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

Di Sha
Graduate Student

Di Yang, Ph.D.
Post-Doctoral Associate

Kaan Ozbay, Ph.D.
Professor

Jingqin Gao, Ph.D.
Research Associate

Fan Zuo
Graduate Student

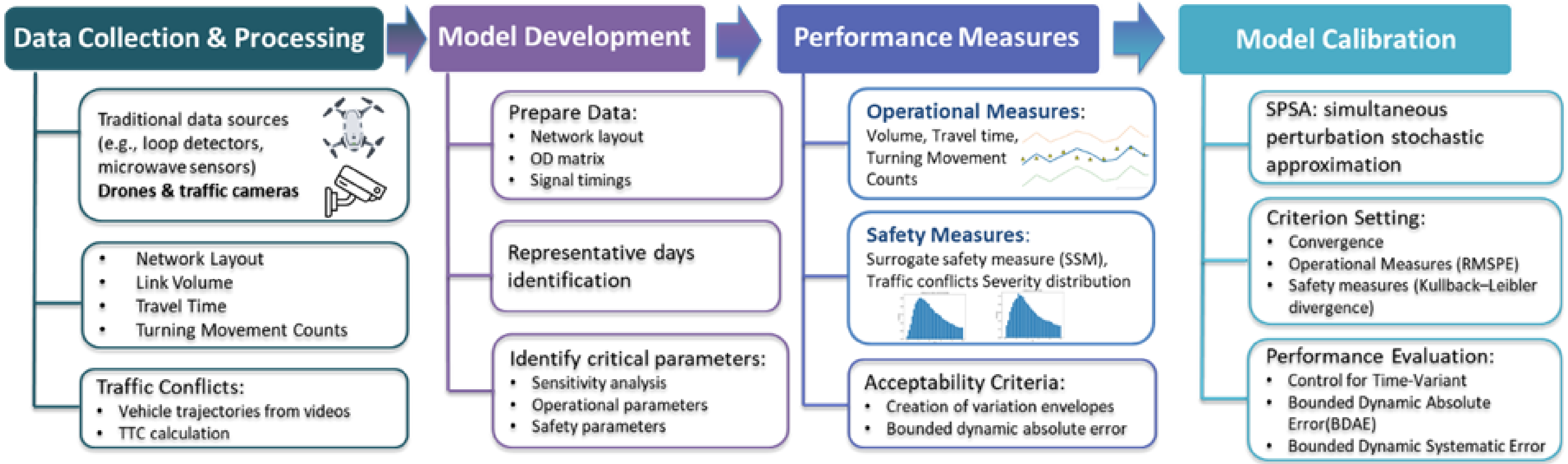
New York University
Jun 10th, 2022

Introduction

- Microscopic simulations have been widely used to assist in the performance evaluation for connected vehicle (CV) technologies.
- Safety evaluation of CV applications demands the simulation models to be calibrated to match real-world safety conditions.
- This study proposes a novel calibration framework which combines traffic conflict techniques and multi-objective stochastic optimization so that the operational and safety measures can be calibrated simultaneously.
- The conflict distribution of different severity levels categorized by time-to-collision (TTC) is applied as the safety performance measure. Real-world traffic conflicts are extracted using vehicle trajectories from both drone and traffic camera videos.
- The proposed calibrated framework is tested on an urban network built in SUMO. 17 parameters are calibrated using the SPSA algorithm.
- The results show that the calibrated parameters can significantly improve the performance of the simulation model to represent real-world traffic conflicts as well as operational conditions.

