Calibration of Driving Behavior Models using Derivative-Free Optimization and Video Data for Montreal Highways

Laurent Gauthier, ing. jr., Master's Student (Corresponding author)
Department of Civil, Geological and Mining Engineering
Polytechnique Montréal, C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7
Phone: +1 (514) 340-4711 ext. 4210
Email: laurent-2.gauthier@polymtl.ca

Nicolas Saunier, ing., Ph.D., Associate professor
Department of Civil, Geological and Mining Engineering
Polytechnique Montréal, C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7
Phone: +1 (514) 340-4711 ext. 4962
Email: nicolas.saunier@polymtl.ca

Sébastien Le Digabel, ing., Ph.D., Associate professor
Groupe d'étude et de recherche en analyse des décisions (GERAD)
Department of Mathematical and Industrial Engineering
Polytechnique Montréal, C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7
Phone: +1 (514) 340-4711 ext. 4291
Email: sebastien.le-digabel@polymtl.ca

Gang Cao, ing. M.Sc.A.
Direction des transports
Ville de Montréal, 801 Brennant, suite 600
Montréal (Québec) Canada H3C 0G4
Phone: +1 (514) 872-5994
Email: gang.cao@ville.montreal.qc.ca

Word count

Text	5739
Tables (3 X 250)	750
Figures (4 X 250)	1000
Total	7489

Date of submission: **August 1st, 2015**

ABSTRACT

Traffic simulation software is commonly used for traffic impact assessment in transportation projects, allowing the selection of adequate solutions to existing complex problems without having to test them in the field. Such software is highly adaptable to different road conditions and driving behaviours by letting engineers modify a large number of parameters. However, this flexibility requires the model to be calibrated for each application in different regions, conditions and settings. Calibration requires data, which can be time-consuming and costly to collect. We propose a calibration procedure for the driving behavior models, which describe how vehicles interact with each other. These calibrated behaviors should be generic for the region regardless of the specific site geometry and the proposed procedure seeks to allow this generalisation by allowing simultaneous simulations on my many networks. To achieve this calibration, a state-of-the-art derivative free optimization algorithm, the mesh-adaptive direct-search algorithm, is used to fit simulations to real world microscopic data acquired via automated video analysis. We propose an implementation of this procedure for the Wiedemann 99 model in the VISSIM traffic micro-simulation software in a case study for the City of Montreal using data collected on a major Montreal highway.

INTRODUCTION

The use of micro-simulation software is now widespread in the traffic engineering practice to evaluate the impact of projected changes to a traffic network such as additions or replacements of existing infrastructure, impacts of a construction site, optimization of traffic light control, etc. The quality of these evaluations relies on the ability of these simulations to adequately represent traffic conditions on the studied network. For VISSIM, one of the most popular traffic micro-simulation pieces of software, Fellendorf and Vortisch found that the driving behavior models included in the software could reliably reproduce traffic flow under different conditions, provided that the model was first adapted to local traffic conditions [1], but without giving much insight on how to proceed.

The act of calibrating such models for a given network can be a tedious exercise that requires data often hard or expensive to gather. Because of this, few attempts have been made at calibrating the driving behavior models of traffic simulations for different locations, and the transferability of parameters calibrated for specific conditions is unknown. As a result, engineers mostly rely on the default parameters provided by the software manual or on guidelines provided by transport agencies [2-4]. This leads to uncertainty on whether or not the traffic conditions in the simulation adequately represent real conditions, or even if they can capture changes in behavior and traffic resulting from the planned changes to the network.

As automated video traffic analysis tools become more readily available [5, 6], gathering microscopic data to calibrate these microscopic models is becoming easier. This method of data collection can track a vehicle on a certain distance and provide trajectory information that is much richer than what can be gathered using traditional counting methods at only one point of the road. Among other things, this type of data lets us study lane changes more accurately by providing the means to calculate the characteristics of accepted gaps and to derive the reasons behind the lane change - that is, if the lane change is necessary for the driver to reach one’s destination or as a means to maintain one’s desired speed over a slower vehicle.

Using such trajectory data, this paper presents a general calibration procedure for micro-simulation software with the aim of providing a set of parameters usable in a wide range of traffic conditions. This is achieved by making use of a state-of-the-art derivative-free optimization algorithm that can reliably explore the search-space and converge to an optimum when comparing the results of simulations with observed data. The presented procedure is used to calibrate the driving behavior models (both car-following and lane-changing models) in VISSIM for highways in the Greater Montreal Area by using microscopic traffic data extracted automatically from video on major highways in Montreal. The proposed parameters are then validated using a separate set of data.

LITERATURE REVIEW

Micro-Simulation Calibration

Calibration is the process by which we seek to adjust a model’s parameter values so it can best reproduce specific traffic conditions [7]. The act of calibration starts with the identification of the study goals and of the model’s relevant parameters. Depending on their number, the optimization algorithm may not be able to handle them all at once, and a sensitivity analysis may become mandatory to select those parameters who have the most impact on the indicators chosen to study the model’s ability to reproduce the field conditions [8]. Assuming this step has been correctly performed, many factors are still to take into consideration: data collection, formulation of the calibration problem, automation of the procedure, measuring fitness, and the need for repeated model runs [9].

