Calibrating Stochastic Traffic Simulation Models for Safety and Operational Measures Based on Vehicle Conflict Distributions Obtained from Drone and Traffic Camera Videos

Di Sha[1], Jingqin Gao, Di Yang, Fan Zuo, Kaan Ozbay

Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY, USA, ds5317@nyu.edu
Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY, USA, jingqin.gao@nyu.edu
Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY, USA, dy855@nyu.edu
Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY, USA, fz380@nyu.edu
Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY, USA, kaan.ozbay@nyu.edu

1. Introduction

One key objective in the performance evaluation of the CV applications is to evaluate their safety performance using observational data collected from both CV probes and microscopic traffic simulation models. The advantage of using microscopic simulation models is that they allow for confounding factors to be controlled in the simulation environment. In addition, the deployed CV applications can be implemented in the simulation environment and tested under different scenarios. Simulation-based surrogate safety measure (SSM) has been proved to be an effective tool for conducting safety assessment of traffic systems (*1*) and microscopic simulation models have been widely used in the literature for Connected and Automated Vehicles (CAVs) applications evaluation. To effectively evaluate the safety benefits of CV applications, the simulation models should not only be calibrated to match the real-world operational conditions, but also the safety conditions (*2*). This paper proposes a novel calibration framework which combines traffic conflict techniques and multi-objective stochastic optimization into the model calibration process. The proposed calibration framework is used to calibrate a stochastic microscopic simulation model of an urban roadway network in Brooklyn, NY. High-resolution drone-based and traffic camera videos are utilized for vehicle trajectories and traffic conflicts extraction during the calibration process.

2. Methodology

The proposed model calibration framework (Figure 1) (*3*) is implemented using SUMO, an open-source microscopic road traffic simulation package short for "Simulation of Urban MObility" (*4*). Various data sources are leveraged to develop the base model network, including traffic volumes, travel times, turning movement counts, and traffic conflicts quantified from SSM based on video records from closed circuit television (CCTV) cameras and drones. Then model calibration is conducted using SPSA in an iterative manner with the goal of minimizing the error between simulation measures and the observed traffic measures. Multiple key parameters, such as acceleration and minimum gap, are considered random variables and are calibrated as probability distributions based on the "trajectory data" extracted from the drone videos to capture the real-world conditions. The simulation model is based on a 1.6-mile road segment on Flatbush Avenue between Tillary Street and Grand Army Plaza in Brooklyn, NY. The studied time period is the morning peak period (between 6:00 AM and 10:00 AM). Roadway geometry, lane usage, link capacity, speed limit, lane and turn connectivity, and other parameters are thoroughly calibrated in the base model. Besides network topology, other basic road network information (such as signal timings and bus stops) is obtained and integrated into the development of the simulation network. The simulation network built using SUMO is shown in Figure 2. As suggested by Federal Highway Administration (FHWA) (*5*), representative days were identified using cluster analysis before calibrating the simulation model. The representative days are used to typify system performance dynamics associated with the collection of days with similar travel condition. 96 weekdays for the study area are identified in this collection. These days can be seen as samples of the general traffic condition of the simulation network based on the available data sources.

[1] * Corresponding author. Tel.: +001-917-476-2380;
E-mail address: ds5317@nyu.edu.

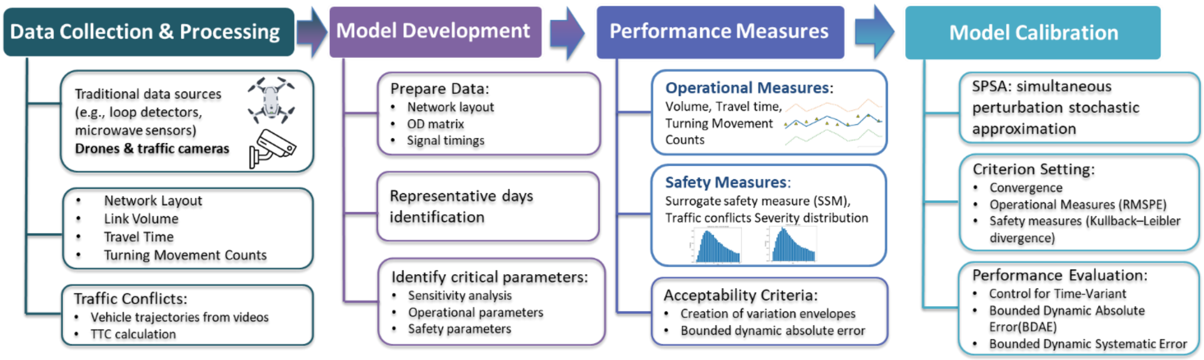

Figure 1: Proposed model calibration framework.