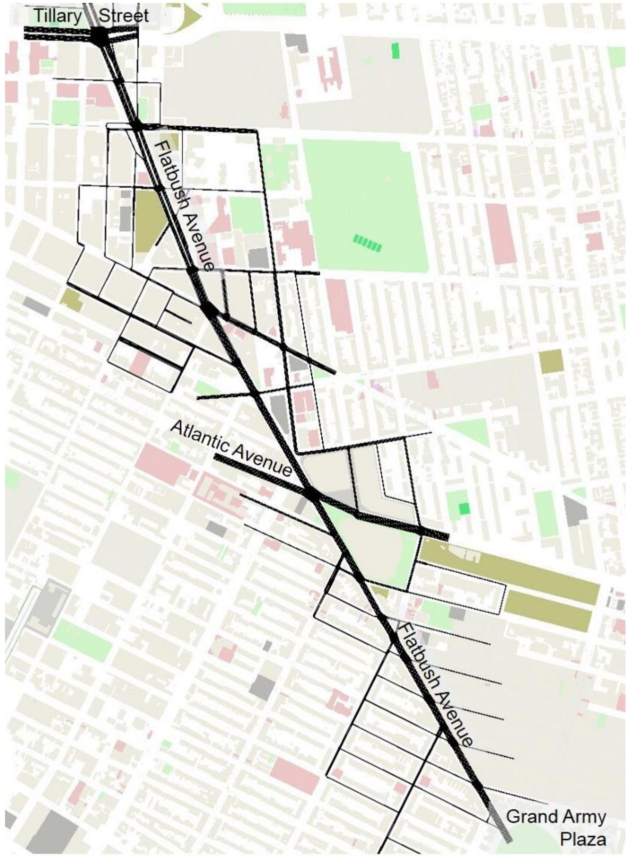
Methodology

- The proposed model calibration framework consists four steps:
  - ➢ Data collection and processing
  - ➢ Base model development; Representative days; Critical parameters identification
  - ➢ Performance measures and acceptability criteria determination
  - ➢ Model calibration using SPSA to minimize the total error between simulation and observation



Methodology - Base Model Development

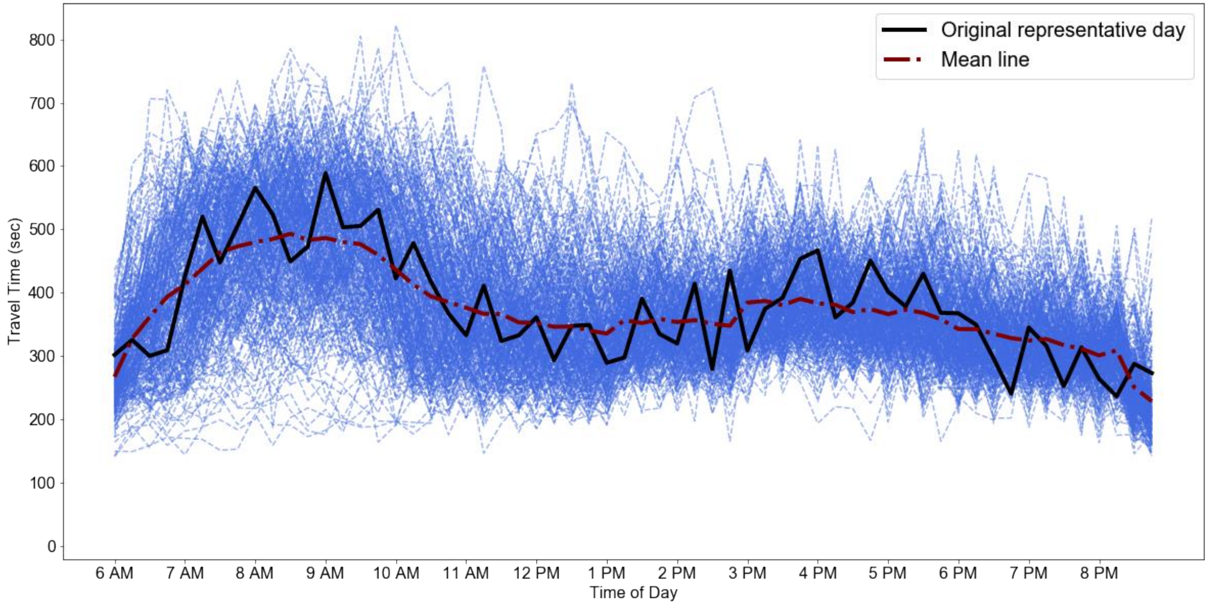
- The proposed model calibration framework is implemented using SUMO:
  - Open-source microsimulation software
  - Traffic Control Interface (TraCI) allows users to control and interact with a running simulation model online
  - Can be run in a parallel mode to reduce simulation running time
- Simulation network: Flatbush Avenue in Brooklyn, NY
  - 1.6-mile urban road segment
  - Flatbush Avenue: bi-directional, North-South urban corridor with eight lanes (four in each direction)
  - West-East minor streets are constructed for the area of two or three blocks from Flatbush Avenue
  - Studied time period: morning peak period (6:00 - 10:00 AM)
  - Traffic signal and bus information is obtained and integrated



Flatbush Avenue simulation network developed in SUMO

Methodology – Representative Days

- FHWA suggests identifying representative days before calibrating the simulation model
    - ➢ The identification results enable simulation validation using the variation envelopes.
    - ➢ For practical purpose, traffic counts data of different intersections/links collected on different days can be supplemental to each other if these days are identified belonging to the same traffic condition.
- A whole year's travel time data are used to identify representative days
    - ➢ Remove weekends and holidays.
    - ➢ Select the original representative day as the closest one to the mean of samples.
    - ➢ Compare the rest of the days with the original representative day and calculate the similarity based on K-S test statistics.
    - ➢ Determine representative days based on similarities. 96 workdays are identified for the simulated traffic condition.