According to [7], the type of data, its temporal and spatial resolution and coverage should be evaluated based on the project scope and the model that is calibrated. Since acquiring that data is often a challenge, their availability may affect the feasible scope of the calibration process. Data such as flow, density, speed, and number of turning vehicles are commonly used in calibration studies [7, 9]. Another aspect of data collection to consider is the need to avoid overfitting, which happens when a model is over calibrated to fit specific conditions and its ability to reproduce other conditions; e.g. different traffic demands or road designs, is reduced. Consequently, data should be collected according to the possible scenarios that a calibration aims to model [9].

The formulation of the problem is its mathematical representation through which the calibration can be solved. It is normally formulated as an optimization problem and includes an objective function to minimize. Additionally, the feasible range of the parameters and the constraints in the optimization process should be formulated. In a review of 18 calibration studies made on various micro-simulation software, Hollander and Liu found that some optimization approaches are commonly used in those studies: most use genetic algorithms as their optimizer, others use the Nelder-Mead Simplex Method or Box's Complex Algorithm [9]. Other types of algorithms such as Particle Swarm optimization [10] are found in the literature, while some studies rely on statistical sampling of the search space with methods such as an Exhaustive search method [11] or a Latin Hypercube sampling technique [12], which are not optimization algorithms by themselves. These techniques all have in common that they do not use derivatives of the calibration problem, as its explicit formulation is either unknown, too complex, or not differentiable (as with minimax problems). They are also all implemented as part of an automatic process that feeds the micro-simulation software with parameters to simulate before comparing the results against the field data and making the decision on what new set of parameters to try. In that sense, [9] states that "innovative optimization methods are intensively discussed in the literature in Mathematics and in Operational Research" and that they "advise traffic analysts to constantly review recent developments in this field and seek improved solution methods for the [micro-simulation] calibration problem."

The choice of the fitness function is of a critical importance during the calibration process as it is also either the objective function itself fed to the optimization algorithm or a part of that objective function (such as when many outputs are aggregated through the use of a minimax or weighting function). In [7], it is stressed that the choice of the measure of fitness is influenced by the study objectives and the available data, while reciprocally the data that the study requires can be influenced by the choice of the measure of fitness. As such, a good definition of the problem is important.

A good review of the different available fitness functions can be found in [9]. Many choices are available both to assess the difference between distributions or between aggregated data. One of the questions to answer while choosing a fitness function is how to treat small errors around the mean value as two main groups of functions can be formed based on that criterion. It is also important to note that the fitness function used to evaluate the model calibration does not have to be linked to the function used thereafter to assess traffic impact of a projected change to the network using the calibrated model [7].

Finally, because of the stochastic nature of micro-simulation software, the question of the required number of replications of a simulation instance (for a fixed set of parameters) is also a critical question that must be answered. Since the result of a single simulation run depends on the seed number given to a pseudorandom number generator, two repetitions of the same simulation given the same parameters and different initial seed numbers will generate two different results. As a consequence, one must run several replications with different seed numbers in order to guarantee a level of statistical confidence in the results. Since most studies found in [9] are interested for their calibration in the mean value of an indicator, such as mean travel time or delay, they can safely use the iterative Student test:

$$N \geq R = \left(\frac{s \cdot t_{\alpha/2;N-1}}{\varepsilon}\right)^2$$

where N is the number of replications performed, R is the minimum number of replications needed to estimate the mean of a given traffic measure within the tolerance ε, s is the standard deviation of that measure, $t_{\alpha/2;N-1}$ is such that for a random variable t following the Student distribution with *N-1* degrees of freedom $P(|t| > t_{\alpha/2;N-1}) = \alpha$ and α is the confidence level. However, this technique cannot be used when the measure of fitness is not based on a mean value. A technique is proposed in [13] that performs several runs and empirically observes the number of replications for a given measure of fitness to converge to a stable value.

The described procedure is summarized in the following list. It is largely inspired by the summary presented in [9] and we refer to that paper for additional precisions on each step of the procedure. Steps 1 to 4 depend on each other and can be resolved in any order, the first one to be fixed influencing the decisions in the other steps.

1. Defining the scope of the problem;
2. Determining the parameters relevant to the calibration and their feasible range, either through a sensitivity analysis or a literature revue;
3. Determining and collecting the needed data;
4. Choosing the indicators of the calibration process and the fitness function, also called objective function in the optimization literature, used to compare field data and simulated data;
5. Determining the calibration process, including the choice of algorithm, the optimization constraints, and the evaluation budget;
6. Determining the required number of replications of each trial point (set of simulation parameters);
7. Validating the calibrated parameters onto an independent set of data.

The last part of the procedure is the validation of the calibrated parameters which aims to ensure that the predictive power of the calibrated model is transposable to another set of data. As proposed in [9], the indicators and measures of fitness used in the calibration can be used for the validation process and they should not be more demanding for the latter. The data used in the validation process should not have been included in the calibration process and can include data from other time sets or other locations. To verify if the calibrated model can be used on other traffic conditions than the one used in the calibration, then the data selected for the validation process should include a sample of those traffic conditions, since if the validation process is undertaken in the same conditions than the calibration, then the calibrated model can only reliably be used under those conditions [9].

The Mesh Adaptive Direct Search algorithm

The MADS algorithm [14] is a derivative-free method specialized for constrained black box optimization. Its general principle is to generate candidate solutions on a discretization of the *n*-dimensional space of the black box parameters called the mesh, and to evaluate these candidates with the black box. Each iteration is decomposed into two main steps, the "search" and the "poll".

