A Sensitivity-Analysis-Based Approach for the Calibration of Traffic Simulation Models

Biagio Ciuffo and Carlos Lima Azevedo

***Abstract*—In this paper, a multistep sensitivity analysis (SA) approach for model calibration is proposed and applied to a complex traffic simulation model with more than 100 parameters. Throughout this paper, it is argued that the application of SA is crucial for true comprehension and the correct use of traffic simulation models, but it is also acknowledged that the main obstacle toward an extensive use of the most sophisticated techniques is the high number of model runs usually required. For this reason, we have tested the possibility of performing a multistep SA, where, at each step, model parameters are grouped on the basis of possible common features, and a final SA on the parameters pertaining to the most influential groups is then performed. The proposed methodology was applied to an urban motorway case study simulated using MITSIMLab, a complex microscopic traffic simulator. The method allowed the analysis of the role played by all parameters and by the model stochasticity itself, with 80% fewer model evaluations than the standard variance-based approach. Ten model parameters accounted for a big share in the output variance for the specific case study. A Kriging metamodel was then estimated and integrated with the multistep SA results for a global calibration framework in the presence of uncertainty. Results confirm the great potential of this approach and open up to a novel view for the calibration of a traffic simulation model.**

***Index Terms*—Calibration, global sensitivity analysis (SA), traffic simulation, uncertainty management.**

I. Introduction

CALIBRATION and validation procedures are increasingly ascending transportation modelers' and practitioners' top priorities, as the use of such tools is quickly spreading and its models progressively improved. Furthermore, the access to both new and advanced modeling techniques and detailed traffic and behavioral data is increasing the level of detail of improved traffic simulation models [1], [2]. Furthermore, traffic simulators are increasingly being applied in many different traffic situations, and consistency with the available data needs to be assured. These challenges have been linked to the need of a consistent understanding of the simulators performance, along with the appropriate calibration and validation procedures to monitor its uncertainty. Sensitivity analysis (SA) is the tool used with this aim [3].

The one-at-a-time (OAT) approach remains the most adopted method when dealing with microscopic simulation models. OAT measures are based on the estimation of partial derivatives and assess how uncertainty in one factor affects the model output, keeping the other factors fixed to a nominal value. The main drawback of this approach is that interactions among factors cannot be detected, since they require the inputs to be simultaneously changed [3]. In addition, this approach pertains to a family of SA techniques usually referred to as local SA, which are used to derive information on the behavior of the model around a certain point rather than for exploring its input space. However, its simplicity and parsimony makes it the preferred choice for practitioners. The OAT approach has been applied with microscopic simulation model by [4] and [5] in order to rank model parameters in terms of their effects on the model outputs and to select the parameters to be calibrated, respectively. In [6], the same approach is followed in order to get insight on the meaning of the values of parameters resulting from the calibration of car-following models.

A more advanced method also found in the literature is the analysis of variance (ANOVA). For details on experimental design techniques and ANOVA, one may refer to technical books such as [7]. ANOVA has been used in [8] and [9] to draw inference about the first-order effect of a set of traffic microsimulation model parameters. Interaction effects were not captured since a two-level full factorial design was adopted in both studies. A three-level factorial design was used in both [10] and [11], but the second-order interaction effects could be only evaluated in [11], as a fractional design was adopted in [10]. Another study that used ANOVA to undertake a model SA was carried out by [12]. In this study, five levels per parameter were taken into account, and a Latin hypercube sampling algorithm was used to define the experimental design. However, also in this case, the interaction effect of the parameters was not evaluated.

Alternatively, a more efficient SA method based on variance decomposition may be also found in the literature. In traffic modeling, the variance decomposition approach has recently been used by [13] for the SA of car-following models.

All the aforementioned work refer to applications with either one or two behavioral models with few parameters or considering just a subset of them. In particular, when dealing with complex traffic simulation models, it is a common practice to make a prior selection of the parameters to be involved in the analysis. The selection is based on *a priori* knowledge of

Manuscript received July 15, 2013; revised November 14, 2013; accepted January 9, 2014. Date of publication March 5, 2014; date of current version May 30, 2014. The Associate Editor for this paper was C. Antoniou.

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

C. Lima Azevedo is with the Department of Transportation, National Laboratory for Civil Engineering (LNEC), Lisbon 1700-066, Portugal (e-mail: cmazevedo@lnec.pt).

Digital Object Identifier 10.1109/TITS.2014.2302674

the model, on advice, or on common sense by developers [3], [13]. This is a fairly dangerous practice, as many interactions among groups of parameters may remain hidden even to the most expert model users. The problem is that complex traffic simulation models involve dozens of parameters, and a full SA would require too many model evaluations. Although the importance of the accurate identification of this subset of parameters was identified in the past [14], only a few number of recent studies focus on the systematization of such procedures [13]. Global methods may provide far more insight for nonlinear models, but they can require large numbers of model evaluations. In addition, the interpretation of such global sensitivity tests can be difficult due to the high number of parameters involved and the potential complexity of the causal relationships.

To overcome these issues, we propose a multistep SA approach. In this framework, the process of identifying sensitive parameters is broken down into separate SA with increasing detail as the most sensitive parameters are being identified. This process is based on grouping the parameters and applying SA techniques at each step and applying a final SA on the individual parameters belonging to the subset of the most influential groups. The SA technique used in this case study is the variance-based approach based on Sobol's decomposition of variance [15]. The general proposed framework is described in detail in the next section, and the mathematical details of the variance-based formulation are presented in Section III. The method was tested in the SA of a simulated urban motorway using the microscopic traffic simulator MITSIMLab [16] (described in Section IV), and the results are presented in Section V. The multistep SA framework was then extended to the global calibration of MITSIMLab using a Kriging-based approach (see Section VI). Finally, the main conclusions from this work are presented in Section VII.

II. Multistep Global SA

To overcome the mentioned limitations of generic SA approaches, a multistep approach for complex traffic simulation models SA is proposed.

In the first step, parameters are grouped with respect to the submodels they are part of (e.g., same submodel, same physical interpretation, etc.), and an SA is carried out considering the different groups rather than the different parameters. Then, the most influential groups on the model outputs are identified, and a new SA on all the parameters of these groups is carried out. If from the first analysis still too many parameters result, a new grouping and an additional group SA may be carried out. After a further reduction of the parameters set, a final SA can identify the subset of final model inputs to be estimated with particular care (see Fig. 1). The proposed approach applies to any type of traffic simulation model and, in general, to any modeling framework composed of different independent submodels interacting among each other. It is thought for models in which the total number of parameters makes the direct application of the selected SA technique unfeasible. It is worth pointing out that the increased design efficiency of the group SA method comes at the cost of information about the interaction between parameters belonging to the same group. In fact, only the full variance-based SA will account for this effect. If one suspects the presence of such interaction within groups identified as nonsensitive, it is, however, possible to test it with full SA for the specific submodel at the burden of a higher number of simulations.

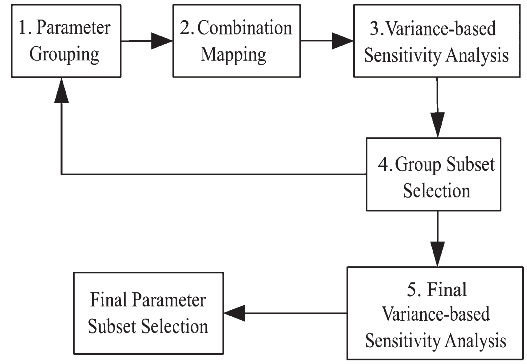

Fig. 1. General framework for the multistep global SA.

As shown in Fig. 1, the SA step is based on the computation of first- and total-order sensitivity indices with a variance-based approach [3]. This approach requires the evaluation of the model $N \times (k+2)$ times, where k is the number of model parameters, and N is the dimension of the Monte Carlo experiment [3].

The methodology is composed of the following steps.

1) Group model parameters on the basis of their similarities.
2) Create a map between a number in the range [0, 1] and a combination of the parameters within the same group.
3) Apply variance-based SA to the groups to distinguish those accounting for the highest share of model variance.
4) Consider only the parameters in the influential groups.
   a) If the number is sufficiently small, apply variance-based techniques to the new set of parameters.
   b) If the number is still too high, go back to step 1, applying steps 1–4 to the new parameter subset.
5) Define the final set of parameters to be included in the subsequent analyses.

A key step in the methodology is represented by step 2. The map between a number in the range [0, 1] and a combination of parameters determines the quality of the sensitivity indices. In general, it is necessary to have a sufficient exploration of the parameters space. For a specific mapping, the value to be assigned to each parameter is extracted from a predefined distribution.

