,j,l} + \cdots + \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k} \cdots \sum_{\substack{m=1 \\ m \neq \{i,\ldots\}}}^{k} S_{i,j,\ldots,m} = 1 \quad (3)$$

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

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

$$\mathrm{ST}_i = \frac{E_{\overline{X}_{\sim i}}\left(V_{X_i}(Y|\overline{X}_{\sim i})\right)}{V(Y)} = 1 - \frac{V_{\overline{X}_{\sim i}}\left(E_{X_i}(Y|\overline{X}_{\sim i})\right)}{V(Y)} \quad (4)$$

which 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 in $\sum_{i=1}^{k} \mathrm{ST}_i \geq 1$, where the equality holds only for perfectly additive models (for which $S_i = \mathrm{ST}_i$, $\forall i = 1, \ldots, k$).

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

Following the aforementioned considerations, it is clear that the total sensitivity index is the appropriate measure to address the *factor fixing* quest, that is, "which are the factors that can be fixed at whatever value without affecting the output variance?" Indeed, $\mathrm{ST}_i = 0$ is a necessary and sufficient condition for X_i to be non-influential.

Proof 1: If $\mathrm{ST}_i = 0$, then $E_{\overline{X}_{\sim i}}(V_{X_i}(Y|\overline{X}_{\sim i})) = 0$. As the variance can only be positive, the above relationship implies that $V_{X_i}(Y|\overline{X}_{\sim i} = \overline{x}^*_{\sim i})$ is identically zero for any value of $\overline{x}^*_{\sim i}$. That is, the factor X_i has no influence on the output variance. The necessary condition is obvious.

B. IDM

The car-following model analyzed in this work is the IDM, which belongs to the class of social force models [7]. The social force concept states that the driving behavior 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 [14]. For further details on the model, please refer to [15]. The model formulation is the following:

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

$$\Delta S^*(t) = \Delta S_0 + \max \left\{ \Delta S_1 \cdot \sqrt{\frac{v_f(t)}{V_f^{\mathrm{Max}}}} + T \cdot v_f(t) + \frac{v_f(t) \cdot [v_f(t) - v_l(t)]}{2 \cdot \sqrt{a_f^{\mathrm{Max}} \cdot |b_f|}}, 0 \right\} \quad (5)$$

where the state variables are as follows.

— $v_f(t)$ [m/s] and $a_f(t)$ [m/s^2] are the follower's speed and acceleration at time t, respectively.
— $v_l(t)$ [m/s] is 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 depends on time as the leader vehicle can change over time.
— $\Delta S^*(t)$ [m] is the rear end–front follower's desired distance from the leader.

The model parameters instead are listed next (default values reported in parentheses are those suggested in Treiber *et al.* [7]).

— V_f^{Max} [m/s] is the follower's desired speed (default value: 33.3).
— a_f^{Max} [m/s^2] is the follower's maximum acceleration (default value: 0.73).
— b_f [m/s^2] is the comfort deceleration rate between normal and emergency conditions [7] (default value: 1.67). 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 portion of the desired distance from the leader. It is usually fixed at zero unless special features are required, such as an inflection point in the equilibrium flow density [7]. A zero value has been adopted here, according to the field literature [7].
— T [s] is the minimum time headway between the leader and the follower (default value: 1.6).
— alpha is a constant (default value: 4).

It is worth noting that the $\max\{\cdot\}$ operator in (5) 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., $v_f(t) < v_l(t)$).

III. Methodology

In this section, the methodology developed in this study is described.

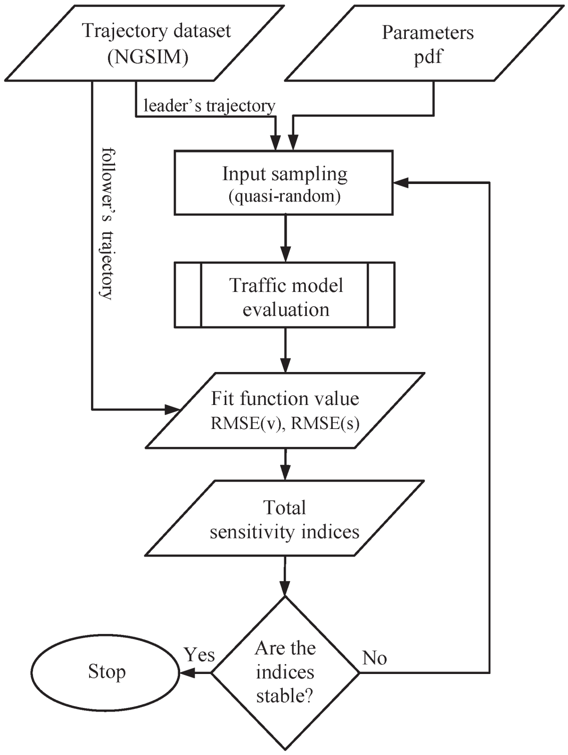

Fig. 2. Flowchart of the Monte Carlo framework for the calculation of sensitivity indices.

A. Sensitivity Analysis in the Monte Carlo Framework

In Fig. 2, the Monte Carlo framework adopted in this paper for the calculation of the sensitivity indices is outlined.