The search is flexible and allows the generation of candidates anywhere on the mesh. Strategies such as latin-hypercube sampling or surrogate modeling can be employed. Another generic search option, used in this work, is the variable neighborhood search (VNS) described in [15], whose goal is to escape from local minima. To do so, it first applies a random perturbation to the current iterate (current best solution), and then a local search similar to the poll step. During the MADS iterations, as long as the current best iterate is not improved, the amplitude of the random perturbation is increased, and each time a new success is made, the amplitude is reset to its smallest value.

The poll is more rigidly defined but ensures the convergence of the algorithm. It consists in constructing mesh trial points using directions with the property of being dense on the unit sphere, once normalized. Once the search and poll candidates have been evaluated, the last step of the iteration checks if the iteration is a success by using the progressive barrier technique [16], which compares the candidates using the objective function and the constraint violations. In case of success, the next iterate is updated and the coarseness of the mesh (the mesh size) is enlarged. Otherwise, the current iterate stays the same and the mesh size is reduced.

As it is expected, near a local optimum, repeated failures eventually lead the mesh size to drop to zero. This is the basis of the MADS analysis which states that, under mild hypotheses, global convergence to a local minimum is ensured. The MADS algorithm is implemented in the free and open source NOMAD software package [17] at www.gerad.ca/nomad.

METHODOLOGY

Problem formulation

The procedure presented in this paper aims at calibrating both the lane-changing model and the car-following model together. Since the two car-following models available in VISSIM, Wiedemann 74 and Wiedemann 99, are mutually exclusive, the calibration consists of two separate subproblems. The objective of this calibration is twofold: firstly, the goal is to develop an easily reusable methodology that can be applied to any traffic simulation software and calibration problem and secondly, to provide a set of parameters that can be used in a large number of traffic conditions in the Greater Montreal area. In this paper, only the calibration of the subproblem including the Wiedemann 99 model is presented.

Black box

To perform the calibration, VISSIM was treated as a black box to be optimized by NOMAD. A wrapper was written in the Python language to allow communication between NOMAD and VISSIM, to adjust the parameters chosen by NOMAD at every trial point, and to calculate the objective function's result and constraint values out of VISSIM's outputs.

The VNS option of NOMAD was activated during the calibration process. Testing showed that it allowed the algorithm to find better points, although at the cost of more iterations.

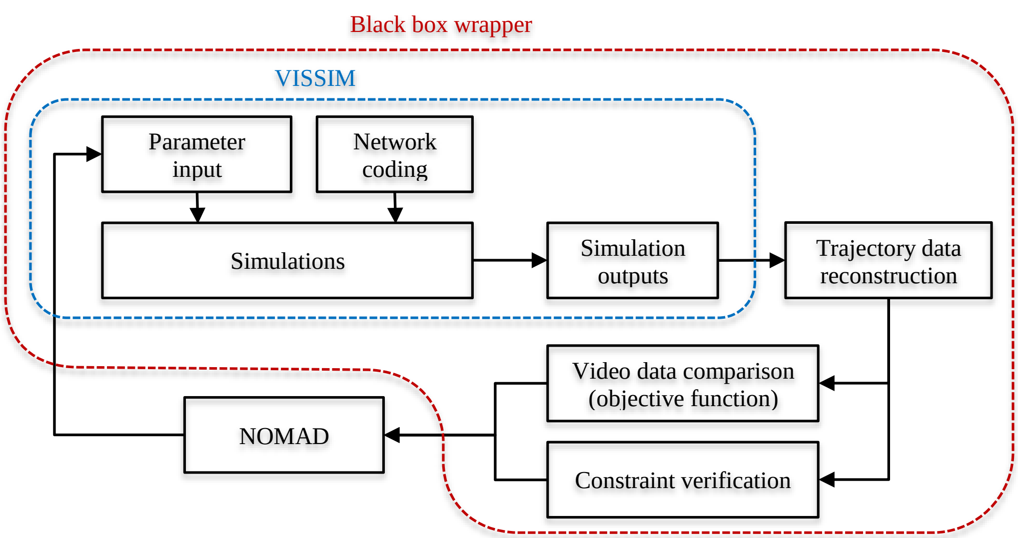

FIGURE 1 : Overview of the method. The loop shown is called once for each trial point selected by NOMAD.

Overview

FIGURE 1 shows the implemented method: it represents a single evaluation from the MADS algorithm asking for a trial point to be tested. VISSIM (in blue) simulates the traffic given the set of parameters (trial point). The outputs are then extracted and transformed by the wrapper (in red) into the objective and constraint values that are returned to NOMAD. The cycle is repeated until convergence or until exhaustion of the calibration budget.

To ensure that a trial point only tests the parameters included in the calibration and that no changes were accidently made to the VISSIM network file in between trials, every parameter that are related to the driving behavior models are overridden by the wrapper at every trial points. The value entered is either the default value or the value selected by NOMAD in the case of parameters included in the calibration.

Site selection and video collection

Data was collected on Highway 13 in Montreal using a GoPro Hero 2 set atop a 20ft telescopic mast attached to fixed structures such as lampposts using the equipment and technique described in [6]. The straight highway section was chosen for its geometric characteristic and for ease of access to fixed structures for the mast. Data was collected in one traffic direction for 12 hours on a week day with clear weather from 10am to 7pm to capture the peak hour in the selected direction.