Original representative day identification based on travel time data

Methodology – Performance Measures

- Operational and safety performance should be measured using different indicators.
- Different measures should be combined in the objective function and calibrated together.
- Operational measures
  - Traffic counts and travel times collected at multiple links
  - Goodness-of-fit measure: root mean square percentage error (RMSPE)

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

  - Besides, the four acceptability criteria suggested by FHWA are adopted to travel time calibration.
    - Control for time-variant outliers
    - Control for time-variant inliers
    - Bounded dynamic absolute error (BDAE)
    - Bounded dynamic systematic error

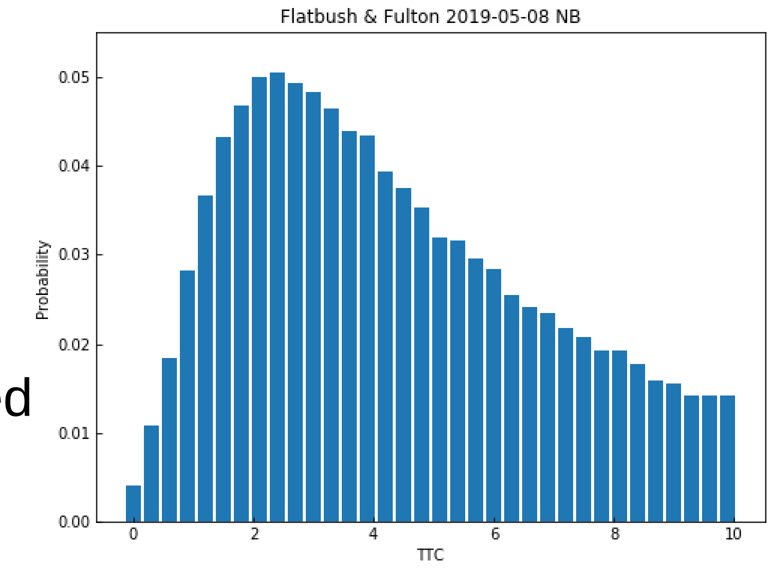
Methodology – Performance Measures

- Operational and safety performance should be measured using different indicators.
- Different measures should be combined in the objective function and calibrated together.
- Safety measures
  - Traffic conflict distribution in terms of different level of severity.
  - The conflict distribution provides more details of the collected conflicts than simply comparing the aggregated number of traffic conflicts.
  - The different levels of severity of traffic conflicts are categorized using time to collision (TTC).
  - Goodness-of-fit measure: Kullback–Leibler divergence

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

- The calibration objective function is defined as the weighted sum of the RMSPE of operational measures and K-L divergence (D_{KL}) for safety measures:

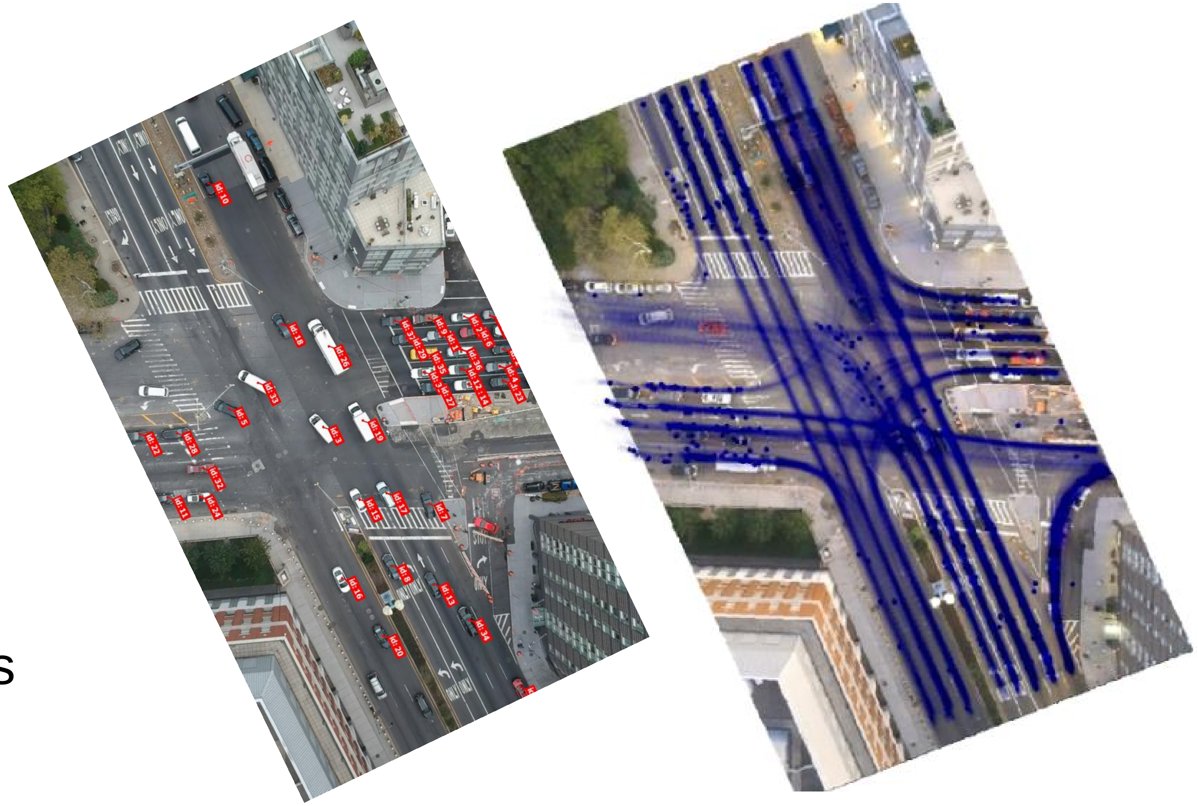
$$\min \quad L(\theta, I) = w_1 \cdot RMSPE_V + w_2 \cdot RMSPE_T + w_3 \cdot D_{KL}$$



An example of the collected conflict distribution

Methodology – Traffic Conflicts from Videos

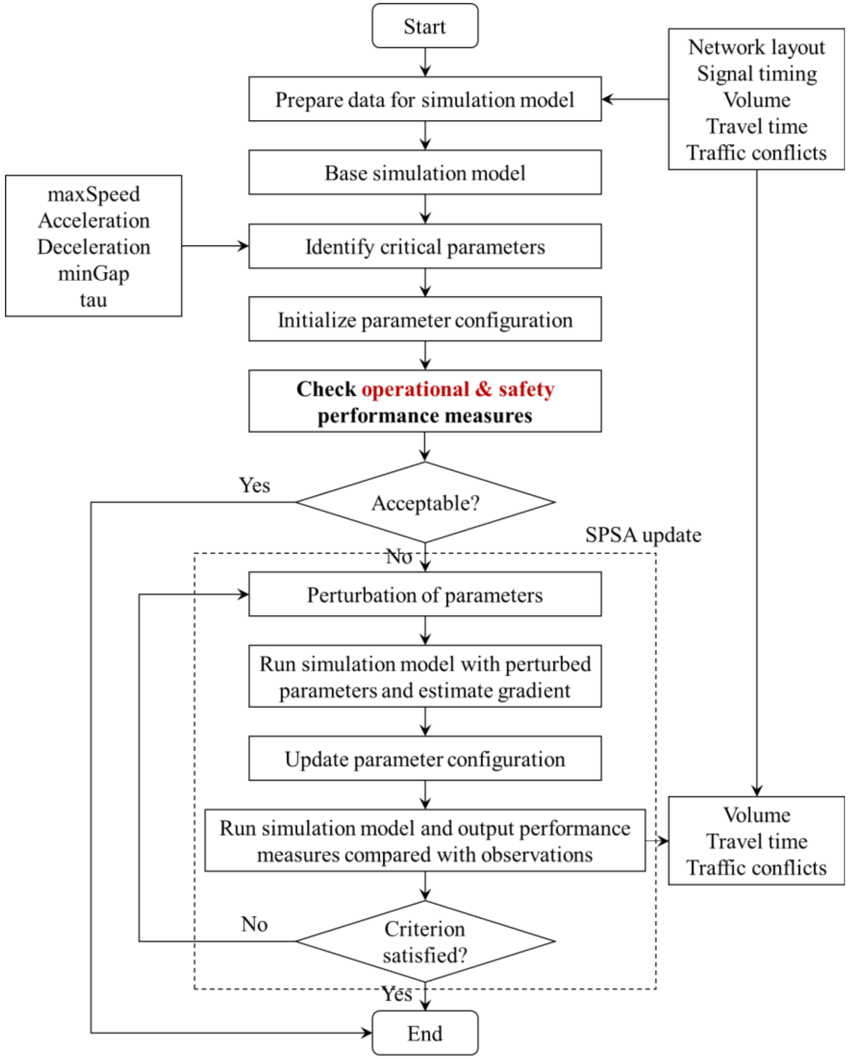
- Traffic surveillance videos and drone videos are both used to extract vehicle trajectories.
- Limitations of the traffic surveillance (street-level) videos:
  - ➢ Analysis can only be conducted at locations where the cameras are installed;
  - ➢ Recorded area is usually confined due to the low height of the camera;
  - ➢ The tilted recording angle can cause errors to vehicle trajectory extraction and the subsequent SSM quantification
- Therefore, drone videos are good alternatives for traffic conflict analysis due to their outstanding flexibility, maneuverability, low cost, and the capability of recording from a bird-eye view.