Once the model inputs are drawn by means of a quasi-random sampling, the traffic model is evaluated, and the distance between the observed and the simulated trajectory is calculated in terms of the root-mean-square error (RMSE) of the instantaneous speed or spacing. The process is iterated until the number of evaluations is sufficient for the calculated indices to be stable.

The quasi-random sampling with low-discrepancy sequences (the so called Sobol's sequences) have been here adopted, guaranteeing faster convergence for the indices calculation than other sampling strategies [12]. The formula by Saltelli *et al.* in [12] for the computation of first-order sensitivity indices and that by Jansen [16] for the computation of the total sensitivity indices were applied in this work.

As a "measure of uncertainty" to base the sensitivity analysis on, an "error measure" was adopted. Such a choice is not ordinary in the sensitivity analysis literature where model outputs are generally used instead. However, because the aim of this sensitivity analysis is to investigate the impact of model parameters on calibration results, the same goodness-of-fit (GOF) function applied in calibration was chosen (see Section III-B). Such measure was calculated both on speed and spacing.

Another peculiarity of the framework in Fig. 2 with respect to previous works [4], [19] is that the leader's trajectory has also been chosen as a factor of the analysis and not only the model parameters. That is, not only the parameter values but also the leader's trajectory is sampled—from the NGSIM data set—at each iteration. In fact, it was conjectured that the variance of the outputs explained by the model parameters was not independent from the leader's trajectory but could 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.

The ID number of each of the 2037 leader/follower couples of trajectories in the reconstructed NGSIM I80-1 data set was therefore set as an additional factor with the name of *PairID*.

The six parameters of the IDM were assumed uniformly distributed over the following intervals: alpha $\in [0.1, 10]$, $T \in [0.1, 3]$, $V_f^{\mathrm{Max}} \in [21.7, 30.7]$, $a_f^{\mathrm{Max}} \in [0.5, 4]$, $b_f \in [0.5, 2.5]$, and $\Delta S_0 \in [0.1, 3]$. The assumption of uniform distribution is customary in the absence of *a priori* information on the parameter probability density functions (pdf's).

The interval bounds were chosen following two criteria. The first criterion, when applicable, relates to the meaning of parameters. When a parameter *establishes a direct constraint on a model output*, its boundaries must be set consistently with the real-world values of such output. For instance, given the formulation in (5), the model output acceleration is always lower or equal to a_f^{Max}. Since vehicles, ordinarily, cannot accelerate at values greater than 4 m/s^2, such a value can be therefore adopted as an upper bound for a_f^{Max}.

The second criterion is based on the comparison between the real-world values of the output variance and those provided by the model. In fact, as the amplitude of a parameter interval affects the output variance, parameter values returning unlikely high variance of the output must be left out from the interval. This was done in a trial-and-error manner through the visual observation of the input–output scatter plots. Such criterion, for example, was applied to define the lower bound of a_f^{Max}. In fact, very low values of a_f^{Max} must be discarded, as they cause the simulated trajectories to be unrealistically far from the real ones and the model output variance to be fictitiously higher than the real-world one.

The parameter V_f^{Max} deserves a further consideration. Following the first criterion, it would be possible to set the bounds based on the measured maximum speed of real trajectories. However, this approach sometimes entails a fictitious increase in the output variance (see the second criterion). This happens when a value lower than the measured maximum speed is sampled for the parameter V_f^{Max} of that vehicle. Therefore, a Gaussian distribution with a coefficient of variation of 0.15 was adopted for the free speeds of the I80 stretch, and its known speed limit (i.e., 65 mi/h) was assumed as the 85th percentile of the distribution. The lower and upper bounds of the parameter were chosen equal to the 5th and 95th percentiles of such distribution.

The ensemble of the previous reasoning that led to hypothesize a model for the input uncertainties is also called "data assimilation" in the "uncertainty management" literature [20].

B. Factor Fixing Setting

The sensitivity analysis setting that aims at identifying the non-influential parameters is called factor fixing setting. In Section II-A, it has been shown that an appropriate measure