III. Variance-Based Methods on the Sobol Decomposition of Variance

The original formulation of the decomposition of variance is due to Sobol [15], who provided the analytical derivation and the Monte-Carlo-based implementation of the concept. The latest setting for its practical implementation, instead, is due to [3]. Given a model in the form $Y = f(Z_1, Z_2, \ldots, Z_k)$, two factors are said to interact when their effect on Y cannot be expressed as a sum of their single effects. Interactions represent important features of traffic models and are more difficult to detect than first-order effects. For example, by using regression analysis tools, it is fairly easy to estimate first-order indices, but not interactions. With Y being a scalar, a variance-based first-order effect for a generic factor Z_i can be written as

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

where Z_i is the ith factor, and $Z_{\sim i}$ is the matrix of all factors but Z_i. Furthermore, it is known that the unconditional variance can be decomposed into main effect and residual, i.e.,

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

Equation (2) shows that for Z_i to be an important factor, we need $E_{Z_i}(V_{\mathbf{Z}_{\sim i}}[Y|Z_i])$ to be small, that it is to say that the closer $V_{Z_i}(E_{\mathbf{Z}_{\sim i}}[Y|Z_i])$ 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 γ as

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

Sensitivity indices as in (3) can be calculated per each factor and per each factor combination. This, however, would need a huge amount of model evaluations. In order to reduce the efforts required, a synthetic indicator to be coupled with the first-order sensitivity index is the total effects index, which is defined as follows [1]:

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

Total effects index of the input factor i provides the sum of the higher and first-order effects (interactions) of factor Z_i. When the total index is $S_{T_i} = 0$, the ith factor can be fixed without affecting the outputs' variance. If $S_{T_i} \sim 0$, the approximation made depends on the value of S_{T_i}. It is worth noting that while $\sum_{i=1}^{k} S_i \leq 1$, $\sum_{i=1}^{k} S_{T_i} \geq 1$, both being equal to 1 only for additive models. Since the analytical feasibility of traffic flow models limits the use of the formulas for the calculation of the variances in (2), the application of this method can be effectively performed in a Monte Carlo setting.

The approach adopted in this paper has been specified in [3] and [13] to avoid brute-force computation of the multidimensional integrals of the Monte Carlo experiment for the space of the input factors and can be summarized in the following points:

1) Generate a $(N, 2k)$ matrix of random numbers (k is the number of inputs) and define two matrices of data ($\mathbf{A}$ and $\mathbf{B}$), each containing half of the sample, using sequences of quasi-random numbers [15]. N is called a base sample.

$$\mathbf{A} = \begin{bmatrix} Z_1^{(1)} & Z_2^{(1)} & \cdots & Z_k^{(1)} \\ Z_1^{(2)} & Z_2^{(2)} & \cdots & Z_k^{(2)} \\ \cdots & \cdots & \cdots & \vdots \\ Z_1^{(N)} & Z_2^{(N)} & \cdots & Z_k^{(N)} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} Z_{k+1}^{(1)} & Z_{k+2}^{(1)} & \cdots & Z_{2k}^{(1)} \\ Z_{k+1}^{(2)} & Z_{k+2}^{(2)} & \cdots & Z_{2k}^{(2)} \\ \cdots & \cdots & \cdots & \vdots \\ Z_{k+1}^{(N)} & Z_{k+2}^{(N)} & \cdots & Z_{2k}^{(N)} \end{bmatrix}.$$

2) Define a matrix $\mathbf{C}_i$ formed by all columns of $\mathbf{A}$, except the ith column, which is taken from $\mathbf{B}$ (with i varying from 1 to k), i.e.,

$$\mathbf{C}_i = \begin{bmatrix} Z_1^{(1)} & Z_2^{(1)} & \cdots & Z_{k+i}^{(1)} & \cdots & Z_k^{(1)} \\ Z_1^{(2)} & Z_2^{(2)} & \cdots & Z_{k+i}^{(2)} & \cdots & Z_k^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ Z_1^{(N)} & Z_2^{(N)} & \cdots & Z_{k+i}^{(N)} & \cdots & Z_k^{(N)} \end{bmatrix}$$

for $i = 1, \ldots, k$.

3) The model is evaluated for all the $[N \times (k+2)]$ combinations of input variables as given by matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ to produce the $(N \times 1)$ vectors of outputs $y_A = f(\mathbf{A})$, $y_B = f(\mathbf{B})$, and $y_{c_i} = f(\mathbf{C}_i)$, for $i = 1, \ldots, k$. These vectors are sufficient for the evaluation of all the first-order S_i and total effects S_{T_i} indices. Because there are k factors, the cost of this approach is $N + N$ runs of the model for matrices $\mathbf{A}$ and $\mathbf{B}$ plus k times N to estimate k times the output vector corresponding to matrix $\mathbf{C}_i$. Hence, the total cost is $N \times (k+2)$, much lower than the N^2 runs of the brute-force method. Since N is usually not lower than 1000, the number of evaluation required by this efficient approach is not, in any case, negligible, particularly for complex and expensive models. For this reason, in the common practice, the approach presented here can be considered relevant.

The sensitivity indices presented in (3) and (4) can be evaluated using the following formulations [3], [17]:

$$S_i = \frac{\frac{1}{N}\sum_{j=1}^{N} y_B^{(j)}\left(y_{C_i}^{(j)} - y_A^{(j)}\right)}{\left(\frac{1}{2N}\sum_{j=1}^{N}\left(y_{A+B}^{(j)}\right)^2 - \left(\frac{1}{2N}\sum_{j=1}^{N} y_{A+B}^{(j)}\right)^2\right)} \tag{5}$$

$$S_{Ti} = \frac{\frac{1}{2N}\sum_{j=1}^{N}\left(y_A^{(j)} - y_{C_i}^{(j)}\right)^2}{\left(\frac{1}{2N}\sum_{j=1}^{N}\left(y_{A+B}^{(j)}\right)^2 - \left(\frac{1}{2N}\sum_{j=1}^{N} y_{A+B}^{(j)}\right)^2\right)}. \tag{6}$$

In the scalar product $y_A \cdot y_{C_i}$, the values of Y computed from $\mathbf{A}$ are multiplied by the values of Y for which all factors but Z_i are resampled while the values of Z_i remain fixed. If Z_i is noninfluential, then high and low values of y_A and y_{C_i} are randomly associated. If Z_i is influential, then high (or low) values of y_A will be preferentially multiplied by high (or low) values of y_{C_i}, increasing the value of the resulting scalar product. The reader should refer to [3] and [17] for a detailed explanation of (5) and (6).

There are no universal recipes for the choice of N. It can vary from a few hundreds to several thousands. In order to assess if the indices calculated for a given N are sufficiently stable, it is worth calculating their confidence interval. This can be easily carried out via a parametric bootstrapping. In practice, in order to calculate the sensitivity indices presented in (5) and (6), per each step of the process in the range [1, N], the term in the summation at the numerator of both equations needs to be available. Performing a parametric bootstrapping of the indices means sampling N' combinations of these terms of the same size N with replacement. The confidence interval will be then created given the distribution of the N' indices. If the confidence interval will result sufficiently small, then the number of model evaluation can be considered sufficient. For this paper, the results of the indices calculation will be presented in the graphical form.

IV. Microsimulation Case Study

The proposed methodology is applied to the identification of the parameters to be considered in the aggregate calibration of a complex traffic simulation model of a specific traffic day in an urban motorway near Porto, Portugal.

A. Microsimulation Model

To test the presented method, an advanced driver behavior model, i.e., MITSIMLab [16], was chosen. Its integrated driver behavior model presented in [16], [18], and [19] is of particular interest due to the high interaction of all advance models describing the driver behavior and the high number of parameters (101). It integrates four levels of decision making, namely, 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. This model has been successfully applied in several traffic scenarios and complex driving behaviors [20]. The description of all 101 parameters considered is presented in the Appendix.

Almost all previous calibrations of MITSIMLab considered all demand parameters [origin and destination (OD) entries] and a small subset of supply (driving behavior) parameters. This subset was typically defined based on the purpose of each calibration without any statistical analysis. SA of the MITSIMLab model were found in a couple of previous studies. In [21], an iterated OAT approach was used to a set of parameters from four specific models (car-following and free-flow acceleration, gap acceptance, and lane utility models), for the analysis of weather effects in a freeway corridor in Virginia, USA. The constant parameters (α^{CL}, α^{RL}) of the car-following model, the desired speed constant parameter (β_{ff}) of the free-flow acceleration, and the gap acceptance constant parameters (α^{lead}, α^{lag}) were found to be significant against loop sensor data (counts and speeds). In [22], experimental design techniques were carried out to test the sensibility of eight parameters of a MITSIMLab car-following model using speed, count, and density loop sensor data in a short congested corridor in California, USA. A small set of parameters of the car-following deceleration model ($h_{\mathrm{cf}}^{\mathrm{lb}}$, $\alpha_{\mathrm{cf}}^{\mathrm{dec}}$, and $\gamma_{\mathrm{cf}}^{\mathrm{dec}}$) was found as sensitive.

B. Network and Traffic Data

The network chosen for this study was the A44 road in the region of greater Porto, Portugal. It is a two-lane urban motorway with 3940 m and five main interchanges (see Fig. 2).

Located in the south bank of the Douro River, this road represents one of the main south entrances for the commuters living in the south-western region of greater Porto and to heavy vehicles heading to the main national port. Each stretch length is less than 1500 m. A44 is a dual-carriageway motorway, with two 3.50-m-width lanes, 2.00-m-width shoulders in each direction, and an additional lane in just one of the five stretches. The main section has acceleration and deceleration lanes in all interchanges, although often as short as 150 m. On- and off-ramps are connected to local roads, generally with tight curves, intersections, or pedestrian crossings, which tend to significantly reduce vehicle speeds.