The videos were then processed to extract trajectory information using the video tracker available in the open source Traffic Intelligence project [6]. These trajectories were assigned to traffic lanes in order to calculate lane changes. A manual inspection of the processed data was performed to subtract camera detection errors such as vehicles being detected twice or trucks being followed on the upper end of their trailer, leading to erroneous projection onto the traffic lanes because of a greater parallax effect than for other vehicles. Finally, a manual count of the videos was performed and compared to the automatic count performed by Traffic intelligence and only videos with a difference of less than 10% between automated and manual counts were kept. Traffic intelligence is freely available at bitbucket.org/Nicolas/trafficintelligence.

CALIBRATION

Parameters selection and bounds

Since the problem is subdivided in relatively small subproblems (20 or fewer parameters), all related parameters of the models were included. The chosen parameters are shown in TABLE 3 with the names used in the COM section of the online help documentation and their descriptions can be found in the VISSIM manual. The VISSIM manual provided a first approach to parameter bounds by providing those that cannot be relaxed without having the software crash. A review of calibration handbooks and of other studies on calibration was done to determine the parameter lower and upper bounds typically used that were not provided in the manual. A process of trial and errors was used to determine the final bounds for the parameters: relaxing a bound that seemed to block the progression of the MADS algorithm, which resulted in many points tried at that value, or tightening bounds when only infeasible points, i.e. points that do not respect the calibration constraints, were found in that range. The final bounds are provided in TABLE 2 in the result section.

Objective function

The optimization process is performed by minimizing the difference between the time headway distributions observed in the video data and those derived from the trajectories simulated by VISSIM and saved in the ".fzp" file. For the car-following models, the headway is calculated on every lane. The point selected is situated at mid-link in VISSIM, but at 20% of the field of view in the videos. This closer point minimizes parallax error and gives better results for the observed data. For the lane-change models, the headway considered is the headway available in the destination lane just prior to the lane change. The comparison of distributions is done using the Kolmogorov-Smirnov d-statistic, which is the largest distance between the two cumulative distribution functions (CDF). To avoid overfitting the observations in a single video, the calibration is performed simultaneously on many videos. To ensure that the new points found during the calibration process are a better result on every data set, the objective function returned to NOMAD is the maximum of all the d-statistic calculated between the simulated data and the videos used. This objective function can be formulated as:

$$f = max_i(d_i)$$

with: $$d_i = \sup_x \left[F_{i,1}(x) - F_{i,2}(x)\right]$$

where i is the video number, $F_{i,1}(x)$ is the CDF of the video data, and $F_{i,2}(x)$ is the CDF of its corresponding simulated data.

Constraints

Calibration constraints were used to discard problematic simulations produced by VISSIM with certain parameters such as vehicles passing through each other or large traffic jams resulting from two vehicles that try to change lane at the same time and block each other and every following vehicle. Two strategies were used to detect these errors: analyzing the error file produced by VISSIM and detecting “collisions” in the vehicle trajectories.

In the error file, three different errors prove to be somewhat related to the problem. In order to build a constraint, the number $count_{ij}$ of each type of error *i* was counted for each replication *j* of the trial point. A trial and error process was used to determine the threshold for each type of error that is associated with a problematic simulation. The errors and their thresholds are summarized in TABLE 1. To “punish” any point where at least one seed number results in a bad simulation, the maximum number of errors for the n replications is returned as the constraint value. For each type of error i,

$$C_i = \max_{j=1\ldots n}\left(count_{ij}\right) - threshold_i \qquad i = 1 \ldots 3$$

These three constraints can then be handled with any type of constraint handling strategy, since the violation can be accurately quantified. In this paper, the constraint strategy used for these was the *Extreme Barrier* approach, which consists of rejecting any point that violates the constraint.

TABLE 1 : Constraints relying on VISSIM errors and their associated threshold

Error	*Threshold*
“vehicle input could not be finished”	35
“deceleration is positive”	0
“acceleration is zero”	0

The other type of errors generated by VISSIM consists of situations where a vehicle is allowed by VISSIM to pass through another vehicle in the same lane. Those situations are impossible in the real world and have to be removed if our simulations are to be representative of real world conditions. To detect these situations, we count each occurrence of a vehicle suddenly becoming ahead of the vehicle it was previously following without any lane change. This operation being a lot more computationally expensive, the current implementation cannot provide the exact number of such occurrences. It is therefore considered as non-quantifiable and the only possible constraint handling strategy is the Extreme Barrier.

Number of replications

Because of our choice of objective function, which does not rely on the mean value but rather on the whole distribution of a traffic variable, the traditional method based on the Student test cannot be used to calculate the number of replications needed. Consequently, we rely on the strategy proposed in [13] to find the point where the objective function stabilizes even though more replications are carried out. Given that the traffic data used for calibration was collected in two categories of traffic conditions as can be seen in TABLE 2, the test is performed for traffic data extracted from one video in each category. FIGURE 2 shows that after 15 to 20 replications, the d-statistic of the Kolmogorov-Smirnov test between observed and simulated data based on the default VISSIM values stabilizes for both traffic conditions. We chose 20 replications as multiples of 10 works well on our computer architecture while using multiprocessing to process VISSIM outputs.