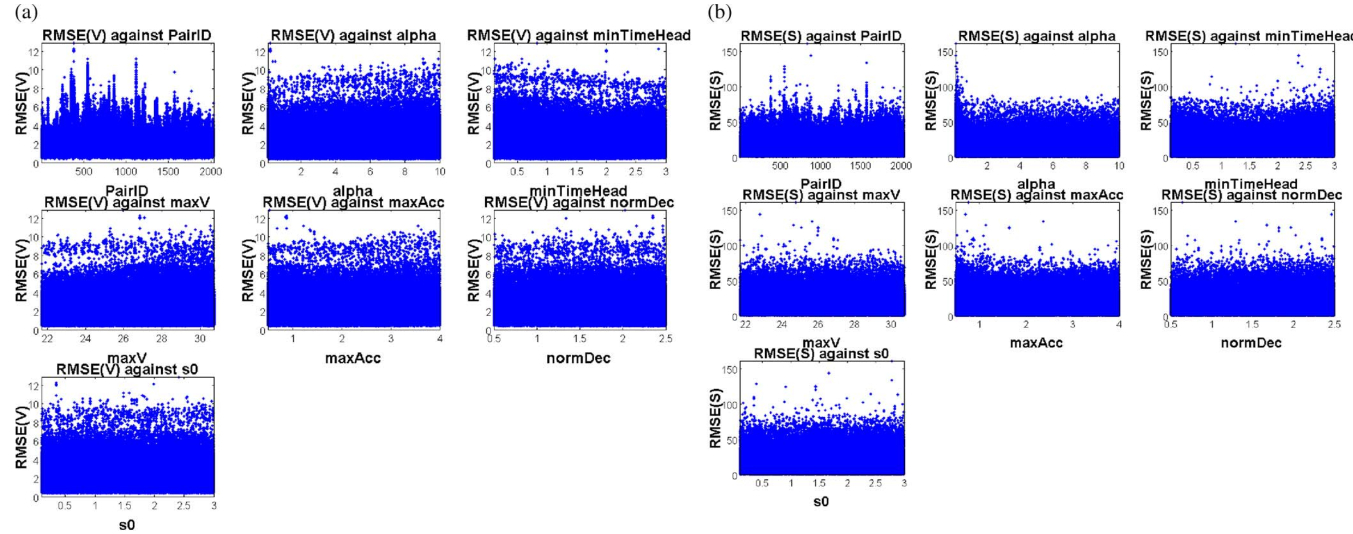

Fig. 3. Scatter plots of the model output against input factors. (a) RMSE(v) [m/s]. (b) RMSE(s) [m].

for such a setting is the "total sensitivity index" ST_i and that $\mathrm{ST}_i = 0$ is both necessary and sufficient for the factor X_i to be non-influential.

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 paper, 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%, that is, explaining less than the 2% of the output unconditional variance could be fixed at any value without sensibly affecting the uncertainty in the model outputs.

C. Calibrations on Reduced and Full 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 [7] and reported in Section II-B were adopted here.

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

The IDM model parameters were calibrated for each individual vehicle (excluding those of type "motorcycle") following the approach reported in [3]. Calibration experiments were run both on speed and on spacing, in order to analyze the effect of using different measurements of performance on the estimation results. The GOF functions were the RMSE on speed, i.e., RMSE(v), and that on spacing, i.e., RMSE(s), whereas the optimization algorithm was the OptQuest Multistart [17]. The upper and lower bounds of parameters in calibration were the same as in the previous sensitivity analysis. Reference [3] proved that the previous calibration setting is the most robust among those applied in the field literature.

TABLE I
FIRST-ORDER AND TOTAL SENSITIVITY INDICES

	RMSE(V)	
Parameter	*First-Order (S) [%]*	*Total (ST) [%]*
PairID	61.88	93.30
alpha	0.57	9.66
minTimeHead	4.14	27.98
maxV	0.09	1.32
maxAcc	0.00	6.81
normDec	0.22	1.33
s0	0.00	0.81
	RMSE(S)	
Parameter	*First-Order (S) [%]*	*Total (ST) [%]*
PairID	35.84	80.40
alpha	6.81	22.47
minTimeHead	6.82	45.27
maxV	0.00	0.91
maxAcc	0.36	9.41
normDec	0.07	1.15
s0	0.02	1.24

IV. RESULTS

A. Sensitivity Analysis

In Fig. 3, the scatter plots of the two error measures against each input factor are presented. Fig. 3(a) relates to RMSE(v), whereas Fig. 3(b) relates to 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 $s0$.

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.

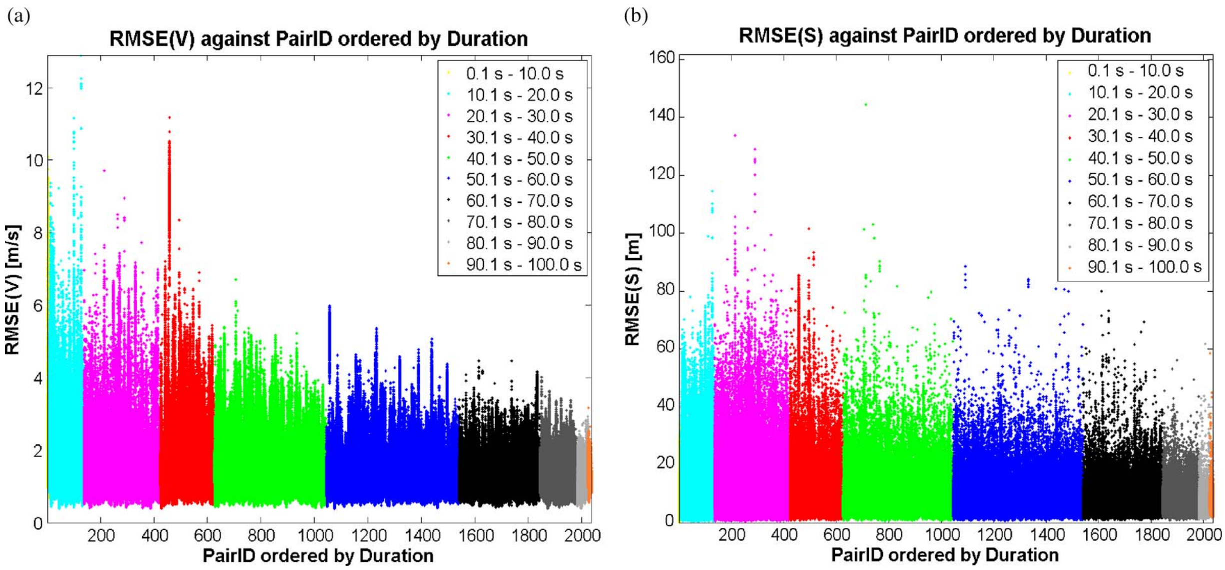

Fig. 4. Scatter plots of the RMSE against the *PairID*. (a) RMSE(v). (b) RMSE(s). The *PairID* values were ordered by duration of the trajectory on the color scale.