Figure 2: Flatbush Avenue simulation network developed and calibrated in SUMO.

High-resolution drone videos have the advantages of outstanding flexibility, maneuverability, low cost, and the capability of recording from a bird-eye view compared with traditional traffic surveillance video data, thus are utilized for data extraction of vehicle trajectories and safety performance measures in this study. The extracted vehicle trajectories include the longitude and the latitude of the center of each vehicle as well as speed and acceleration. Each vehicle is given a unique but anonymous identification number and is classified into one of the following four vehicle types based on its length: passenger car, medium vehicle, heavy vehicle, and bus. An example of the collected drone videos and extracted vehicle trajectories is shown in Figure 3.

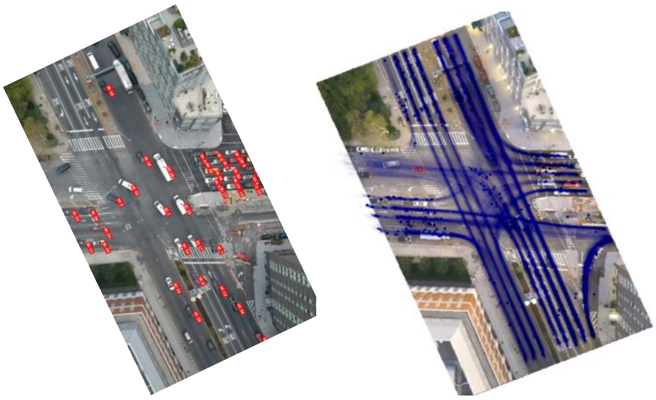

Figure 3: An example of drone videos (left) and extracted vehicle trajectories (right).

2.1. Simulation performance measures

In the proposed model calibration framework, the operational and safety measures may have different indicators, but they need to be combined in the calibration process. Link volumes and travel times are adopted as the operational performance measures in this study. The root mean square percentage error (RMSPE) (*6*) is used as the goodness-of-fit measure for link volume at multiple locations along Flatbush Ave. RMSPE is calculated as:

$$RMSPE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{Y_n^{obs}-Y_n^{sim}}{Y_n^{obs}}\right)^2} \quad (1)$$

where Y_n^{obs} and Y_n^{sim} are the observed and simulated performance measures respectively, and N is the number of links for volume or number of time intervals for travel time collection. There are also four acceptability criteria suggested by FHWA (*5*) being adopted for travel time calibration using the variation envelope approach. Please see (*5*) for more details.

For safety evaluation, traffic conflicts have been frequently used as a performance measure. Instead of simply comparing the aggregated number of traffic conflicts for an intersection or an approach, the conflict distribution in terms of different levels of severity is used as the comparison target in this study since it provides more details of the collected conflicts. The different levels of severity of traffic conflicts are categorized using time to collision (TTC) (*7*). TTC is defined as the time required for two vehicles to collide if they continue along the same path at their present speeds. To measure the goodness-of-fit of the simulated conflict distribution compared to the ground truth one, the Kullback–Leibler divergence (*8*) is used which is a metric that quantifies the "distance" between two distributions. The Kullback–Leibler divergence from Q to P is defined as:

$$D_{\mathrm{KL}}(P\,||\,Q) = \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right) \quad (2)$$

where P and Q are discrete, simulated, and observed conflict severity distributions respectively defined on the same probability space, $\mathcal{X}$.

As the goal is to minimize the errors between simulation results and observed measures, the weighted sum of the RMSPE of operational measures and Kullback–Leibler divergence (D_{KL}) for safety measures is defined as the calibration objective function, which can be represented as:

$$\min \quad L(\theta, I) = w_1 \cdot RMSPE_V + w_2 \cdot RMSPE_T + w_3 \cdot D_{KL} \quad (3)$$

where $L(\theta, I)$ is the total simulation error, $RMSPE_V$ is the simulation error of link volumes, $RMSPE_T$ is the simulation error of travel time, D_{KL} is the simulation error of traffic conflict distributions, and w_1, w_2, w_3 are weights of the error terms. In this paper, the three error terms are scaled to the same magnitude, then assigned equal weights for calibration.