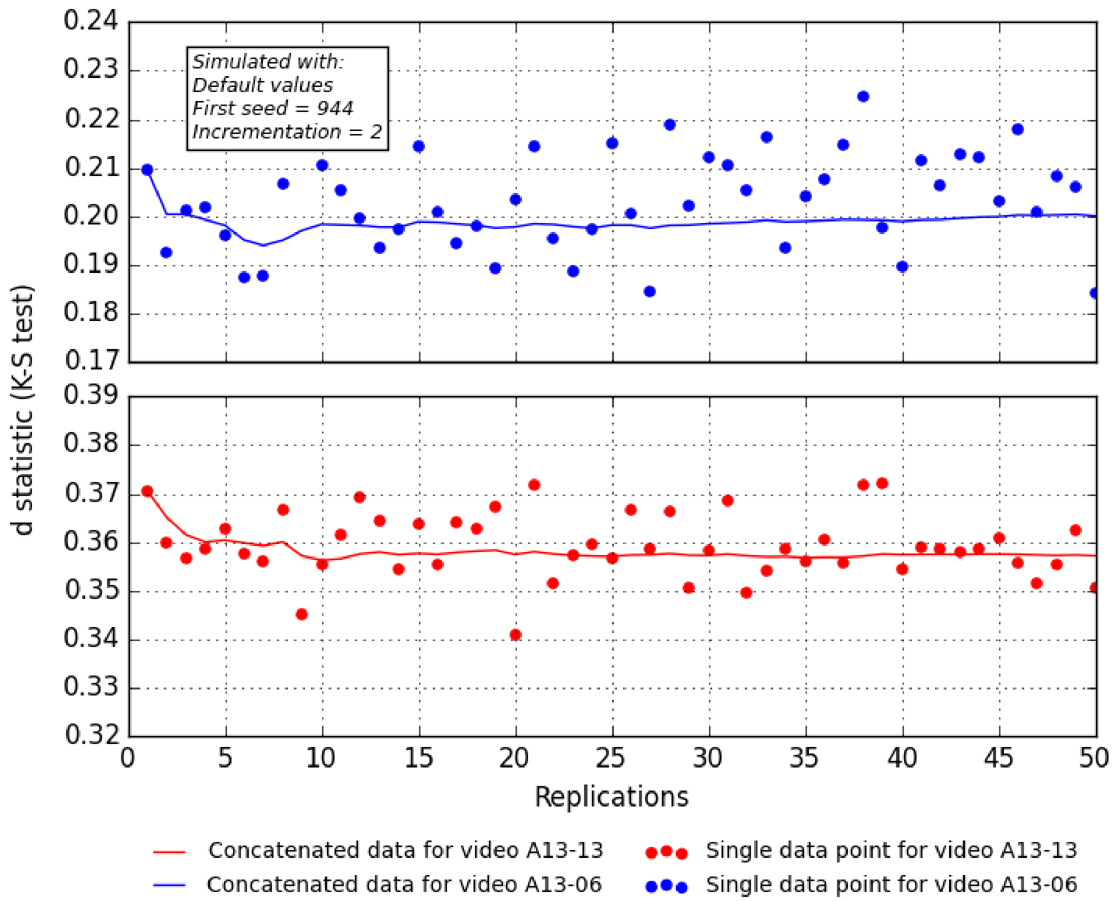

FIGURE 2: Effect of many replications on the Kolmogorov-Smirnov test's d statistic. The seed number for replication i is given by first seed value $+ i \times$ incrementation. The scatter points show the value of the d statistic for the simulated headways for replication i and the video data. The solid line represents the value of the d statistic calculated on the concatenated headway distributions of every replications up to and including i. Flow for the A13-13 video is about twice as high as the flow for the A13-06 video.

Starting point

The default values listed in the VISSIM documentation provide a convenient starting point since it is the most commonly used set of parameters by practitioners. It is therefore meaningful to compare the end results with that point.

Evaluation budget

With 1.5 to 5 minutes per tested point, depending on the number of VISSIM networks to run and whether or not lane change headways are calculated, it rapidly becomes rather costly to perform a complete calibration and finding the right number of points to test becomes important. Several trials have been made to determine the best evaluation budget to allocate, with 1000 points turning out to be the most efficient since most best points are found around the 900 mark.

VALIDATION

The validation process is nearly identical to the method presented in FIGURE 1 except that NOMAD is not a part of the process. As a result, a single iteration of the method is run. Moreover, the validation is performed on videos not part of the calibration process. If an infeasible point is detected during that phase, the whole calibration process must be restarted, and adjustments have to be made to take into account the network producing the invalid simulations.

RESULTS

The described methodology is implemented for the Wiedemann 99 model along with the lane-changing model comparing the car-following headways observed in the simulation and those extracted from the videos taken on a straight 3 lanes segment of Highway 13 (A-13) on the island of Montreal. Speed distributions and total vehicle count (shown as flow per lane in TABLE 2) were extracted from each video sequence and manually verified for the first 5 minutes. All of these data points were in free-flow conditions as the tracker from Traffic Intelligence was not well enough calibrated to provide good tracking results in congested conditions. The simulation time used after 1 minute warm up is also 5 minutes on a stretch of about a kilometer of highway. The included parameters, their lower and upper bounds and the starting and calibrated points are presented in TABLE 3.

TABLE 2 : Values of the objective function for the starting point and the best point found for the different videos used for the calibration and the validation processes

*Video**	*Flow* (veh/h/ln)	*d statistic from the Kolmogorov-Smirnov test* Default values	Calibrated values
Calibration			
A13-06	1252	0.2025	0.0933
A13-11	1352	0.3069	0.1447
A13-12	1556	0.2961	0.1311
A13-13	2140	0.3576	0.0879
Objective function	--	*0.3069*	*0.1447*
Validation			
A13-10	1352	0.185	0.1181
A13-25	1296	0.224	0.0972

**A13-XX stands for the video sequence number*

FIGURE 4 shows the cumulated probability functions of the headways for both the calibrated parameters and the starting parameters. The two CDF are close to each other, but the four graphs show that the beginnings of the calibrated CDF are closer to the field data while the end of the distribution was not substantially changed. Trying to calibrate over four set of field data with different CDF makes it hard to approach any closely, though the figure shows that the range of simulated headways is always smaller than what is observed. There also probably remain some tracking errors for small headways that cannot be simulated. On the other hand, field data show platoons that seem further apart and that VISSIM's vehicle generation model does not seem to be able to replicate.