The variability of the output against an input factor gives graphical information about the first-order effect of that 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, whereas a uniform cloud is a symptom—although not a proof—of a non-influential one (see [6] for an explanation).

From the visual inspection of scatter plots in Fig. 3, a clear pattern in the variance of the RMSE statistics can be appreciated only for the *PairID* factor. This shows that the leader's trajectory is 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 (see Fig. 5 and Table I).

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 is the value of the model RMSE for a specific parameter combination. The higher the RMSE variance, the higher the chance that, for that trajectory, the model yields high errors for non-optimal parameter combinations.

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

The last guess, in particular, was tested in Fig. 4, where the trajectories are ordered by their duration and not by their *PairID* (points were also colored according to the duration intervals reported in the legend). The result clearly confirms the guess: in both the plots, the RMSE variance decreases as duration increases. Thus, the chance of having very high errors 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. This suggests that longer trajectories including different driving regimes (stop-and-go and free-flow) should be used for calibration.

Moving to the analysis of the sensitivity indices, Fig. 5 reports the values of the first-order and total sensitivity indices (jointly with their 90% confidence intervals) of all the input factors calculated on both RMSE(v) and RMSE(s). The numerical values are reported instead in Table I.

A total number of 2^{17} model evaluations were necessary to have clearly stable sensitivity indices (please note that in quasi-random sampling, the size of the experiment has to be a power of 2 [10]).

Most of the error measure variance is explained by the PairID. Comparing the total sensitivity indices of the PairID on both speed and spacing (93.30 and 80.40) with those of all the other factors, it comes out clearly that the model performances can be very low if the model is not calibrated against each trajectory (this is not a critique to the IDM as it holds for all the car-following models).

Regarding the model parameters, their impact on the error measure is mainly due to the interaction effects. In fact, the sum of their first-order effects explains less than 5% and 15% of the total variance of the error on speed and spacing, respectively.

The total sensitivity measures also define the rank of influence of model parameters. The minimum time headway (with its interaction effects) explains 28% and 45% of the total output variance of RMSE(v) and RMSE(s), respectively, followed by alpha with 10% and 22% and by maximum acceleration with 7% and 10%. Each remaining parameter (maxV, normDec, and s0) instead explains less than 1.5% of the total output variance.

According to this ranking, in the second part of the work, we tested the impact on the calibration performances of fixing the non-influential parameters, i.e., those with an ST_i lower than 2%, to values commonly used in the literature.

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

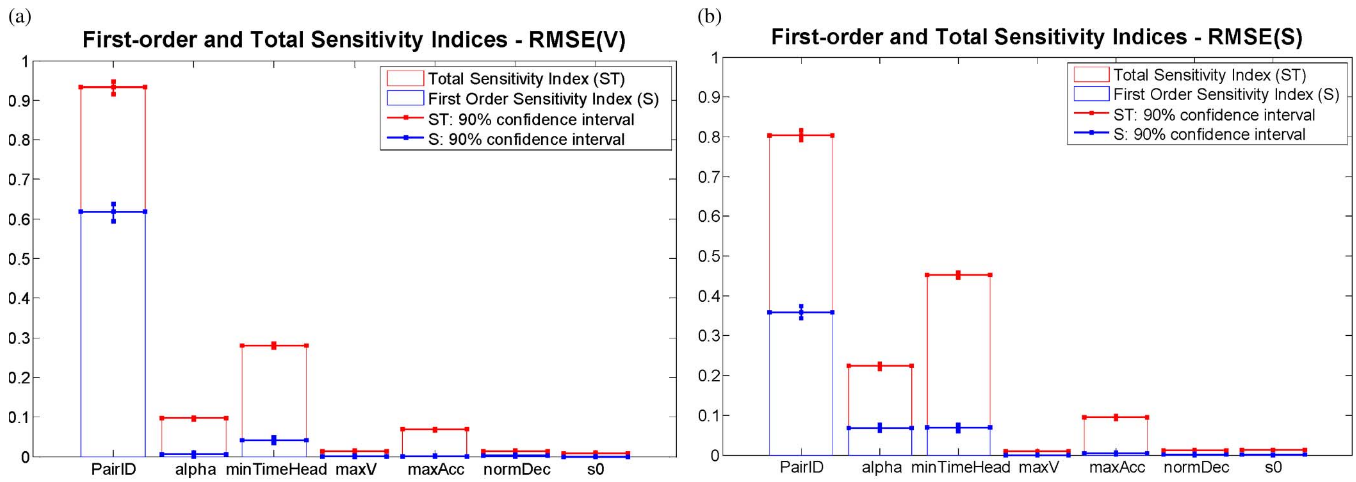

Fig. 5. First-order and total sensitivity indices of input factors calculated on (a) RMSE(v) and (b) RMSE(s).