The demand data were estimated using the generalized least squares (GLS) simultaneous estimation method [23] applied to 5-min aggregated daily loop counts and OD sample data from license plate matching. A simulation for the morning peak period was set up in MITSIMLab (warming period of 30 min). Since the possible measures of performance (MoPs) are all the time series of counts and speeds at the existing eight different detectors (see Fig. 2), a strategy to aggregate them in a single measure needs to be put in place. To this aim, per each simulation, we computed a measure of goodness of fit (GoF) between real and simulated time series. In order to assess the dependence from the GoF measure selected, we used three of them, namely, the root-mean-square error, the root-mean-square percentage error, and the Theil inequality coefficient U, (please prefer to [24] for their formulation and description). Eleven different outputs were considered to assess the differences between the results achieved at different locations: GoF measures computed on each single detector, on all the detectors of each road direction, and on all the detectors of the network. In total, therefore, we performed the SA of 66 different GoF measures.

For the computation of the sensitivity indices task, MITSIMLab was installed under Scientific Linux in a cluster with 80 cores with 1 GB of random access memory. This resource allowed for a faster processing of all simulations.

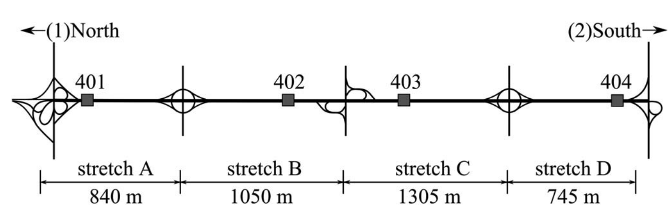

Fig. 2. Schematic of the A44 motorway.

V. Results of the SA

A. Group SA: Variance-Based SA

Fifteen groups were identified on the submodels specified in MITSIMLab (see the Appendix). The parameter grouping was based on the intrinsic formulation and the separate estimation of these submodels during MITSIMLab's development. When mapping the Sobol quasi-random sequences, parameters were assumed to be uniformly distributed with lower and upper values extracted from previous estimations ([16], [18], and [19]). A total of 34 816 ($N = 2048$) simulations (nonreplicated) were carried to compute the group sensitivity indices.

In Fig. 3, results of the group SA considering the Theil inequality coefficient as GoF measure calculated on counts and speeds from all the detectors are reported. The major figures here presented are based on this measure since it provides the most consistent and stable results [24]. It is clear that, from first-order indices (white bar), count profiles are mainly driven by the parameter combinations of Group 2 (Car Following—CF), Group 10 (Driver Heterogeneity), and Group 13 (Lane Utility Model); whereas speed profiles are also influenced by Group 1 (Reaction Time).

The model stochasticity accounts for almost 20%–25% of variance with counts and 10%–15% with speed. Just a few groups are influencing model outputs with their main effect or their interactions: Groups 2 and 10, for example, account for 80% of the output variance for the count data in the south–north direction. The higher impact of model stochasticity using counts when compared with speed suggests the use of the latter for driver behavior calibration. Different directions (different traffic conditions) showed different sensitive parameters, e.g., merging and gap-related models (Groups 5, 9, and 11) had less significance for the noncongested direction (north–south). In addition, group interactions happen to be nonnegligible in the (congested) south–north direction. This was as expected since lane change models are typically defined by several submodels (gap acceptance, gap choice, etc.) and bring more complexity to the calibration procedure.

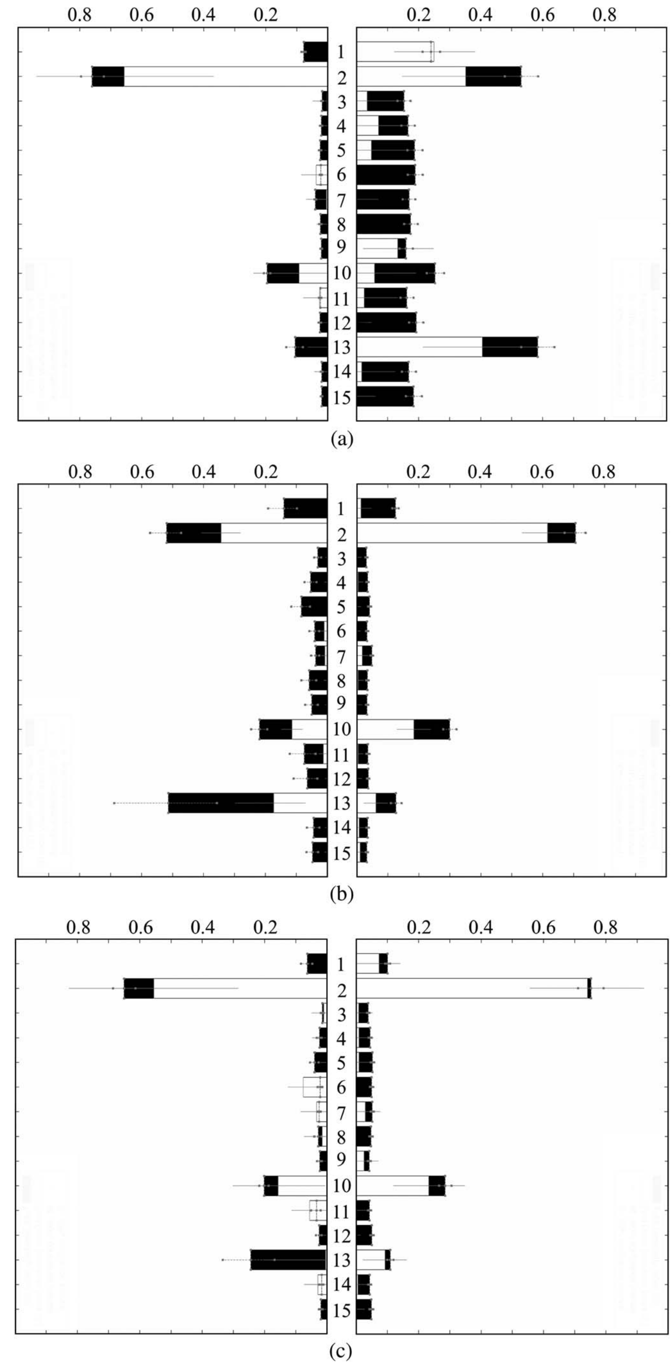

Fig. 3. Group analysis bar plots of total-order (black) and first-order (white) indices on the Theil coefficient GoF, using (left) counts and (right) speeds, with its 90% confidence intervals. (a) South–sorth direction. (b) North–south direction. (c) All sensors.

B. Final SA: Elementary Effects

Four groups influencing the speed and count profiles were selected for further analysis: Group 1 (Reaction Time), Group 2 (CF), Group 10 (Driver Heterogeneity), and Group 13 (Lane Utility Model). These groups account for 39 parameters in total, with a consequent reduction of almost two thirds in the total number of parameters. This number might be considered quite high for a comprehensive variance-based analysis, suggesting further group analysis. However, we considered the possibility of performing a variance-based SA (adopting the elementary effects approach [25]), evaluating only the total-order sensitivity indices. It is particularly useful in computationally costly mathematical models or in models with a large number of inputs. In fact, as clearly pointed out in [3], total-order indices reach stability much sooner than first-order ones, thus requiring less model evaluations. We therefore tried using a Monte Carlo experiment size of 512, resulting in 20 992 model evaluations. Five replications of each combination were considered for the analysis, and the other parameters values were set to the values obtained from the group analysis best solution.

In Fig. 4, the relatively narrow (90%) confidence intervals show the good quality of the indices estimation. It is possible to ascertain that there are six parameters outperforming all the others with regard to the output speed variance, namely, μ_{RT} [#1], $\beta_{\mathrm{cf}}^{\mathrm{acc}}$ [#8], $\alpha_{\mathrm{cf}}^{\mathrm{dec}}$ [#12], $\gamma_{\mathrm{cf}}^{\mathrm{dec}}$ [#13], $\rho_{\mathrm{cf}}^{\mathrm{dec}}$ [#14], and μ_h [#19] (from Groups 1, 2, and 10), both in counts and speed. In addition, three other parameters from the lane changing model (Group 13) account for a significant share of the total