An example of vehicle trajectory extraction from drone videos

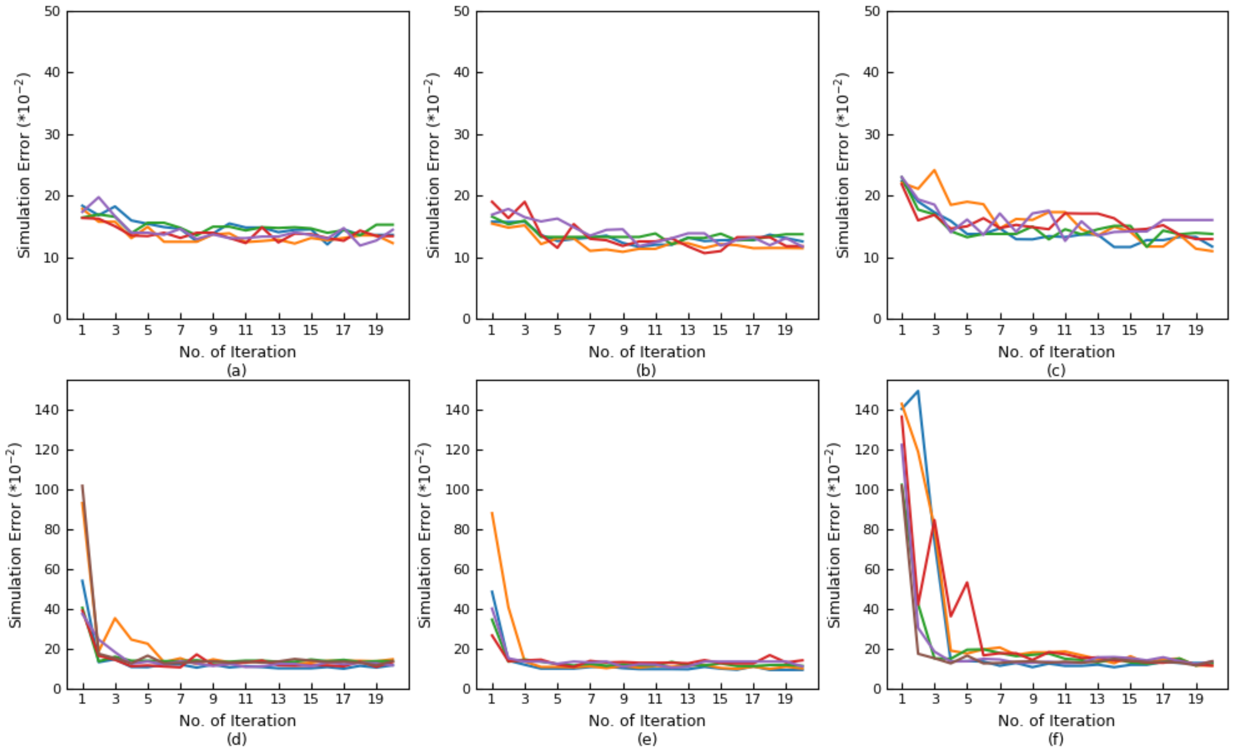
Methodology – Calibration Algorithm

- The calibration process is an iterative process to minimize the error between simulation measures and the observed traffic measures.
- The operational and safety measures are combined and calibrated simultaneously.
- The SPSA algorithm is an efficient calibration algorithm for high-dimensional stochastic simulation models, which needs only two function evaluations for gradient estimation regardless of the parameters size.
- Calibration process:
  - ➢ Assemble data
  - ➢ Develop base simulation model
  - ➢ Initialize parameter configuration
  - ➢ Update parameter values using SPSA
  - ➢ Validate calibration results