The validation also shows good results, with improvements on both tested videos. More importantly, it shows that the simulation using the calibrated parameters can replicate the traffic conditions in these videos as well as for the videos used for the calibration. Since no traffic condition out of the original range used for the calibration could be tested, due to lack of good data outside of these conditions, it is only possible to state that the parameters give good results for flows ranging from 1250 veh/h/ln to 2150 veh/h/ln in non-congested conditions.

Some of the parameter values found by the calibration process differ substantially from their default value provided in VISSIM, while most stayed close to this initial value. The space explored between the given bounds can be visualized in FIGURE 3, which shows that around 80% of the parameter combinations actually result in the violation of at least one of the calibration constraints and are therefore discarded.

TABLE 3 : Starting point, lower and upper bounds and best point found for parameters used in the calibration

	Parameter	*Default value*	*Lower bound*	*Upper bound*	*Calibrated Value*
Car-following model	w99cc0	1.5	0.0*	10.0	0.894
	w99cc1	0.9	0.0*	20.0	0.006
	w99cc2	4.0	0.0*	10.0	0.671
	w99cc3	-8.0	-30.0	30.0	-6.658
	w99cc4	-0.35	-30.0	30.0	0.992
	w99cc5	0.35	-30.0	30.0	0.350
	w99cc6	11.44	-30.0	30.0	17.925
	w99cc7	0.25	-30.0	30.0	-5.117
	w99cc8	3.5	0.1	120.0	5.462
	w99cc9	1.5	0.1	120.0	120.000
	ObsrvdVeh	2	0*	10	8
Lane-changing model	MaxDecelOwn	-4.0	-10.0*	-0.02*	-3.337
	DecelRedDistOwn	100.0	0.0*	200.0	104.472
	AccDecelOwn	-1.0	-10.0*	-1.0*	-1.000
	MaxDecelTrail	-3.0	-10.0*	-0.02*	-3.446
	DecelRedDistTrail	100.0	0.0*	200.0	91.056
	AccDecelTrail	-1.0	-10.0*	-1.0*	-9.397
	DiffusTm	60.0	0.0	200.0	158.387
	MinHdwy	0.5	0.0*	10	8.550
	SafDistFactLnChg	0.6	0.0*	1.0*	0.578