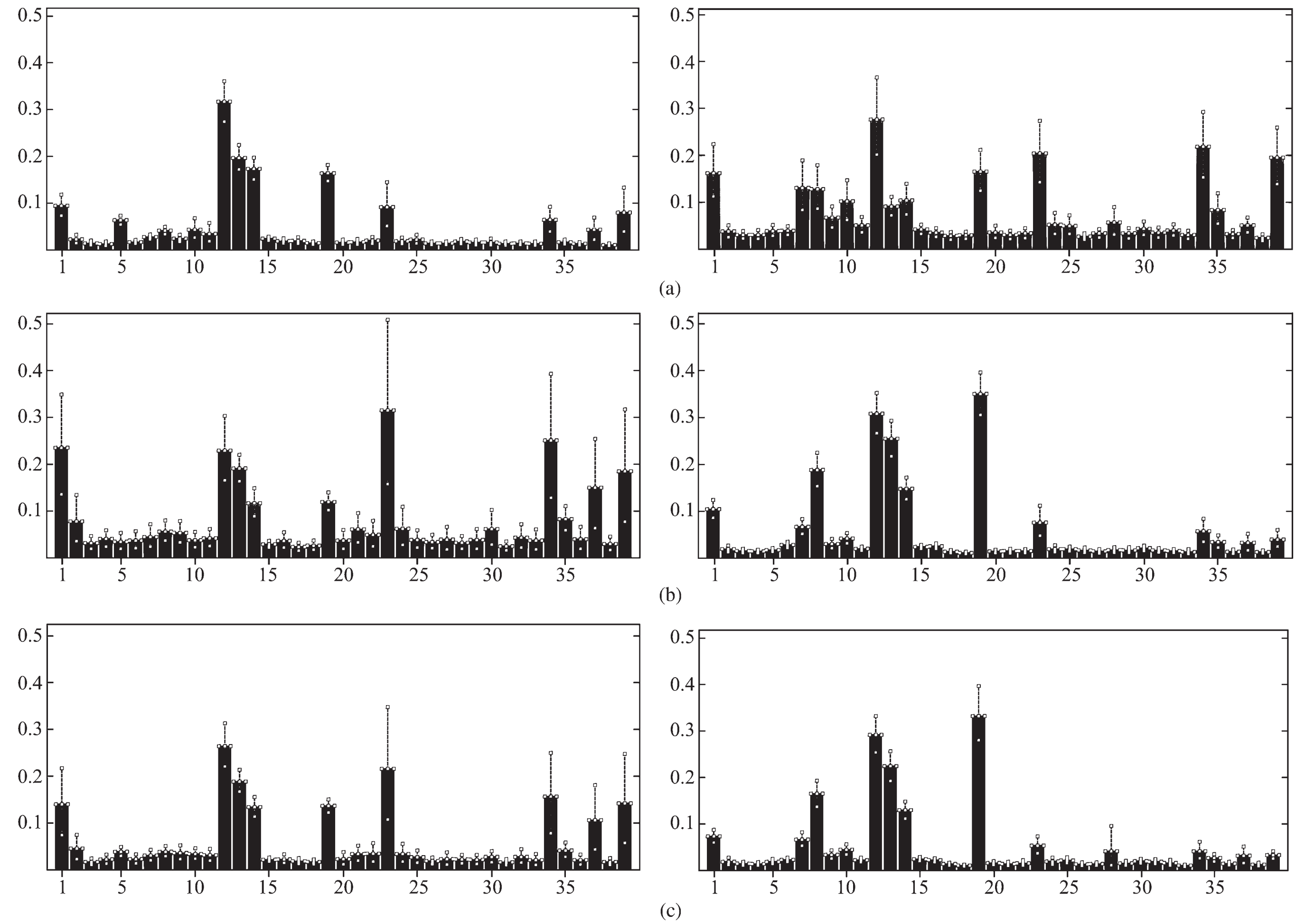

Fig. 4. Elementary analysis bar plots of total-order sensitivity indices for each of the 39 parameters on the Theil coefficient, using (left) counts and (right) speeds, with its 90% confidence intervals. (a) South–north direction. (b) North–south direction. (c) All sensors.

output variance, namely, α^{CL} [#23], $\beta_{1,i}$ [#34], and θ_{MLC} [#39] (Group 13).

Finally, the same analysis was carried out extending the simulation period to the off-peak scenario, resulting in the inclusion of the desired speed add-on parameter, i.e., β_{ff} (Group 3), which affected the free-flow acceleration model and, therefore, the loop sensor speed data.

Together with their interactions, these ten parameters are able to account for a high share (90%) of the output variance, thus sufficient to provide, once correctly estimated, a correct representation of traffic dynamics with just 10% of uncertainty.

- μ_{RT} and μ_h are the mean of the reaction time and headway threshold distributions, respectively. These are known to be important parameters, particularly when analyzing individual models separately. As expected, their share in the outputs variance is clear when analyzing the total sensitivity index, as both of them work as input to other submodels.
- $\alpha_{\mathrm{cf}}^{\mathrm{dec}}$ is the constant parameter of the CF deceleration model. It is a typical parameter considered for any calibration, and its contribution is once again exposed in this analysis.
- $\beta_{\mathrm{cf}}^{\mathrm{acc}}$ is the speed parameter in the CF acceleration model and sensitive for the noncongested speed GoF.
- $\gamma_{\mathrm{cf}}^{\mathrm{dec}}$ and $\rho_{\mathrm{cf}}^{\mathrm{dec}}$ are the gap and speed differences between the subject and the leader vehicles of the CF deceleration model. Although $\gamma_{\mathrm{cf}}^{\mathrm{dec}}$ was already found as significant in the previous SA of MITSIMLab, it is clear that both parameters should be taken into account.
- α^{CL}, $\beta_{1,i}$, θ_{MLC} are parameters of the lane changing model. The characteristics of the A44 motorway clearly conditioned this outcome, as the low number of lanes in our network configuration limits this behavior.
- β_{ff} is the constant parameter of the desired speed function. It is a typical parameter to be considered in the calibration of free-flow speeds.

C. General Remarks

The results sustained the importance of model parameters related to the deceleration regime of the car-following model, as already stated in previous studies. Other parameters usually excluded from calibration procedures appear to be quite significant. The intrinsic physical meaning of certain parameters, such as the reaction time and headway threshold distributions parameters or even the main constant parameter of the Lane Utility Model, points to the importance toward the aggregate calibration, and their total sensitivity indices results proved as such. Another important conclusion is the importance of the SA itself. The identification of calibration parameters is very sensitive to each case study configuration and observed traffic conditions. Fig. 4 shows a clear difference between GoF measures of a dense traffic scenario [see Fig. 4(a)] and a noncongested one [see Fig. 4(b)]. Similar conclusions were obtained for GoF measures disaggregated by loop sensors, where sensors near ramps revealed lane change and merging models parameters to be much more relevant for a calibration process.

The group analysis has allowed individuating the most important submodels, namely, the reaction time model, the car-following model, the lane utility model, and the drivers' heterogeneity model. In addition, it has allowed choosing among different possible measures of GoF and among different traffic measures those able to better depict specific traffic dynamics. The final SA has then been performed with the last 39 model parameters and has allowed identifying a group of ten parameters (out of 101), accounting for almost the 90% of the output's variance, with a consequent significant simplification of the calibration and estimation phase. Moreover, the methodology required 55 808 parameter combinations instead of 421 888 (−80%) otherwise required for applying variance-based techniques to the whole set of parameters.

Regarding the specific SA results, the importance of deceleration parameters along with driver specific distribution parameters was confirmed. However, these parameters should always be considered altogether as their interaction is nonnegligible. In addition, lane changing models should not be left aside in comprehensive calibration procedures, particularly when using counts as MoPs.

VI. Final Kriging-Based Calibration

The calibration of a traffic simulation model is usually based on aggregated data, namely, local counts, speeds, occupancies, or, when available, path travel times. Here, we test if such common calibration approach results in appropriate estimates of not only aggregate variables but also detailed variables typically used in behavioral and safety studies. For this purpose, MITSIMLab was calibrated for a specific weekday (May 11, 2011), and simulated detailed traffic statistics were compared with the observed ones.

In typical calibration methodologies, the number of simulations to be performed may quickly reach an unfeasible number when dealing with a large set of MoPs, GoF measures, replications, and parameter sets, representing a major obstacle in the whole problem [26].

In recent years, it has been demonstrated that the use of metamodels may significantly reduce the computational burden of the calibration and validation task of traffic simulation models [26]. By definition, a metamodel is an approximation of the input/output function defined by the simulation model. This approach has been widely used in general simulation and optimization fields and particularly suited for the purpose of our statistical validation. Thus, per each MoP/GoF combination, a surrogate of the simulation model could be computed and used for parameter calibration.

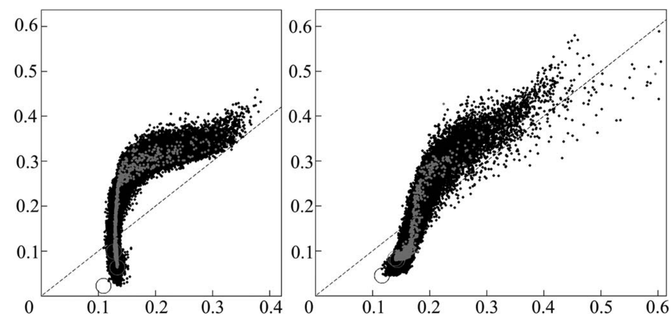

Fig. 5. Speeds (horizontal axis) versus counts (vertical axis) Theil inequality coefficient values for the MITSIMLab (gray) and Kriging (black) estimates in the (left) morning peak and (right) off-peak.

Kriging models have been used in several fields for the simulation optimization of expensive simulation models. They have been recently applied with success in the evaluation of different calibration algorithms of a microsimulation application [26], [29]. The Kriging approach was originally developed in the geostatistics field [27] and may be viewed as an estimator based on the value of neighbor points. The basics of the Kriging model can be found in [27] and [28].