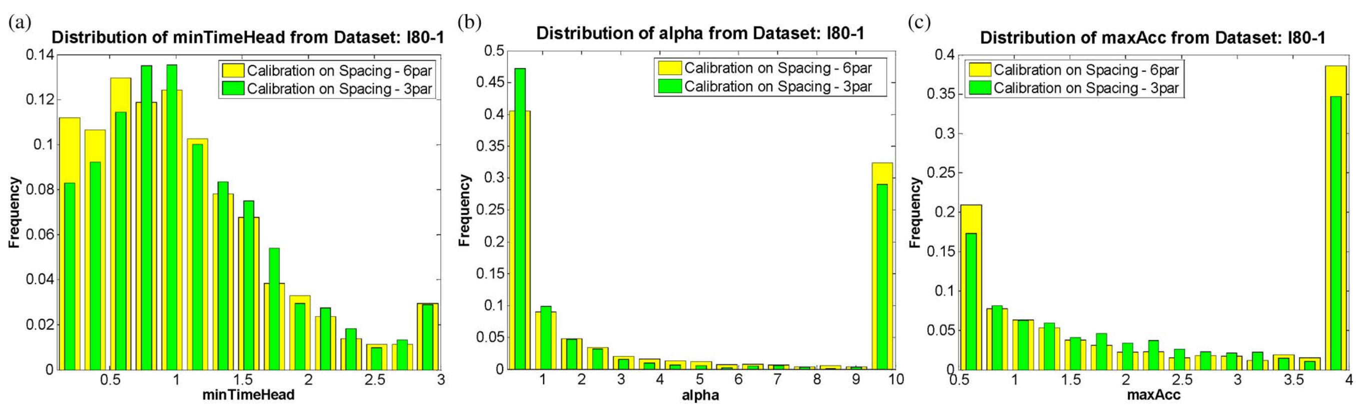

Fig. 6. Empirical distributions of parameter values after calibration against RMSE(s), for the full model (yellow bars) and the three-parameter reduced model (green bars). (a) Distribution of minimum time headway. (b) Distribution of alpha. (c) Distribution of maximum acceleration.

B. Calibration Results

A comparison of the histograms of the estimated parameters for the full model and the reduced model is presented in Fig. 6 (the figure shows results on spacing, although very similar ones hold on speed). Shapes and values are very similar between the two model configurations. This was expected, given the very low degree of interaction between the influential (calibrated) parameters and the non-influential (fixed) ones, explained by the sensitivity analysis.

Concerning the alpha and maximum acceleration histograms [see Fig. 6(b) and (c)], the estimated distributions from model calibration would be different if other boundary values were adopted. For example, for many trajectories, the optimal parameter value would lay above the upper bound here adopted—we recall that such boundary values were those chosen in the data assimilation phase to satisfy the physical expectations on the output (see the first criterion in Section III-A). It is worth underlining here that measured values of "physical" parameters can be used only to set the boundary values (as just recalled). However, optimal values for such parameters must be obtained from calibration: this is evident from Fig. 6(c), where we can observe that the estimated distribution of "maximum acceleration" is clearly different from what we could measure in reality (i.e., if we measured the empirical distribution of vehicles' maximum accelerations). In other words, this is a further empirical evidence that also parameters that are measurable in reality (e.g., maximum acceleration) need to be calibrated in order to cover modeling uncertainties.

Fig. 7 presents the comparison between the distributions of the minimum error after calibration of the full model and the reduced model. Fig. 7(a) refers to the calibration on speed, whereas Fig. 7(b) refers to that on spacing.

A general consideration is that the increase in the calibration error is intuitively expected of the same magnitude of the variance explained by the three parameters that have been fixed. However, this is not true. Indeed, in sensitivity analysis, the total variance of the model error is investigated, whereas here, the focus is on a very small part of this variance, i.e., the variation of the residual error after calibration, between the full model and the reduced model.

Moving to such results, in the case of calibration on speed [see Fig. 7(a)], fixing non-influential model parameters produced very little effects on the capability of the model to reproduce the follower trajectory, as compared with the full model. Indeed, the mean value of RMSE(v) increased, on average, by 6.85% only. Although distributions are statistically different at