Results Analysis – Calibration Convergence

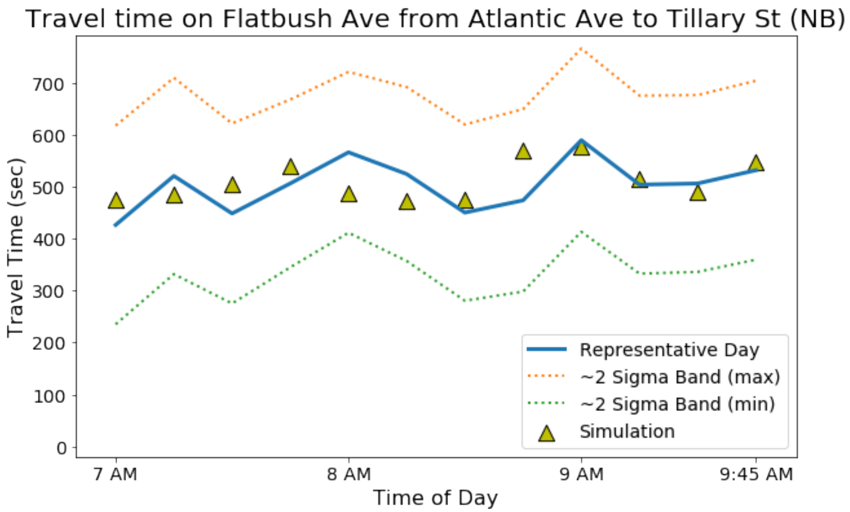
- Two aspects of stochasticity exist in the calibration process
  - ➢ Randomness of the initialized parameter values
  - ➢ Stochastic nature of microsimulation and the variance of input data
- To account for the stochasticity:
  - ➢ Multiple sets of initial parameter values are used as different calibration scenarios.
  - ➢ Multiple runs with different random seeds are performed to ensure the statistical significance of the simulation results.
  - ➢ The numbers of initial parameter sets and random seeds are determined by a sequential statistical approach.

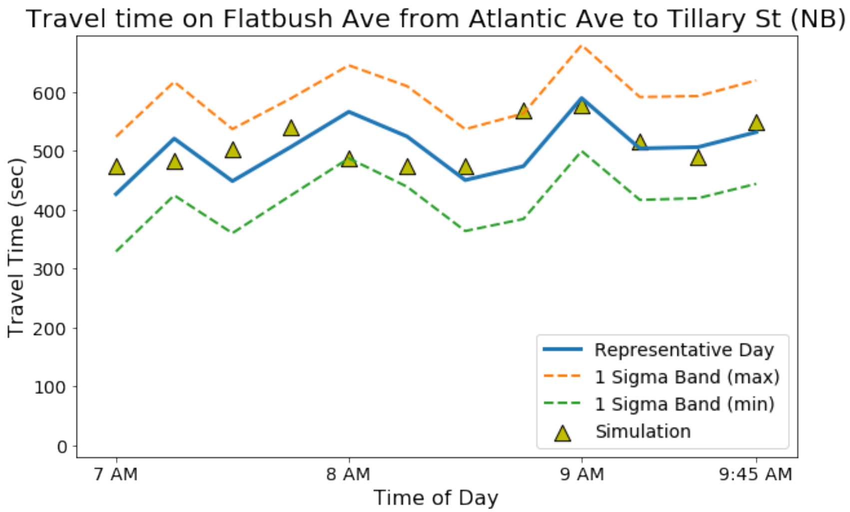


Calibration convergence diagrams for different sets of initial parameter values

Results Analysis – Calibration Results

- Operational measures: link volume & travel time.
  - RMSPEs of volumes for 26 NB/SB links: 0.23% ~ 16.61% with an average value of 11.01%
  - Check the four acceptability criteria suggested by FHWA for travel time results.
    - Control for time-variant outliers
    - Control for time-variant inliers





NB travel time results validation with respect to the first two acceptability criteria suggested by FHWA