The Kriging metamodels of MITSIMLab were estimated on the basis of the results of 1024 combinations of the ten sensitive parameters (ten replications per each simulation). Again, the OD for the specific weekday was estimated using the GLS simultaneous estimation method [23]. The Kriging estimation was based on the publicly available MATLAB toolbox DACE [30]. The estimated metamodels were then used to explore the objective function of the calibration problem, by using its output to compute several thousand GoF measures of different parameter combinations. Fig. 5 shows scatterplots of how both counts and speed toward a minimum in the Theil inequality coefficient as the parameter combinations vary. They also show that while the Kriging model is almost able to nullify the distance between real and simulated counts, this is not the case for the average speed. This is due to the prior GLS estimation of the demand parameters (OD matrix), which have a direct influence on loop sensor count measurements.

For what concerns the final calibration approach, in the authors' opinion, a single best solution should not be defined for several reasons: the Kriging approximation might not capture small changes existing in the true model, a single best option may easily change considering the variability of traffic data measurements, and the best solution for speed-related GoF may not be the best for counts-related GoFs. For these reasons, the 30 best sets of parameter combinations with comparable (speed) GoF performances were kept for the validation test, and ten replications of each were carried out in MITSIMLab for stochasticity control. The number of combinations was set based on a user-predefined GoF threshold for the generic full-day calibration ($U \leq 0.085$).

A. Aggregated Data From Loop Sensors

The 30 best combinations obtained from the SA-based calibration presented in the previous sections managed to appropriately replicate the observed loop sensor counts and speeds.

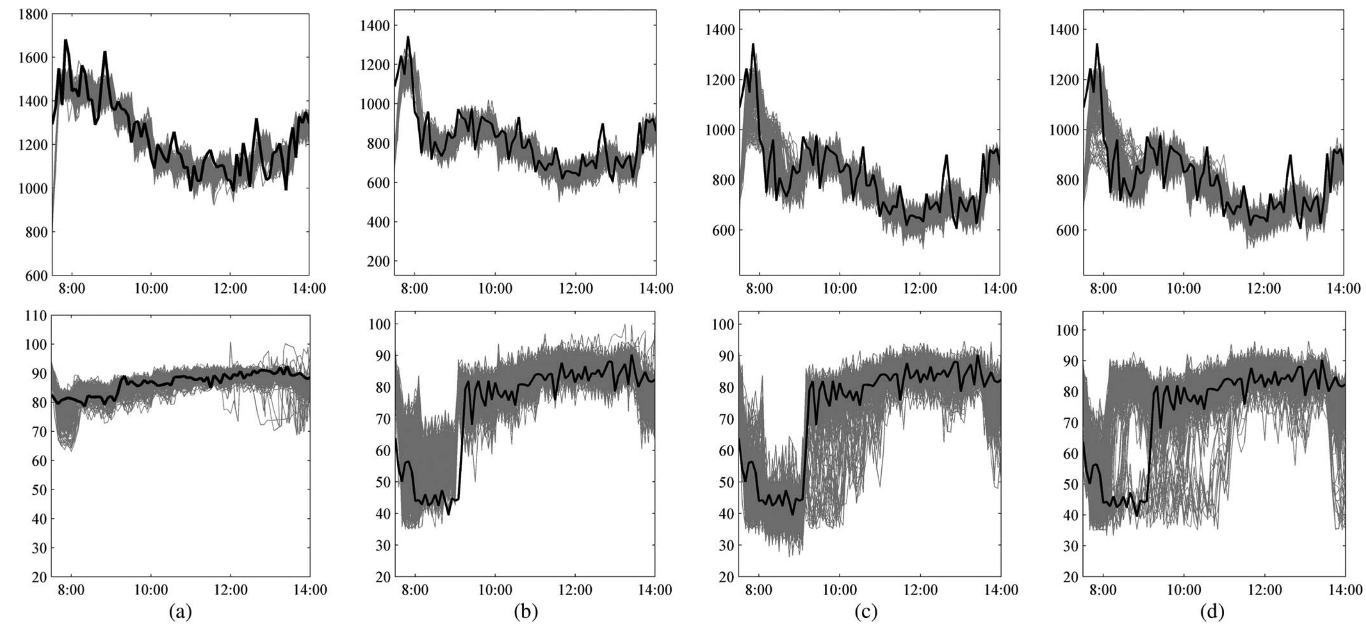

Fig. 6. Simulated (gray) versus true (dark) (top) counts and (bottom) average speed. (a) Full network data for the best parameter set of full-day calibration (b) Loop 401-1 data for the best parameter set of full day calibration (c) Loop 401-1 data for the best parameter set of AM peak calibration (d) Loop 401-1 data for the best parameter set of off peak calibration.

Total loop sensor counts and average speeds in the entire network showed a good and stable fit along the daily variations [see Fig. 6(a)], with a Theil coefficient of 0.083 for a generic calibration of a full day using all sensor data. However, the selection of different MoPs will always have a strong impact on the simulation results. As expected, local calibration resulted in better local fitting results ($U = 0.066$, for a calibration of the MoP using sensors in the south–north direction during the morning peak period, and $U = 0.056$ for the calibration of the MoP using sensors in the north–south direction during the off-peak period), but failed to capture traffic characteristics in noncalibrated scenarios. In Fig. 6(b)–(d), for example, the different calibration results for sensor 401-1 (in the south–north direction) using a full network MoP calibration and two calibrations using the MoPs based on the sensor measurements in the peak and in the off-peak are presented.

B. Detailed Data From Trajectories

It is common practice by practitioners and even researchers to use simulators calibrated with aggregated data to extract further detailed traffic information from the transportation system under analysis. A common example is the extraction of accelerations and surrogate safety measures for vehicle emissions and safety modeling, respectively. This practice is generally wrong, particularly when the detailed variables or driving behaviors scrutinized are far from the ones used in the original model estimation and the calibration process. It might be the case where the appropriate conditions are met, but one should always test these simulation outputs against their real counterparts. To this aim, simulated trajectories obtained using the aforementioned 30 best parameter combinations for the off-peak period were compared with real trajectories collected on site through aerial remote sensing [31]. For consistency, the model was calibrated with the proposed method for the specific day of trajectory extraction. Cumulative distribution functions (cdfs) of a set of six detailed variables, namely, speed, headway, acceleration, deceleration, and two safety-related surrogate measures, i.e., the time to collision (TTC) and the deceleration rate to avoid crash (DRAC), were extracted (see Fig. 7).

For the entire A44 motorway, including its entry and exit links, it is clear that the majority of the detailed variables could not be appropriately simulated. Although loop sensor speeds were used for aggregate calibration, speeds on acceleration and deceleration lanes, ramps, and access links are far from being replicated. In Fig. 8, it is clear that under dense traffic conditions, both speed and headway for specific road sections were considerably underestimated by the model.

Fig. 9 shows a very good fit of simulated speeds and headways in a very specific road section group and under light traffic conditions. In fact, 37% of the loop sensor observations belong in such specific groups, resulting in a much better fitting of speeds.

Finally, the simulated accelerations and safety-related surrogate measures cannot be used without their appropriate calibration using their on-site counterparts.

VII. Conclusion and Discussion

The common approach usually adopted in calibration of simulation models consists in solving an optimization problem in which the distance between some measured and simulated traffic measures is minimized by changing the value of their parameters. This approach has several shortcomings as it does not take into account the high level of uncertainty in the traffic demand and in the simulation model parameters. In addition, the entire problem is made more complex as what is measured is just one of the possible traffic realizations due to the same

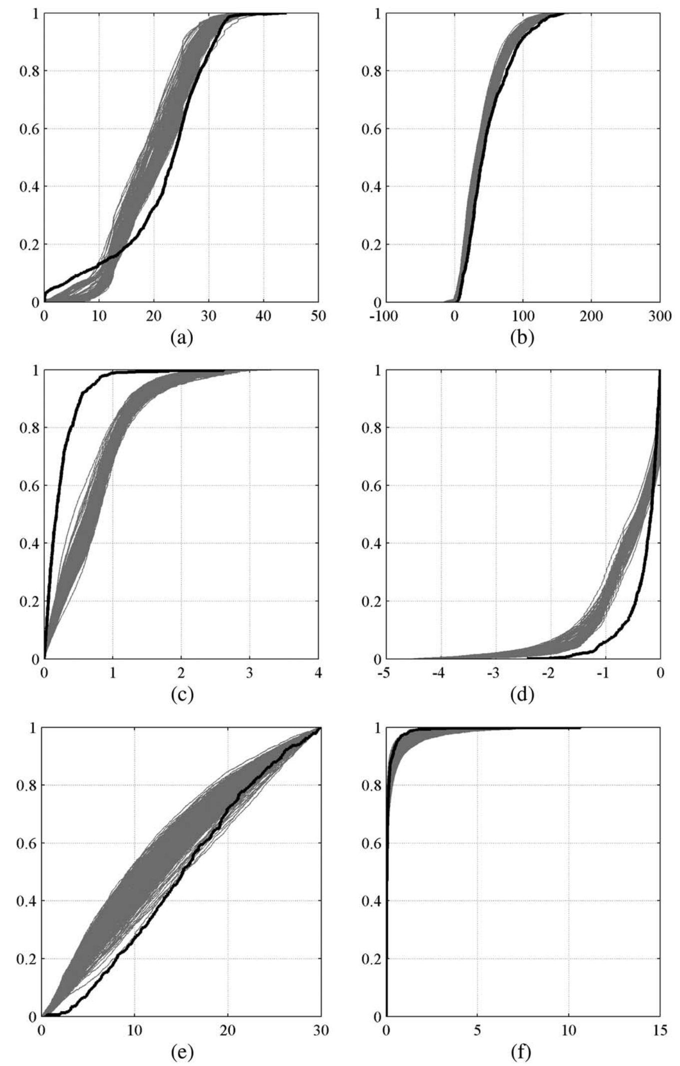

Fig. 7. cdfs for different detailed traffic variables. (a) Speed$(\mathrm{m/s})$ (b) Headway(m) (c) Acceleration$(\mathrm{m/s}^2)$ (d) Decelaration$(\mathrm{m/s}^2)$ (e) TTC (s) (f) DRAC$(\mathrm{m/s}^2)$.