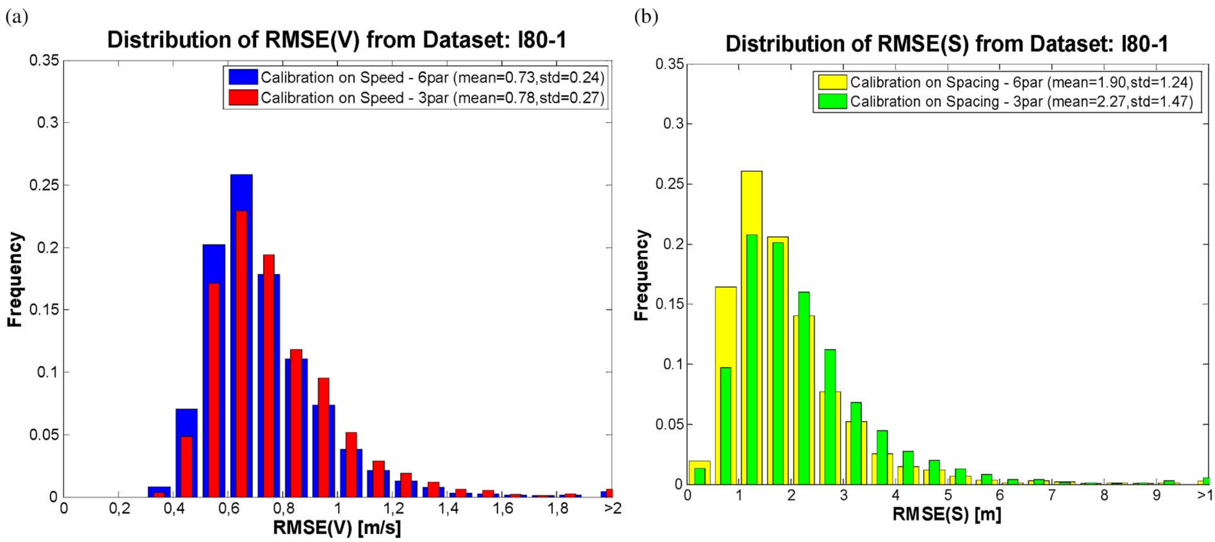

Fig. 7. Histograms of model errors after calibration against (a) RMSE(v) and (b) RMSE(s), for all the vehicles in the reconstructed NGSIM I80-1 data set. In (a), the blue bars refer to the full model (i.e., six parameters) and the red ones refer to the reduced model (i.e., three parameters estimated: T, alpha, and a_f^{Max}). In (b), the yellow bars refer to the full model, and the green ones refer to the reduced model.

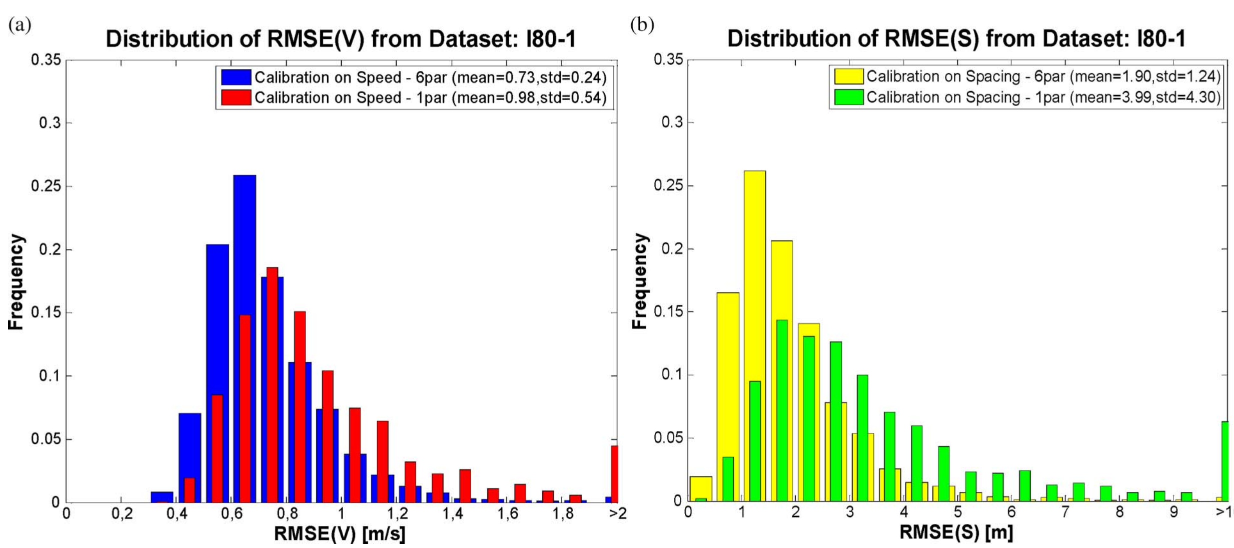

Fig. 8. Histograms of model errors after calibration against (a) RMSE(v) and (b) RMSE(s), for all the vehicles in the reconstructed NGSIM I80-1 data set. In (a), the blue bars refer to the full model (i.e., six parameters), and the red ones refer to the reduced model (i.e., one parameter estimated: T). In (b), the yellow bars refer to the full model, and the green ones refer to the reduced model.

the level of confidence of 5%, it can be asserted that, from a practical point of view, the simplified model performed mostly the same as the full model.

When calibrating on spacing, instead, the increase in the average residual error is more sensible: about 19% [see Fig. 7(b)]. This is the consequence of the integral nature of the spacing error that makes the full model calibration on spacing more accurate than on speed (see [18]) and emphasizes the increase in the calibration error when moving to the simplified model. Notwithstanding, in the absolute value, such an increase corresponds to an additional error of 37 cm, on average. It is claimed here that such an error is negligible in practical applications.

In terms of computational time, the benefit of calibrating the simplified model is evident as the number of model evaluations for the convergence decreased by about 80% (from 24 700 to 3900 for speed and from 24 400 to 4900 for spacing).