**Bound value fixed by the software. Violation causes VISSIM to crash.*

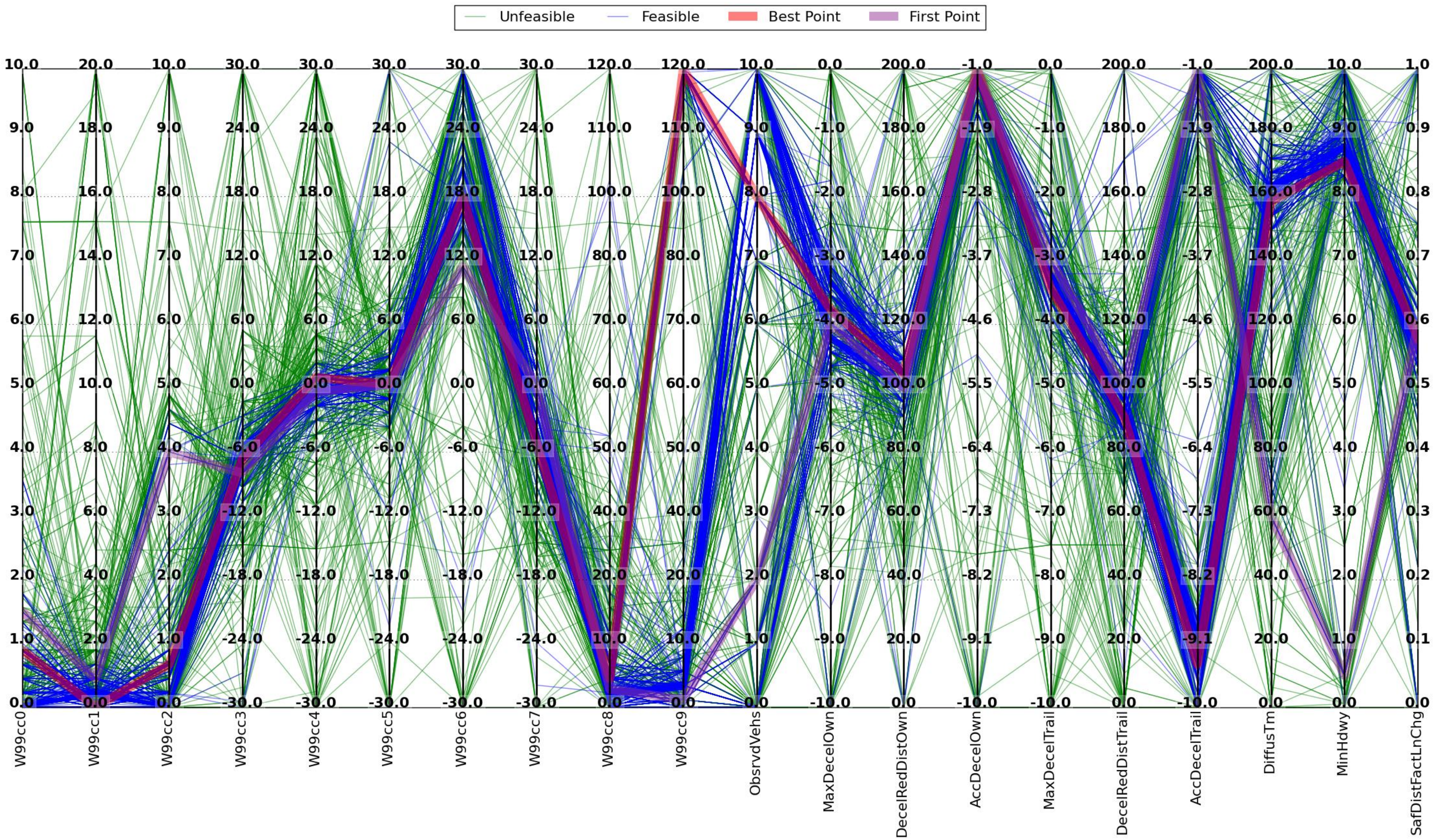

FIGURE 3 : Visualization of the explored area of the parameter search space given to NOMAD. Points that violate at least one constraint are presented in green while feasible points are shown in blue. The starting point and the calibrated points are shown in different colors than the other feasible points.

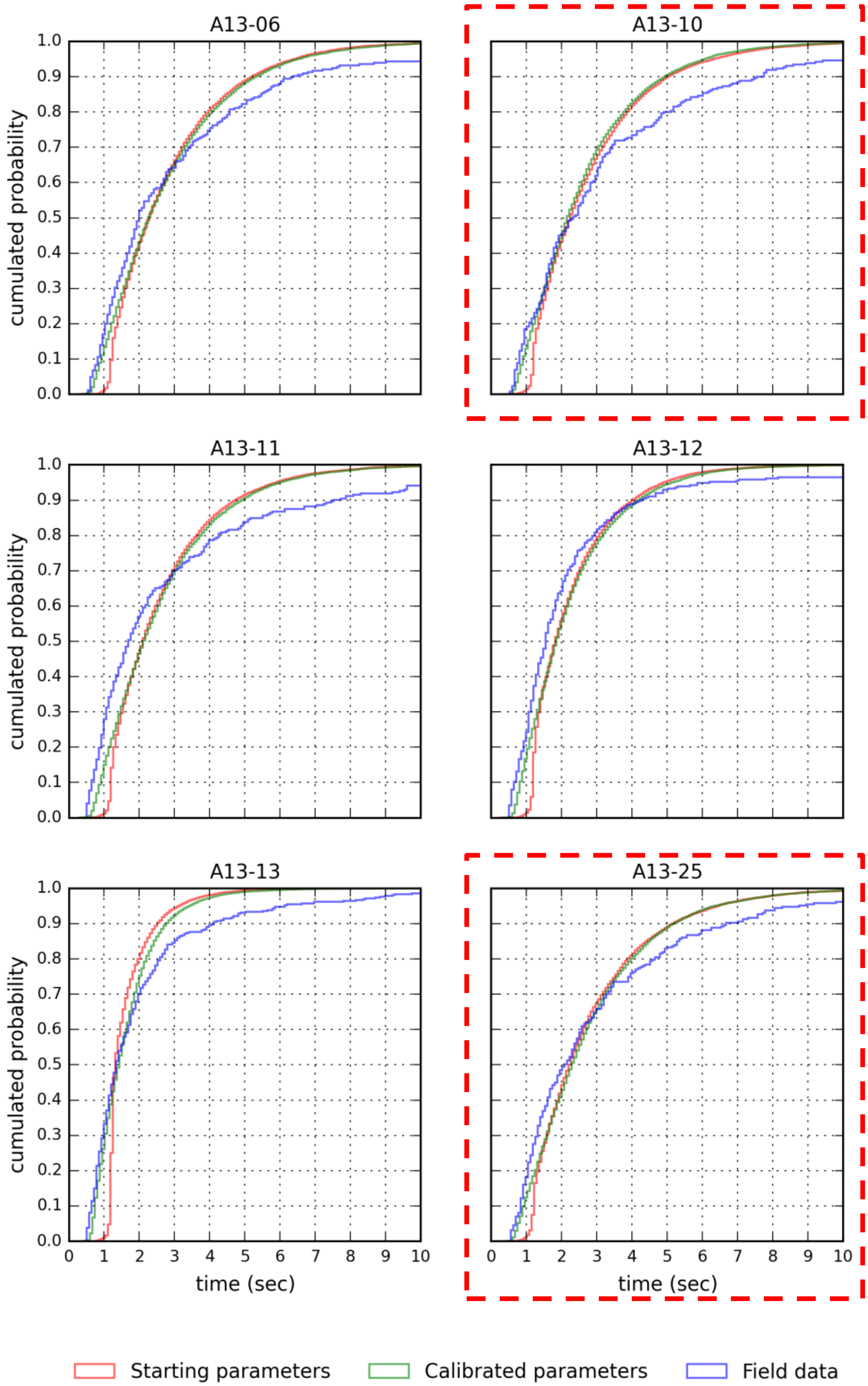

FIGURE 4 : CDFs of the headways associated with the starting and calibrated points, as well as the field video data. A13-06, A13-11, A13-12, and A13-13 are the data set used in the calibration process while A13-10 and A13-25 correspond to the data set used in the validation process (in dashed red boxes).