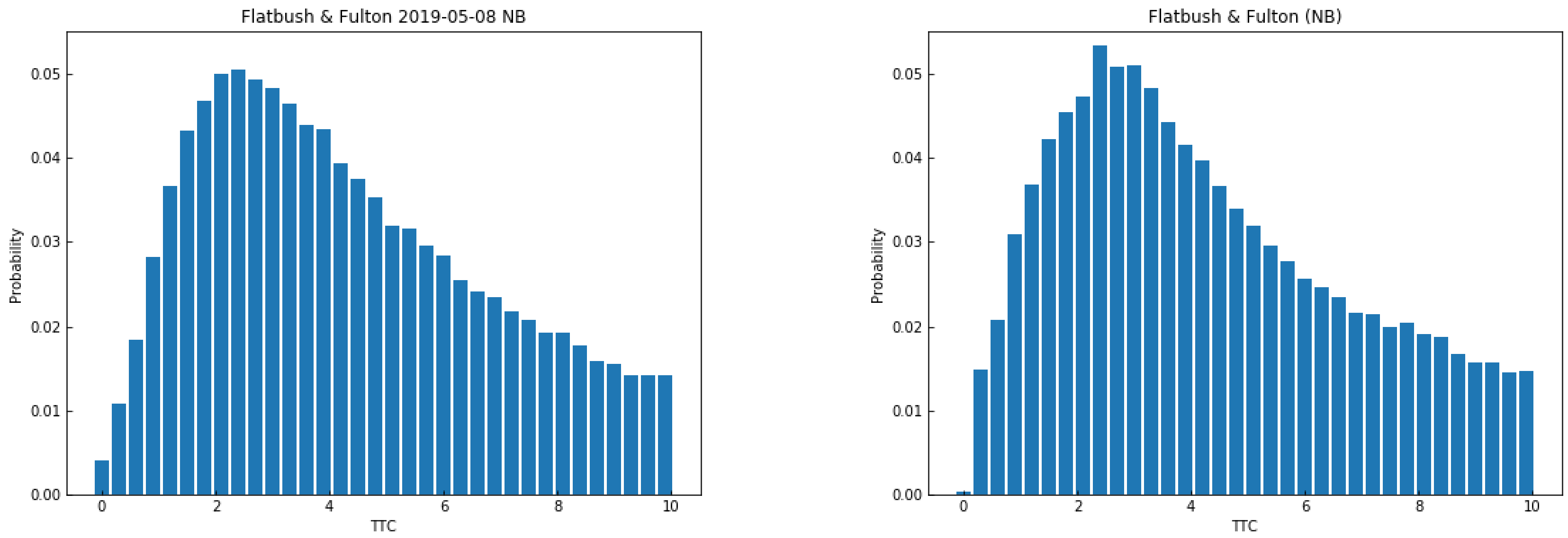
Results Analysis – Calibration Results

- Operational measures: link volume & travel time.
  - RMSPEs of volumes for 26 NB/SB links: 0.23% ~ 16.61% with an average value of 11.01%
  - Check the four acceptability criteria suggested by FHWA for travel time results.
    - Bounded dynamic absolute error (BDAE)
    - Bounded dynamic systematic error

Time	Simulation (s)	Observation (s)	Absolute difference (s)	Difference (s)
7:00 AM	474.51	426.12	48.39	48.39
7:15 AM	483.76	520.39	36.63	-36.63
7:30 AM	503.48	448.18	55.30	55.30
7:45 AM	540.26	505.97	34.29	34.29
8:00 AM	486.13	565.84	79.71	-79.71
8:15 AM	473.18	524.00	50.82	-50.82
8:30 AM	474.33	449.86	24.47	24.47
8:45 AM	569.13	473.43	95.70	95.70
9:00 AM	577.44	589.14	11.70	-11.70
9:15 AM	514.55	503.61	10.94	10.94
9:30 AM	489.53	505.86	16.33	-16.33
9:45 AM	547.68	531.36	16.32	16.32
		Average	40.05	7.52
BDAE Threshold: 82.15, $\frac{\sum_t \lvert c_r(t)-\tilde{c}_r(t)\rvert}{N_T}$ = 40.05, $\left\lvert\frac{\sum_t c_r(t)-\tilde{c}_r(t)}{N_T}\right\rvert$ = 7.52				
CRITERION III ($\frac{\sum_t \lvert c_r(t)-\tilde{c}_r(t)\rvert}{N_T} \leq BDAE\ Threshold$) is met.		CRITERION IV ($\left\lvert\frac{\sum_t c_r(t)-\tilde{c}_r(t)}{N_T}\right\rvert \leq \frac{1}{3} \times BDAE\ Threshold$) is met.		