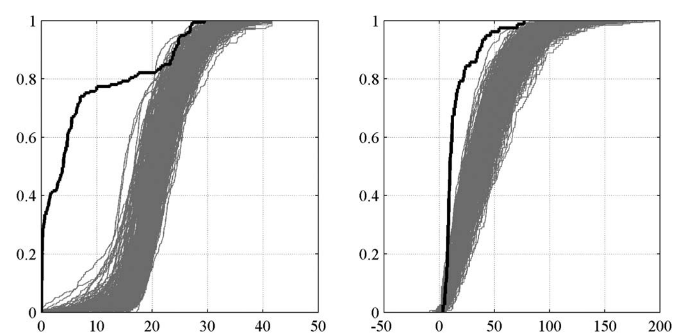

Fig. 8. (Left) Speed and (right) headway cdfs in the center lane of three-lane sections (with acceleration lane) (grade higher than $+2\%$, speed limit over 100 km/h, and dense traffic). Negative simulated headways are due to biased estimation of simulated vehicle lengths.

demand, supply, and composition of the population (e.g., due to the changes in the departing times, etc.).

In this paper, the entire calibration problem has been formulated under the light of uncertainty management. In this respect, the SA is crucial to individuate the most important sources of modeling uncertainties. A multistep approach for the SA of computationally expensive and high-dimensional traffic simulation models is presented and applied to the MITSIMLab model. Then, rather than finding the parameter combination that best reproduces the real measures, one should look at a number of combinations for which the model behaves relatively well. At this point, the entire model uncertainty due to the different combinations and to the model stochasticity should be considered (and not fixed as with the calibration) in the form of distributions of the model outputs. In this framework, the validation of the model should be carried out by checking whether the individuated uncertainty is sufficient to cover the uncertainty of the real system.

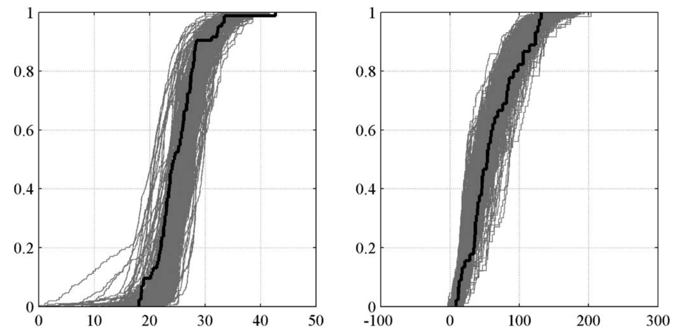

Fig. 9. (Left) Speed and (right) headway cdfs in the right lane of two-lane sections (grade higher than $+2\%$, speed limit over 100 km/h, and light traffic).

Results show the robustness and usefulness of the approach, by which it was possible to save as much as the 80% of model evaluations without necessarily paying in terms of result accuracy.

Many aspects of this novel view need further research. The evaluation of the best combination set selection criteria and the different possible grouping structures with regard to the final parameter values should be carried out. Other types used in different MoP formulations should be also tested. In addition, alternative SA methods and screening methods at each step of the proposed methodology should be studied, particularly when relevant variance-based designs still rely in an excessive number of simulations. Finally, for computational efficiency assessment, the proposed methodology should be applied to different simulation model formulations. However, the results presented throughout this paper might aim at slightly modifying how the reliability of traffic simulation is perceived.

APPENDIX
INTEGRATED DRIVING BEHAVIOR MODEL

The integrated driving behavior model proposed by [16] is integrated in the microscopic traffic simulator MITSIMLab [18], [19]. A brief review of all the models considered in the current analysis is here presented. The reader should, however, refer to [16], [18], and [19] for its full description. All the parameters considered for potential calibration were classified in the following 15 different groups.

1) *Reaction time*: When a new vehicle enters the network, it is randomly assigned an update step size, which specifies the frequency with which drivers update their driving

behavior. This value is drawn from a truncated normal distribution with mean, standard deviation, and lower and upper bounds, i.e., μ_{RT}, σ_{RT}, and $\mathrm{lb}_{\mathrm{RT}}$ and $\mathrm{ub}_{\mathrm{RT}}$, respectively (Group 1, four parameters).

2) *Car-following acceleration*: Different models describe the acceleration behavior under the various situations. The stimulus-sensitivity framework [32] was adapted for all the acceleration models considered in MITSIMLab. The car-following model, for both acceleration and deceleration ($g \in \{\mathrm{acc}, \mathrm{dec}\}$) is given by

$$a^g_{\mathrm{cf}}(t) = a^g_{\mathrm{cf}} \left[V(t)^{\beta^g_{\mathrm{cf}}} \Delta\chi(t)^{-\gamma^g_{\mathrm{cf}}} \right] \times k(t)^{\delta^g_{\mathrm{cf}}} \Delta V(t-\tau_{\mathrm{RT}})^{\rho^g_{\mathrm{cf}}} + \varepsilon^g_{\mathrm{cf}}(t) \quad \text{(A1)}$$

where V is the speed of the subject vehicle; Δx and ΔV are the gap and speed differences between the lead and subject vehicles, respectively; k is the traffic density downstream of the subject vehicle; τ_{RT} is its driver reaction time; and $\varepsilon^g_{\mathrm{cf}}$ is the random error term. The car-following state depends on the headway between the subject and the front vehicle. In MITSIMLab, each vehicle has its own headway threshold (see Group 10) However, general thresholds such as the minimum response distance $\Delta x^{\min}_{\mathrm{cf}}$ and the general headway lower bound $h^{\mathrm{lb}}_{\mathrm{cf}}$ are also considered for this specific model (Group 2, 11 parameters).

3) *Free-flow acceleration*: When the headway between the subject and the lead vehicle is big enough and the subject vehicle speed is higher than a threshold $V_{\min}$, the vehicle free-flow acceleration is given by

$$a_{\mathrm{ff}} = \alpha_{\mathrm{ff}} \left[V_{\mathrm{DS}}(t-\tau_{\mathrm{RT}}) - V(t-\tau_{\mathrm{RT}}) \right] + \varepsilon_{\mathrm{ff}}(t) \quad \text{(A2)}$$

$$V_{\mathrm{DS}}(t) = \beta_{\mathrm{ff}} + V_{\mathrm{lim}} + \gamma_{\mathrm{ff}} \delta^h_s + \delta^k_{\mathrm{ff}} k(t) + \rho^{V_f}_{\mathrm{ff}} V_f(t) \quad \text{(A3)}$$

where V is the speed of the subject vehicle, V_{DS} is the desired speed of the subject vehicle, V_{lim} is the local speed limit, δ^h_s is 1 if the subject vehicle is heavy and 0 otherwise, k is 1 if the traffic density downstream is equal or less than a threshold θ_{ff} and 0 otherwise, $V_l(t)$ is the front vehicle speed, τ_{RT} is the driver reaction time, and $\varepsilon_{\mathrm{ff}}(t)$ is the random error term (Group 3, seven parameters).

4) *Merging*: When a vehicle reaches a lane dropping area, it may be tagged as a merging vehicle. The acceleration is then calculated relaxing the car-following gap limitation and restricting overtaking when using the dropping lane. Upstream (Δx_u) and downstream (Δx_d) lengths from the beginning section of the dropping lane set the total area where a vehicle can be tagged as merging. The probability of being tagged is given by p_0 and only if the number of vehicles in the merging area is less than $n_{\max}$ (Group 4, four parameters).

5) *Mandatory lane change state* (MLC): MLC is derived from previous models of MITSIMLab [18]. When the general lane changing model proposed in [16] cannot be applied due to the lack of acceptable gaps (dense traffic conditions), an MLC may be initiated, limiting the lane alternatives in the lane choice and gap acceptance models. A vehicle may switch to the MLC state only if its current lane is ending or does not connect to the next link in its path. The probability of initiation of such state is derived from the following equation when the distance to the downstream node is less than $\Delta x_{\min}$:

$$P^{\mathrm{MLC}} = \exp\left(\frac{-\Delta x^2}{\left(\alpha^{\mathrm{MLC}}_0 + \alpha^{\mathrm{MLC}}_{n_{\mathrm{lc}}} n_{\mathrm{lc}}(t) + \alpha^{\mathrm{MLC}}_k k(t)\right)^2} \right) \quad \text{(A4)}$$

where Δx is the distance to the downstream node limited by the lower bound Δx_{lb}, n_{lc} is the number of lane changes required to reach the target lane, and k is the lane density. $\Delta t_{\min}$ is an additional parameter that sets the minimum time in the lane when tagged for MLC (Group 5, five parameters).