2.2. Model calibration algorithm

The microscopic model calibration and validation process is an iterative process which involves searching for an optimal set of input parameters such that the error between simulation measures and the observed traffic measures is minimized (*9*). When incorporating the safety-related measures into the calibration process, both stochasticity and dimensionality of the calibration problem increase. Therefore, the simultaneous perturbation stochastic approximation (SPSA) (*10*) method is used as the calibration algorithm in this study due to its superior efficiency (*11-13*) when calibrating high-dimensional stochastic simulation models. The SPSA algorithm is an extension of the generic stochastic approximation (SA) algorithm. Compared with the traditional finite-difference (FD) method (*14*), SPSA only needs two function evaluations for gradient approximation regardless of the parameter dimension. The proposed calibration process is shown in Figure 4. By iterating the parameter updating process, the differences between simulated and observed measures are minimized until a pre-defined stopping criterion is satisfied. Moreover, multiple sets of initial parameter values and multiple runs of each of these sets are tested to account for the randomness of the initialized parameter values, the stochastic nature of microscopic simulation and the variance of the input data.

3. Analysis and Results

The model calibration process starts with six sets of different initial values of model parameters. These six sets of initial values are generated from the initial pool defined by the boundary conditions of parameters using the Latin hypercube sampling (LHS) method (*15*). The calibrated critical parameters include vehicle's maximum speed (*maxSpeed*), acceleration, deceleration, and driver's reaction time (*tau*) for different vehicle types. For each set of initial values, the simulation model is calibrated for multiple replications with different random seeds until a pre-set level of stooping criterion based on a variance reduction method is satisfied. The convergence results reveal that the calibrated parameters achieve similar levels of simulation accuracy for different replications although the starting simulation errors may vary due to the difference between initial parameter values and random seeds. The converged simulation errors lie between 10% and 15%, which is satisfactory considering the scale of the simulation network and the dimension of the calibration problem. This result also confirms the robustness of the proposed calibration process since the converged condition is stable even if the starting conditions are varying.

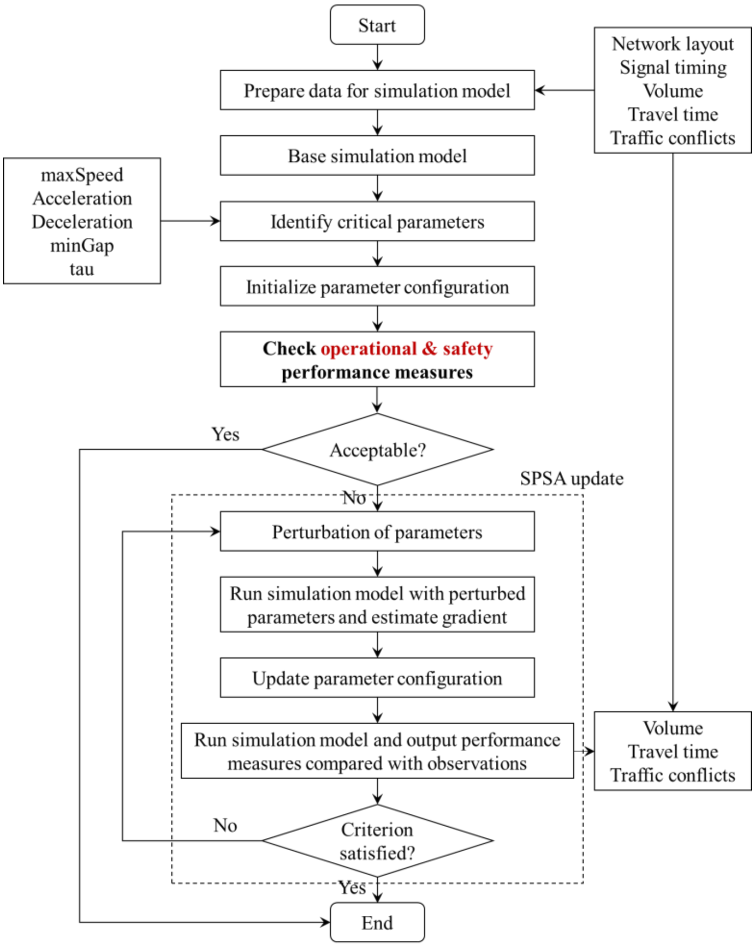

Figure 4: Flowchart of the proposed calibration process with operational and safety considerations.