Results Analysis – Calibration Results

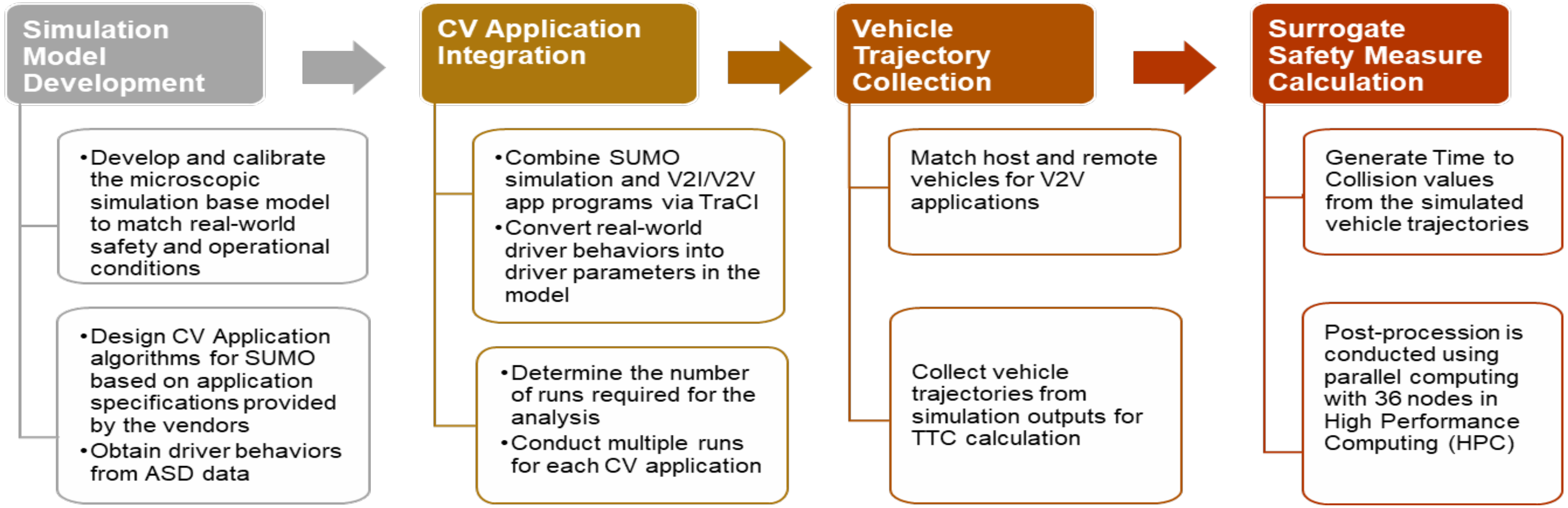
- Safety measures: traffic conflict distribution.
  - The observed and simulated conflict distributions are compared for seven approaches from multiple intersections on Flatbush Avenue.
  - The Kullback–Leibler divergence is calculated to quantify the discrepancy of distributions for each of the seven approaches.



An example of observed (left) and simulated (right) conflict distributions comparison (D_{KL}=0.0047)

Discussion

- The ultimate goal of the development and calibration of the Flatbush Ave model is to be used as a test platform for multiple CV applications.
- When incorporating the CV applications into the simulation model, drivers’ behavior under the CV environment needs to be further calibrated before the safety impact of CV technologies can be sufficiently evaluated.
- Future works include the simulation based SSM analysis for CV applications when the entire before and after data from the NYC CV pilot deployment becomes available.



Conclusions

- This study proposes a novel calibration framework is proposed for the purpose of evaluating CV applications using microscopic simulation models.
- The proposed methodology combines operational and safety measures to be calibrated simultaneously as a multi-objective stochastic optimization problem using the SPSA algorithm.
- The traffic conflict distribution of different severity levels categorized by TTC is applied as the safety performance measure.
- Traffic conflicts are recognized using vehicle trajectories extracted from high-resolution drone and traffic camera videos. Drone videos have the advantages of flexibility, maneuverability, low cost, and the capability of recording from a bird-eye view.
- The methodology is empirically tested in an urban network built in SUMO. The calibrated model performs well to represent real-world traffic conflicts and operational conditions.
- The well calibrated base scenario model will benefit future safety impact evaluation of CV applications by controlling the changes in drivers' behavior arising from the CV technology and eliminating the impacts of confounding factors.

Thank you!

https://c2smart.engineering.nyu.edu/

UrbanMITS Laboratory
C2SMART Connected Cities with Smart Transportation
Department of Civil & Urban Engineering
Center for Urban Science + Progress (CUSP)
@ New York University (NYU)