Given these promising results and that only one parameter explains the biggest part of the output variance, therefore, one may wonder which would be the impact of calibrating only that parameter, i.e., the minimum time headway T. Results are shown in Fig. 8.

The picture offered is in accordance with the previous findings. On the one hand, it confirms that calibration on spacing is more challenging than on speed [18], resulting in a higher increase in the calibration error. On the other hand, the spread between model performances and calibration burden is clearly larger than in the previous case.

V. Conclusion

This paper aimed to give a contribution on calibration of microscopic traffic flow models.

As automated calibration is arduous for many reasons, including the computational complexity (in black-box optimization settings) and the asymmetry in the sensitivity of model parameters, the main idea and contribution of this paper is to provide a robust methodology to simplify models.

Following on previous authors' investigations [4], [19], this paper is based on the application of variance-based techniques for the sensitivity analysis of car-following models, in a "factor fixing" setting.

Among the original methodological contributions are as follows:

i) a novel formulation for the factor fixing setting, where the "model performance," instead of the "model output," is adopted as a 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 nonparametric input of car-following models that is the leader's trajectory;
iii) general criteria for "data assimilation" in car-following models, i.e., to set the parameter bounds for the model sensitivity analysis and calibration.

Such methodology was applied to the IDM and to the complete set of vehicle trajectories from the "reconstructed" NGSIM I80-1 data set.

First, sensitivity analysis showed that the leader's trajectory is considerably more important than the parameters in affecting the variability of model performances. Sensitivity analysis 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 behaviors, model parameters need calibration, and that long trajectories are required for robust estimation.

Sensitivity analysis also returned the "importance ranking" of IDM parameters. The ranking was the same when using speed or spacing as measure of performance, although the magnitude of parameter sensitivity resulted different between the two. Among the six model parameters, the "minimum time headway" explained most of the variance of the error measure, followed by "alpha" and "maximum acceleration."

Basing on these results, a reduced version of the IDM model was formulated by fixing three (out of six) parameters at default literature values. Both the model versions were then calibrated against trajectories.

Calibration performances of the reduced model resulted comparable with those of the full model, in face of a sensibly faster convergence. This holds both for the case of calibration on speed and for the more challenging (and effective) one of calibration on spacing. Inspired by the "worst practice" of calibrating traffic simulation models manually, a quantitative estimate of the IDM performances in case of calibrating only the most influential parameter, i.e., the "minimum time headway," was also provided.

Eventually, the analysis unveiled the "true" meaning of model calibration. It was shown that also parameters that can be measured in reality (e.g., maximum acceleration) need to be calibrated if modeling uncertainties are to be covered.

Acknowledgment

The authors would like to thank S. Tarantola and S. Kucherenko for the invaluable discussions on the setting of the sensitivity analysis technique and the anonymous reviewers for the precious comments. Research contained within this paper benefited from participation in EU COST Action TU0903 MULTITUDE—Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation models.

References

[1] S. Ossen, S. P. Hoogendoorn, and B. G. H. Gorte, "Interdriver differences in car-following. A vehicle trajectory-based study," *Transp. Res. Rec.*, vol. 1965, pp. 121–129, 2006.
[2] A. Kesting and M. Treiber, "Calibrating car-following models by using trajectory data," *Transp. Res. Rec.*, vol. 2088, pp. 148–156, 2008.
[3] V. Punzo, B. Ciuffo, and M. Montanino, "Can results of car-following model calibration based on trajectory data be trusted?" *Transp. Res. Rec.*, vol. 2315, pp. 11–24, 2012.
[4] V. Punzo and B. Ciuffo, "Sensitivity analysis of car-following models: Methodology and application," presented at the 90th TRB Annual Meeting, Washington, DC, USA, Jan. 23–27, 2011.
[5] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, "Global sensitivity analysis for importance assessment," in *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*., 1st ed. Chichester, U.K.: Wiley, 2004, pp. 31–61.
[6] A. Saltelli *et al.*, "Introduction to sensitivity analysis," in *Global Sensitivity Analysis—The Primer*., 1st ed. Chichester, U.K.: Wiley, 2008, pp. 1–51.
[7] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Nonlin. Soft Matter Phys.*, vol. 62, no. 2, pp. 1805–1824, Aug. 2000.
[8] NGSIM—Next Generation SIMulation, Jul. 2013. [Online]. Available: http://www.ngsim-community.org
[9] R. I. Cukier, C. M. Fortuin, K. E. Schuler, A. G. Petsheck, and J. H. Schailbly, "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients," *J. Chem. Phys.*, vol. 26, no. 8, pp. 1–42, Oct. 1973.
[10] I. M. Sobol, "Sensitivity analysis for non-linear mathematical models," *Math. Models Comput. Exp.*, vol. 1, no. 4, pp. 407–414, 1993.
[11] I. M. Sobol, S. Tarantola, D. Gatelli, S. Kucherenko, and W. Mauntz, "Estimating the approximation error when fixing unessential factors in global sensitivity analysis," *Reliab. Eng. Syst. Safety*, vol. 92, no. 7, pp. 957–960, Jul. 2007.
[12] A. Saltelli *et al.*, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," *Comput. Phys. Commun.*, vol. 181, no. 2, pp. 259–270, Feb. 2010.
[13] A. M. Mood, F. A. Graybill, and D. C. Boes, "Random variables, distribution functions, expectation," in *Introduction to the Theory of Statistics*., 3rd ed. New York, NY, USA: McGraw-Hill, 1974, pp. 51–72.
[14] H. Wang, W. Wang, J. Chen, and M. Jing, "Car-following model calibration and analysis of intra-driver heterogeneity," *Adv. Mater. Res.*, vol. 108–111, pp. 805–810, 2010.
[15] M. Treiber, A. Kesting, and D. Helbing, "Delays, inaccuracies and anticipation in microscopic traffic models," *Phys. A, Stat. Mech. Appl.*, vol. 360, no. 1, pp. 71–88, Jan. 2006.
[16] M. J. W. Jansen, "Analysis of variance designs for model output," *Comput. Phys. Commun.*, vol. 117, no. 1/2, pp. 35–43, Mar. 1999.
[17] "MXLINDO: A Matlab Interface," in *LINDO API User Manual 2.0*. Chicago, IL, USA: LINDO Systems, Inc., 2003, pp. 269–338.
[18] V. Punzo and F. Simonelli, "Analysis and comparison of microscopic traffic flow models with real traffic microscopic data," *Transp. Res. Rec.*, vol. 1934, pp. 53–63, 2005.
[19] B. Ciuffo, V. Punzo, and M. Montanino, "Global sensitivity analysis techniques to simplify the calibration of traffic simulation models. Methodology and application to the IDM car-following model," *IET Intell. Transp. Syst.*, to be published.
[20] E. De Rocquigny, N. Devictor, and S. Tarantola, *Uncertainty in Industrial Practice: A Guide to Quantitative Uncertainty Management*. New York, NY, USA: Wiley, 2008.