Based on the calibration results, the output travel times of the calibrated simulation model on both northbound (NB) and southbound (SB) Flatbush Avenue satisfy all of the four aforementioned acceptability criteria suggested by FHWA (*5*). For demonstration purpose, the northbound calibration results are shown in Figure 5. For the validation of link volumes, the above four acceptability criteria are not applicable due to the lack of multiple days' traffic counts data. Therefore, the observed and simulated link volumes are compared for 26 NB/SB segments on Flatbush Ave. The RMSPEs range from 0.23% to 16.61%, with an average value of 11.01%. These results confirm that the calibrated simulation model generates traffic volumes that are in good agreement with the observed counts.

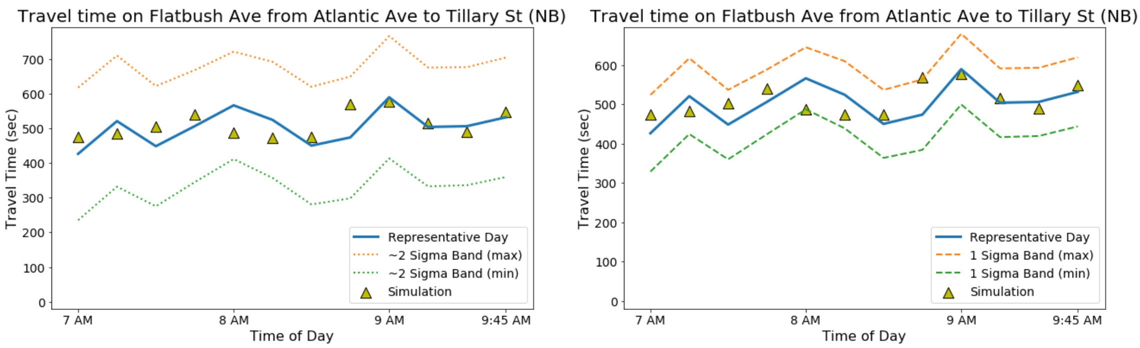

Figure 5: NB travel time (Atlantic Ave – Tillary St) validation with respect to the variation envelope.

For the validation of safety measures, the Kullback–Leibler divergence between observed and simulated conflict distributions is computed. Based on the availability of field-collected traffic conflict data, 7 approaches from multiple locations on Flatbush Avenue are investigated. These traffic conflict data are extracted using vehicle trajectories from a total of 14 hours' CCTV cameras and drone videos. Figure 6 shows an example of observed and simulated conflict distributions comparison at Flatbush Ave & Fulton St Northbound approach. The significant similarity between the traffic conflict distribution from the simulation model and the observed one can be confirmed by the relatively low Kullback–Leibler divergence values for all the 7 approaches that are investigated, with an average value of 0. 0223. These results indicate that the calibrated simulation model can reproduce the real-world traffic conditions in both operational and safety aspects.

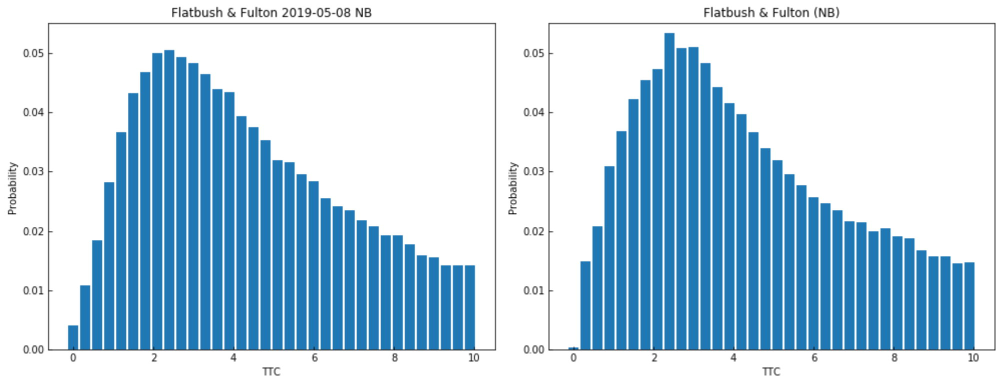

Figure 6: Observed (left) and simulated (right) conflict distributions after calibration of NB approach at Flatbush Ave & Fulton St intersection (D_{KL}=0.0047).