DISCUSSION

It was shown that the calibration methodology could substantially improve the fit of the simulated headway distribution with that of the observed field headway distributions. Using video collection and automated trajectory extraction permitted an easier access to the microscopic data needed to calibrate these models. Due to a lack of data in congested traffic conditions and on other straight segments, both on other 3 lanes segments of other highways and on straight segments with different numbers of lanes, how the calibrated parameters perform generally on the highway network of Montreal could not be completely evaluated. Also because of a lack of data, it was not possible to verify the validity of the parameters through the day, although [18] showed that driver's aggressiveness varies between the morning peak hour and the evening peak hour which may affect the calibration results or may require different parameters for different times of the day. However, the success of the method on the different available traffic conditions demonstrates that it is indeed possible to reach generalisation with minimal additional efforts.

The methodology presented is general and will be used again to calibrate other models and software. One of the first extensions of the proposed method will be to consider the lane changing headways and the lane change counts into the objective function to attest that the new parameters do not deteriorate these aspects of the simulation while calibrating for the car-following headways. Then, the calibration tool will be ready to be applied to other types of roads such as arterials and smaller residential streets as well as downtown area segments, provided that the microscopic trajectory data can be obtained in heavier traffic conditions. The technique could also be used in other cities to find the values associated with driver behaviors on their respective networks with relatively small efforts: video data collection and analysis is a fast and easy process once automated.

The presented methodology could also be used for other types of road users available in VISSIM such as the pedestrian and cyclist models to make urban traffic analyses that depend on these parameters more faithful to field conditions. Ultimately, the method will be ported to other micro-simulation software and released under an open source license for independent replication and contribution, as well as wider adoption by researchers and practitioners.

ACKNOWLEDGMENTS

This work has been done as part of a project funded by the City of Montreal. The authors also wish to acknowledge the work of Karla Gamboa who helped with data collection and video analysis.

REFERENCES

[1] Fellendorf, M. and P. Vortisch, *Validation of the Microscopic Traffic Flow model VISSIM in different Real-World Situations*, in *Transport Research Board 89th Annual Meeting*. 2001: Washington, DC.

[2] ODOT, *Protocol for VISSIM Simulation*. 2011, Oregon Department of Transportation: Salem, USA.

[3] WSDOT, *Protocol for VISSIM simulation*. 2014, Washington State Department of Transportation: Seatle, USA.

[4] Transport for London, *Traffic Modelling Guidelines, TfL Traffic Manager and Network Performance Best Practices. V3.0*. 2010, Transport for London: London, UK.

[5] Jodoin, J. P., G. A. Bilodeau, and N. Saunier. *Urban Tracker: Multiple object tracking in urban mixed traffic*. in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*. 2014.

[6] Jackson, S., et al., *A Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioural Analysis.* Transportation Research Record: Journal of the Transportation Research Board, 2012. Vol. 2365: p. 90-98.

[7] Hellinga, B. R., *Requirements for the calibration of traffic simulation models.* Proceedings of the Canadian Society for Civil Engineering, 1998. Vol. 4: p. 211-222.

[8] Ciuffo;, B., 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 Intelligent Transport Systems, 2014. 8 (5): p. 479 – 489.

[9] Hollander, Y. and R. Liu, *The principles of calibrating traffic microsimulation models.* Transportation, 2008. Vol. 35 (3): p. 347-362.

[10] Aghabayk Eagely, S., et al. *A novel methodology for evolutionary calibration of vissim by multi-threading*. in *36th Australasian Transport Research Forum*. 2013. Bristbane, Australia.

[11] Chatterjee, I., et al., *Replication of work zone capacity values in a simulation model.* Transportation Research Record: Journal of the Transportation Research Board, 2009. No. 2130: p. 138-148.

[12] Park, B. B. and J. D. Schneeberger, *Microscopic Simulation Model Calibration and Validation - Case study of VISSIM Simulation Model for a Coordinated Actuated Signal System.* Transportation Research Record: Journal of the Transportation Research Board, 2003. Vol. 1856: p. 185-192.

[13] Hollander, Y. and R. Liu, *Estimation of the distribution of travel times by repeated simulation.* Transportation Research Part C: Emerging Technologies, 2008. Vol. 16 (2): p. 212-231.

[14] Audet, C. and Dennis, Jr., J. E., *Mesh Adaptive Direct Search Algorithms for Constrained Optimization.* SIAM Journal on Optimization, 2006. Vol. 17 (1): p. 188-217.

[15] Audet, C., V. Béchard, and S. Le Digabel, *Nonsmooth Optimization through Mesh Adaptive Direct Search and Variable Neighborhood Search.* Journal of Global Optimization, 2008. Vol. 41 (2): p. 299-318.

[16] Audet, C. and J. Dennis, J E, *A progressive barrier for derivative-free nonlinear programming.* SIAM Journal on Optimization, 2009. Vol. 20 (1): p. 445-472.

[17] Le Digabel, S., *Algorithm 909: NOMAD: Nonlinear Optimization with the MADS Algorithm.* ACM Transactions on Mathematical Software, 2011. Vol. 37 (4).

[18] Dixit, V., V. Gayah, and E. Radwan, *Comparison of Driver Behavior by Time of Day and Wet Pavement Conditions.* Journal of Transportation Engineering, 2012. 138 (8): p. 1023-1029.