6) *Yielding probability*: When a vehicle is in nosing state, the lag vehicle is set to yielding with probability p_{no} if it was not previously yielding and p_{yes} otherwise (Group 6, two parameters).

7) *Nosing probability*: When a vehicle has decided to change lanes and is in MLC state, a merging model that captures merging by gap creation, either through courtesy yielding of the lag vehicle or nosing of the subject vehicle, may be applied. The probability of a subject vehicle being set to the nosing state is given by

$$P^{\mathrm{nos}} = 1 \Big/ \left[1 + \exp\left(\alpha^{\mathrm{nos}} + \beta^{\mathrm{nos}}_{\Delta V\_} \Delta V\_(t) + \beta^{\mathrm{nos}}_{I_{\Delta\chi}} I_{\Delta\chi}(t) + \beta^{\mathrm{nos}}_{l_{\mathrm{gap}}} l_{\mathrm{gap}}(t) + \beta^{\mathrm{nos}}_{n_{\mathrm{lc}}} n_{\mathrm{lc}}(t) \right) \right] \quad \text{(A5)}$$

where $\Delta V\_$ is the relative speed between the subject vehicle and the lead vehicle on the target lane, $I_{\Delta x}$ is an impact factor depending on both the remaining distance to the point at which the lane change must be completed and on a parameter λ^{nos}, l_{gap} is the total gap length, and n_{lc} is the number of lane changes required to reach the target lane (Group 7, six parameters).

8) *Nosing rules*: The application of the nosing model is also restricted by a maximum waiting time before nosing $t^{\mathrm{nos}}_{\max}$, a maximum and a minimum distance for nosing $\Delta x^{\mathrm{nos}}_{\max}$ and $\Delta x^{\mathrm{nos}}_{\min}$, and a maximum yielding time $t^{\mathrm{yield}}_{\max}$ for the lag vehicle (Group 8, four parameters).

9) *Courtesy yielding probabilities*: They are modeled as fixed probabilities. p^{cyield}_0, p^{cyield}_1, p^{cyield}_2, and p^{cyield}_3 are the probabilities to yield to 0, 1, 2, and 3 vehicles when tagged as MLC (Group 9, four parameters).

10) *Driver heterogeneity*: The acceleration model error terms for the car-following and free-flow behavior follow a normal distribution with mean, zero, and standard deviation $\sigma^{\mathrm{acc}}_{\mathrm{cf}}$, $\sigma^{\mathrm{dec}}_{\mathrm{cf}}$, and σ_{ff}, respectively; the headway threshold, which rules the choice between car-following and free-flow acceleration models, is obtained from a truncated normal distribution with parameters μ^h_{dv} and σh^h_{dv} and lower and upper bounds $\mathrm{lb}^h_{\mathrm{dv}}$ and $\mathrm{ub}^h_{\mathrm{dv}}$ (Group 10, seven parameters).

11) *Target gap acceleration*: Captures the behavior of drivers who target a lane change and already chose the corresponding target gap. This formulation is part of the model proposed in [16], i.e.,

$$\alpha^{\mathrm{TG}} = \alpha_g^{\mathrm{TG}} \left[D^{\mathrm{TG}}(t-\tau)\beta_D^{\mathrm{TG}} \cdot \exp\left(\beta_{\Delta V_+}^{\mathrm{TG}} \Delta V_+^{\mathrm{TG}}(t)\right) \cdot \exp\left(\beta_{\Delta V\_}^{\mathrm{TG}} \Delta V\_^{\mathrm{TG}}(t)\right) + \varepsilon_g^{\mathrm{TG}}(t)\right] \quad \text{(A6)}$$

where D^{TG} is the distance to the desired position for the target gap and has different formulations for each of the possible TG $\in \{\text{backward, adjacent, forward}\}$; ΔV_+^{TG} and $\Delta V_-^{\mathrm{TG}}(t)$ are the positive and negative relative target lane leader speeds, respectively; τ is the driver reaction time; and $\varepsilon^{\mathrm{TG}}(t) \sim N(0, (\sigma^{\mathrm{TG}})^2)$ is the random error term (Group 11, 13 parameters).

12) *Gap acceptance model*: It evaluates the adjacent gaps in the target lane model and decides to switch lanes immediately or not. The adjacent gap is split into lead and lag gaps, which both need to be acceptable for the lane change action. A gap is acceptable if it is greater than the corresponding critical gap, whose mean is modeled as a random variable following a lognormal distribution, i.e.,

$$\ln\left(G_n^{l,\mathrm{cr}}(t)\right) = \alpha^l + \beta_{\Delta V_+}^l \Delta V_+^l(t) + \beta_{\Delta V\_}^l \Delta V_-^l + \beta^{\mathrm{EMU}}\mathrm{EMU}^l(t) + \alpha_v^l v_n + \varepsilon^l(t) \quad \text{(A7)}$$

where $G_n^{l,\mathrm{cr}}$ is the critical l ($l \in \{\text{lead, lag}\}$) gap; ΔV_+^l and ΔV_-^l are the positive and negative speed differences, respectively, between the subject vehicle and the l vehicle on the target lane limited by a threshold $\Delta V_{\max}$; EMU^l is the expected maximum utility of the target gap l; v_n is the individual specific error term; and $\varepsilon^l \sim N(0, (\sigma^l)^2)$ is the random error term (Group 12, eight parameters).

13) *Lane utility model*: At the top of the drivers' decision tree is the lane choice model. Modeled as a discrete choice problem, the probability of choosing a target lane is computed through a logit formulation using the following utility function:

$$\begin{aligned} U^{\mathrm{TL}} = {} & \alpha^{\mathrm{TL}} + \beta_{\mathrm{RML}}^{\mathrm{TL}}\delta_{\mathrm{RML}} + \beta_{V_l}^{\mathrm{TL}} V_l(t) + \beta_{\Delta\chi}^{\mathrm{TL}}\Delta\chi(t) + \beta_b\delta_b \\ & + \beta_h\delta_h(t) + \beta_k\delta_k(t) + \beta_{\mathrm{tail}}^{\mathrm{TL}}\delta_{\mathrm{tail}} + [\Delta\chi_{\mathrm{exit}}(t)]^{\theta_{\mathrm{MLC}}} \\ & \times \sum\left(\beta_{\mathrm{nlc},i}\delta_{\mathrm{nlc},i}(t)\right) + \beta_{\mathrm{next}}\delta_{\mathrm{next}}(t) + \beta_{\mathrm{add}} n_{\mathrm{add}}(t) \\ & + \beta_{\mathrm{gap}}\mathrm{EMU}^{\mathrm{TL}}(t) + \alpha_v^{\mathrm{TL}} v_n + \varepsilon^{\mathrm{TL}}(t) \end{aligned} \quad \text{(A8)}$$

where α^{TL} is a constant parameter for the target lane TL $\in \{\text{left, current, right}\}$, δ_{RML} is a dummy variable equal to 1 if TL is the rightmost lane, V_l is the speed of the lead vehicle on TL, Δx is the gap between the lead and subject vehicles, δ_h is a dummy equal to 1 if the traffic density in TL is higher than a threshold k_{ceil}, δ_b and δ_h are dummy variables equal to 1 on the presence of bus and heavy good vehicles in TL, δ_{tail} is a dummy variable that captures the drivers' tendency to move out of their current lane if they are being tailgated and is equal to 1 if the backward gap is less than $\Delta x_{\mathrm{floor}}^{\mathrm{back}}$, Δx_{exit} is the distance from the subject vehicle to the next exit, $\delta_{\mathrm{nlc},i}$ are i dummy variables equal to 1 for each i number of lane changes required to reach TL, δ_{next} is a dummy for the need of exiting on the next off-ramp; n_{add} is a dummy for the number of lane changes required from the TL to the off-ramp, $\mathrm{EMU}^{\mathrm{TL}}$ is the maximum utility of the available gaps in TL given by the target gap model, v_n is the individual specific error term capturing correlations between observations, and $\varepsilon^{\mathrm{TL}}$ is the random error (Group 13, 17 parameters).

14) *Target gap model*: When a driver has decided to switch lanes, the target gap model captures the drivers' intention on the lane changing decision process, when the adjacent gap is rejected. The subject vehicle will then adjust its speed and position depending on the chosen target gap. Similarly to the lane choice model, the probability of choosing a target gap is modeled as a logit model using the following utility equation:

$$\begin{aligned} U^{\mathrm{TG}} = {} & \alpha^{\mathrm{TG}} + \beta_{\Delta\chi_{\mathrm{TG}}}\Delta\chi_{\mathrm{TG}}(t) + \beta_{l_{\mathrm{TG}}} l_{\mathrm{TG}}(t) + \beta_{\delta_f}\delta_f(t) \\ & + \beta_{\Delta V_{\mathrm{TG}}}\Delta V_{\mathrm{TG}}(t) + \alpha_v^{\mathrm{TG}} v_n + \varepsilon^{\mathrm{TG}}(t) \end{aligned} \quad \text{(A9)}$$

where Δx_{TG} is the distance to the target gap TG $\in \{\text{backward, adjacent, forward}\}$; l_{TG} is the effective gap length; δ_f^{TG} is a dummy for the presence of a front vehicle on the current lane; ΔV_{TG} is the relative gap speed; v_n is the individual specific error; and $\varepsilon^{\mathrm{TG}}$ is the random error (Group 14, six parameters).