4. Conclusions

In this study, a novel calibration framework is proposed for the purpose of evaluating CV applications using a microscopic simulation model. The main contribution of the proposed calibration framework is to calibrate the simulation model for both operational and safety measures simultaneously using traffic conflict techniques and multi-objective stochastic optimization. The conflict distribution of different severity levels categorized by TTC is applied as the safety performance measure along with traditional operational measures including traffic volume and segment travel times. The simultaneous perturbation stochastic approximation algorithm, SPSA, is used to calibrate 17 key parameters, such as acceleration and minimum gap, and address the increasing stochasticity and dimensionality of the calibration problem when incorporating safety measures. The proposed calibration methodology is empirically tested in a SUMO-based simulation network of the Flatbush Avenue corridor in Brooklyn, NY with new stream of data from high-resolution drone and traffic camera videos. The case study demonstrates that the calibrated parameters can significantly improve the performance of the simulation model to represent real-world traffic conflicts as well as operational conditions. It also demonstrates the usefulness of drone data and the applicability of the proposed novel model calibration framework for calibrating both operational and safety measurements simultaneously by setting up the calibration problem as a multi-objective stochastic optimization problem.

Acknowledgment

This study was partially funded by C2SMART, a Tier 1 University Transportation Center at Tandon School of Engineering, New York University. The work was developed as part of the New York City Connected Vehicle Pilot Deployment (NYC CVPD) project which is funded by the U.S. Department of Transportation (USDOT). The contents of this paper only reflect the view of the authors who are responsible for the facts and do not represent any official views of any sponsoring organization or agencies.

References

1. Ozbay, K., H. Yang, B. Bartin, et al. Derivation and validation of new simulation-based surrogate safety measure. *Transportation Research Record,* Vol. 2083, No. 1, 2008, pp. 105-113.
2. Xie, K., D. Yang, K. Ozbay, et al. Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. *Accident Analysis & Prevention*, Vol. 125, 2019, pp. 311-319.
3. Gaigano, S., M. Talas, K. Opie, et al. Connected vehicle pilot deployment program Phase 2: Performance measurement and evaluation support plan - New York City, U.S. Department of Transportation, 2021.

4. Krajzewicz, D., J. Erdmann, M. Behrisch, et al. Recent development and applications of SUMO - Simulation of Urban MObility. *International Journal on Advances in Systems and Measurements*, Vol. 5, No. 3 & 4, 2012, pp. 128-138.
5. Wunderlich, K., M. Vasudevan, and P. Wang. *Traffic analysis toolbox volume III: Guidelines for applying traffic microsimulation modeling software 2019 update to the 2004 version*, U. S. D. o. T. Federal Highway Administration. Federal Highway Administration, U.S. Department of Transportation, 2019.
6. Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts, fourth ed*. Irwin McGraw-Hill, Boston, MA, 1997.
7. Hayward, J. C. Near miss determination through use of a scale of danger. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 384, 1972, pp. 24-34.
8. Kullback, S., and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, Vol. 22, No. 1, 1951, pp. 79-86.
9. Yu, M., and W. D. Fan. Calibration of microscopic traffic simulation models using metaheuristic algorithms. *International Journal of Transportation Science and Technology*, Vol. 6, 2017, pp. 63-77.
10. Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, Vol. 37, No. 3, 1992, pp. 332-341.
11. Lee, J.-B., and K. Ozbay. New calibration methodology for microscopic traffic simulation using enhanced simultaneous perturbation stochastic approximation approach. *Transportation Research Record: Journal of Transportation Research Board*, No. 2124, 2009, pp. 233-240.
12. Mudigonda, S., and K. Ozbay. Robust calibration of macroscopic traffic simulation models using stochastic collocation. *Transportation Research Part C: Emerging Technologies*, Vol. 59, 2015, pp. 358-374.
13. Sha, D., K. Ozbay, and Y. Ding. Applying bayesian optimization for calibration of transportation simulation models. *Transportation Research Record: Journal of Transportation Research Board*, Vol. 2674, No. 10, 2020, pp. 215-228.
14. Dennis Jr., J. E., and R. B. Schnabel. Chapter I A view of unconstrained optimization. *Handbooks in Operations Research and Management Science*, Vol. 1, 1989, pp. 1-72.
15. Loh, W.-L. On Latin hypercube sampling. *The Analysis of Statistics*, Vol. 24, No. 5, 1996, pp. 2058-2080.