[21] M. Montanino and V. Punzo, “Making NGSIM data usable for studies on traffic flow theory: A multistep method for vehicle trajectory reconstruction,” *Transp. Res. Rec. 2390*, pp. 99–111.

[22] M. Montanino and V. Punzo, “Reconstructed NGSIM I80-1,” Cost Action TU0903—Multitude, Mar. 2014. [Online]. Available: http://www.multitude-project.eu/exchange/101.html

Vincenzo Punzo received the M.Sc. degree in civil engineering and the Ph.D. degree in transportation engineering from University of Naples “Federico II,” Naples, Italy, in 1998 and 2002, respectively.

From 2003 to 2004 he was a Postdoctoral Fellow and since 2005 he has been an Assistant Professor with the Department of Civil, Architectural and Environmental Engineering, University of Naples “Federico II.” On temporary leave since 2011, he joined the European Commission Joint Research Centre, Ispra, Italy, as a Senior Researcher. His research interests include the modeling and simulation of road and rail transportation systems, intelligent transportation systems (ITS), driving simulation, and road safety, reflected by more than 50 papers in peer-reviewed journals and conference proceedings.

Dr. Punzo is a member of Editorial Boards of journals (*Transportation Research—Part C* and the *International Journal of Transportation*) and of the Transportation Research Board (TRB)’s Traffic Flow Theory Committee (AHB45) and an Associate Editor of IEEE Conference on Intelligent Transportation Systems. He is the Chairman of the COST Action TU0903 MULTITUDE on the topic of the uncertainty management in traffic simulation. In 2012 he received the Greenshields Prize of the TRB of the National Academies, USA, and, in 2013, of the Best Simulation Paper of the TRB Joint Traffic Simulation Subcommittee.

Marcello Montanino was born in Naples, Italy, in 1987. He received the M.S. degree in transportation engineering and the Ph.D. degree from University of Naples “Federico II,” Naples, in 2010 and 2014, respectively.

He is a Postdoctoral Fellow with the Department of Civil, Environmental and Architectural Engineering, University of Naples “Federico II”. His main research topic is the management of uncertainty in modeling and decision-making processes. In particular, he focused on the application of variance-based techniques for global sensitivity analysis of traffic simulation models and on the application of multiagent multicriteria decision-making analyses for the comparison and assessment of alternative project options.

Dr. Montanino has been a member of the COST Action TU0903 MULTITUDE since 2010. He received the Greenshields Prize of the Transportation Research Board (TRB) of the National Academies, USA, in 2012 and of the Best Simulation Paper of the TRB Joint Traffic Simulation Subcommittee in 2013.

Biagio Ciuffo received the M.Sc. degree in engineering for environment and land use in 2004 and the Ph.D. degree from University of Naples “Federico II,” Napoli, Italy.

He is currently a Scientific Officer with the Institute for Energy and Transport, European Commission Joint Research Centre, Ispra, Italy. He is the author of more than 20 papers in peer-reviewed journals and conference proceedings in transportation engineering. His main research interests concern the management of modeling uncertainties, traffic simulation and control, emissions estimation from maritime traffic, sustainable development, and sustainability appraisal.

Dr. Ciuffo is a member of the COST Action MULTITUDE as a Working Group Leader. He received the Greenshields Prize of the Transportation Research Board (TRB) of the National Academies, USA, in 2012 and of the Best Simulation Paper of the TRB Joint Traffic Simulation Subcommittee in 2013.