15) *Demand stochasticity*: The OD matrix is a key input on the variability of the simulation output. In this paper, the common stochasticity of the OD matrix was analyzed by considering a common variance (σ_{OD}^2) for all OD paths and a distribution factor (β_{OD}), which determines the percentage of vehicles departing randomly (Group 15, two parameters).

ACKNOWLEDGMENT

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).

REFERENCES

[1] K. Koskinen, I. Kosonen, T. Luttinen, A. Schirokoff, and J. Luoma, "Development of a nanoscopic traffic simulation tool," *Adv. Transp. Stud.—Int. J.*, vol. 17, pp. 89–96, 2009.

[2] K. Basak, S. Hetu, Z. Li, C. Lima Azevedo, H. Loganathan, T. Toledo, R. Xu, Y. Xu, L. S. Peh, and M. E. Ben-Akiva, "Modeling reaction time within a traffic simulation model," in *Proc. 92nd Annu. Meet. Transp. Res. Board*, Washington, DC, USA, Jan. 2013, p. 19.

[3] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis: The Primer*, 1st ed. Hoboken, NJ, USA: Wiley, Feb. 2008.

[4] N. E. Lownes and R. Machemehl, "Vissim: A multi-parameter sensitivity analysis," in *Proc. Winter Simul. Conf.*, Monterey, CA, USA, 2006, pp. 1406–1413.

[5] T. V. Mathew and P. Radhakrishnan, "Calibration of microsimulation models for non lane-based heterogeneous traffic at signalized intersections," *J. Urban Plann. Develop.*, vol. 136, no. 1, pp. 59–66, Mar. 2010.
[6] A. Kesting and M. Treiber, "Calibrating car-following models using trajectory data: Methodological study," *Transp. Res. Rec.*, vol. 2008, pp. 148–156, Mar. 2008.
[7] A. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, Dec. 1999.
[8] B. Bartin, K. Ozbay, O. Yanmaz, and N. Rathi, "Modeling and simulation of an unconventional traffic circle," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 13–15, 2005, pp. 384–389.
[9] Z. Li, H. Liu, and K. Zhang, "Sensitivity analysis of paramics based on 2K-P fractional factorial design," in *Proc. Int. Conf. Transp. Eng.*, 2009, pp. 3633–3638.
[10] A. Beegala, J. Hourdakis, and P. Michalopoulos, "Methodology for performance optimization of ramp control strategies through microsimulation," *Transp. Res. Rec.*, vol. 1925, pp. 87–98, Jan. 2005.
[11] B. Ciuffo, V. Punzo, and V. Torrieri, "Comparison of simulation-based and model-based calibrations of traffic-flow microsimulation models," *Transp. Res. Rec.*, vol. 2088, pp. 36–44, 2008.
[12] B. Park and H. Qi, "Development and evaluation of a procedure for the calibration of simulation models," *Transp. Res. Rec.*, vol. 1934, pp. 208–217, Jan. 2005.
[13] V. Punzo and B. Ciuffo, "How parameters of microscopic traffic flow models relate to traffic dynamics in simulation," *Transp. Res. Rec.*, vol. 2124, pp. 249–256, 2009.
[14] Y. Hollander and R. Liu, "The principles of calibrating traffic microsimulation models," *Transportation*, vol. 35, no. 3, pp. 347–362, Mar. 2008.
[15] I. M. Sobol, "Uniformly distributed sequences with an additional uniform property," *USSR Comput. Math. Math. Phys.*, vol. 16, no. 5, pp. 236–242, 1976.
[16] T. Toledo, H. N. Koutsopoulos, and M. Ben-Akiva, "Integrated driving behavior modeling," *Transp. Res. C*, vol. 15, no. 2, pp. 96–112, Apr. 2007.
[17] F. Campolongo, J. Cariboni, and A. Saltelli, "From screening to quantitative sensitivity analysis: A unified approach," *Comput. Phys. Commun.*, vol. 182, no. 4, pp. 978–988, Apr. 2001.
[18] Q. Yang, H. N. Koutsopoulos, and M. Ben-Akiva, "Simulation laboratory for evaluating dynamic traffic management systems," *Transp. Res. Rec.*, vol. 1710, pp. 122–130, 2000.
[19] K. I. Ahmed, "Modeling drivers' acceleration and lane changing behaviors," Ph.D. dissertation, Dept. Civil Environ. Eng., Mass. Inst. Technol. (MIT), Cambridge, MA, USA, 1999.
[20] V. Vaze, C. Antoniou, Y. Wen, and M. Ben-Akiva, "Calibration of dynamic traffic assignment models with point-to-point traffic surveillance," *Transp. Res. Rec.*, vol. 2090, pp. 1–9, 2009.
[21] E. Sterzin, "Modeling influencing factors in a microscopic traffic simulator," M.S. thesis, Dept. Civil Environ. Eng., Mass. Inst. Technol., Cambridge, MA, USA, 2004.
[22] M. Kurian, "Calibration of a microscopic traffic simulator," M.S. thesis, Dept. Civil Environ. Eng., Mass. Inst. Technol., Cambridge, MA, USA, 2000.
[23] E. Cascetta, D. Inaudi, and G. Marquis, "Dynamic estimation of origin–destination matrices using traffic counts," *Transp. Sci.*, vol. 27, no. 4, pp. 363–373, 1993.
[24] Y. Hollander and R. Liu, "The principles of calibrating traffic microsimulation models," *Transportation*, vol. 35, no. 3, pp. 347–362, May 2008.
[25] M. D. Morris, "Factorial sampling plans for preliminary computational experiments," *Technometrics*, vol. 33, no. 2, pp. 161–174, May 1991.
[26] B. Ciuffo and V. Punzo, "'No free lunch' theorems applied to the calibration of traffic simulation models," *IEEE Transact. Intell. Transp. Syst.*, issue 99, pp. 1–10, 2013.
[27] G. Matheron, "Principles of geostatistics," *Econom. Geol.*, vol. 58, no. 8, pp. 1246–1266, 1963.
[28] J. P. C. Kleijnen, "Kriging metamodeling in simulation: A review," *Eur. J. Oper. Res.*, vol. 192, no. 3, pp. 707–716, Feb. 2009.
[29] B. Ciuffo, J. Casas, M. Montanino, J. Perarnau, and V. Punzo, "From theory to practice: Gaussian process meta-models for sensitivity analysis of traffic simulation models: Case study of Aimsun mesoscopic model," forthcoming in *Transp. Res. Rec.*, 2014.
[30] S. Lophaven, H. Nielsen, and J. Sondergaard, DACE: A MATLAB Kriging toolbox, IMM—Technical University of Denmark, vol. 18, Kgs. Lyngby, Denmark 2002.
[31] C. Lima Azevedo, J. Cardoso, M. Ben-Akiva, J. P. Costeira, and M. Marques, "Automatic vehicle trajectory extraction by aerial remote sensing," presented at the 16th Euro Working Group Transportation, Procedia, Soc. Behavioral Sciences, Porto, Portugal, 2013.
[32] D. Gazis, R. Herman, and R. Rothery, "Nonlinear follow-the-leader models of traffic flow," *Oper. Res.*, vol. 9, no. 4, pp. 545–567, Jul./Aug. 1961.

Biagio Ciuffo received the Ph.D. degree in hydraulic, transportation, and land use systems engineering from the University of Naples Federico II, Naples, Italy, in 2008.

He held a three-year postdoctoral position with the Joint Research Centre (JRC), European Commission, Ispra, Italy, working at the sustainability assessment of traffic- and transport-related measures and policies. He is currently an Official of the JRC. He has authored over 30 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, emission estimation from maritime traffic, sustainable development, and sustainability appraisal.

Dr. Ciuffo serves as a reviewer for several international journals and conferences. He is currently part of the European Cooperation in Science and Technology Action "Methods and tools for supporting the use, calibration and validation of traffic simulations models," as a Leader of one working group. For his research activities on the calibration of simulation models, he was the recipient of the 2012 Greenshields Prize from the Transportation Research Board of the U.S. National Academy of Science.

Carlos Lima Azevedo received the B.Sc. degree in structural engineering in 2004 and the M.Sc. degree in transportation engineering in 2008 from Instituto Superior Técnico, Lisbon, Portugal, where he is currently working toward the Ph.D. degree in transportation systems, under the MIT Portugal Program and in partnership with the National Laboratory for Civil Engineering (LNEC).

From 2009 to 2010, he was a Visiting Student with the Intelligent Transportation Systems Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. During the summer of 2011, he was a Visiting Researcher with Singapore-MIT Alliance for Research and Technology. Since 2005, he has been a Research Scholar with LNEC, Lisbon. His main research interests are transportation modeling, traffic simulation, behavioral research, traffic flow theory, and road safety.