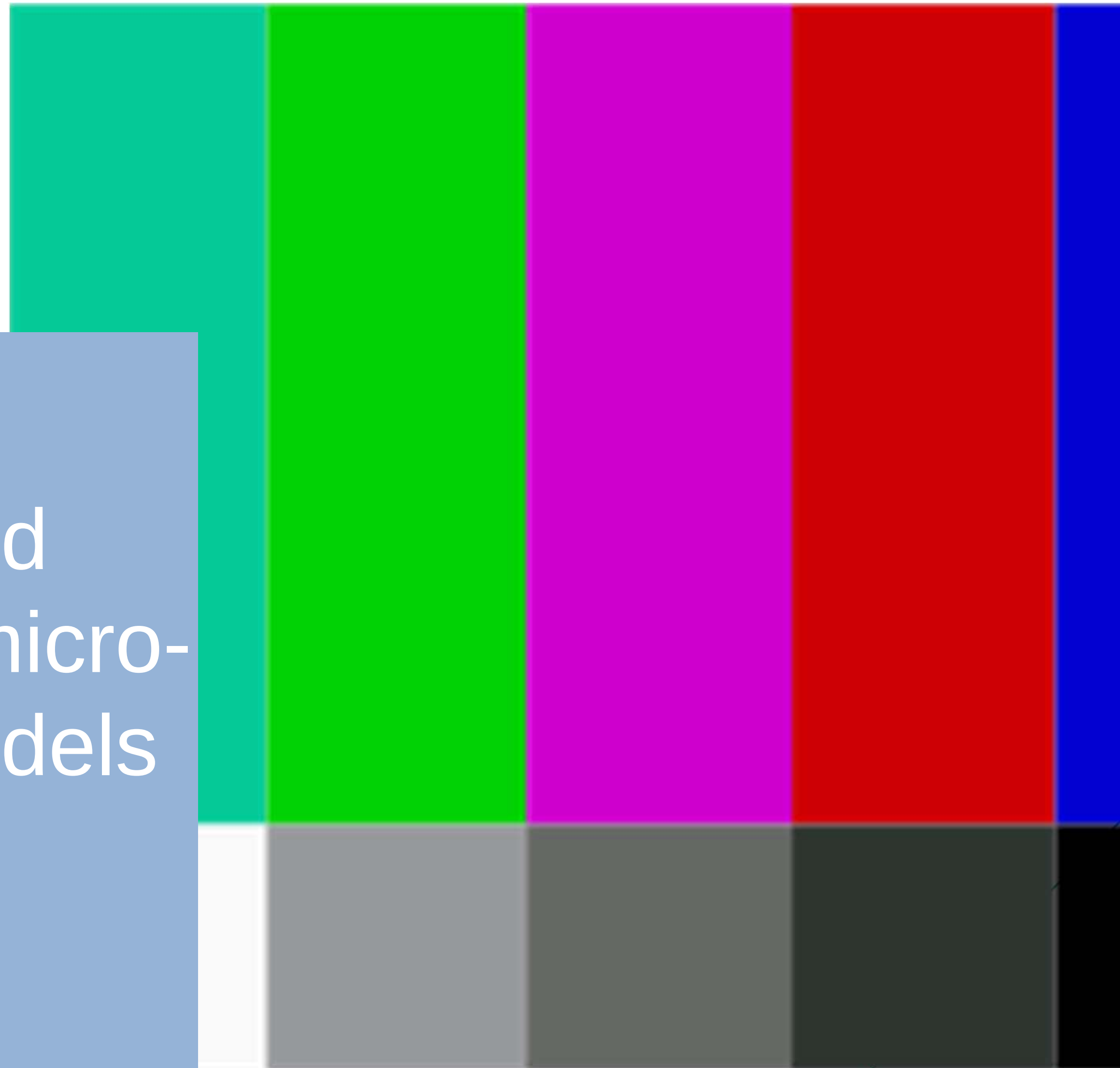
Calibration and validation of micro-simulation models

François Bélisle, WSP
Laurent Gauthier, Polytechnique Montréal
Nicolas Saunier, Polytechnique Montréal



Presentation outline

→ **Introduction**

→ **Definitions**

→ **State of the practice**

→ **Issues with calibration and validation**

→ **Unmet needs**

→ **Case study**

→ **Conclusion**

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

WSP

Introduction

1 - Introduction

- → Traffic simulation tools have been used for several decades to evaluate and support transportation projects
- → They are **complex and** include **a large number of parameters**
- → There is limited or **contradictory guidance** to choose their values
- → A ‘state of the practice’ **survey** done in 2011 (Antoniou et al. 2014) to examine how practitioners were using models uncovered that
  - → 19% of practitioners polled conducted **no calibration** of their models,
  - → of those that did, 45% based their decisions on **personal experience**

1 - Introduction

→ Expectations of models and project stakes are high, but there is a lack of confidence

→ The City of Montreal, the TRB Traffic Flow Committee, and others have started thinking about writing micro-simulation guidelines

→ We want your collaboration and participation in the thought process to make models robust and trust worthy

Definitions

2 - Definitions

Calibration

The act of adjusting a model so that it represents reality as closely as possible

- Optimization algorithm: method to choose the best set of calibration parameters
- Calibration parameters: Subset of software parameters used for a given calibration
- Objective function: function that measures the distance between observed data and simulated results (Measure of performance [MOP])
- Calibration constraints: Contraints that define the subset of plausible parameter values

2 - Definitions

Objective function

Function that measures the distance between observed data and simulated results (Measure of performance [MOP])

The function must be chosen according to the nature of the data : simple data or statistical distribution from a deterministic or stochastic process, etc.

For example:

Simple Data		**Statistical Distribution**	
Mean Error	$\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)$	Komolgorov-Smirnov test	$\max_x(\lvert F_{obs} - F_{sim}\rvert)$

Ex : saturation Flow

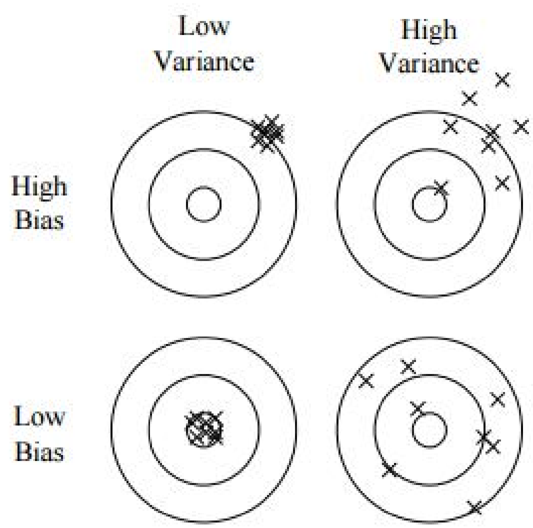
Ex : Travel time

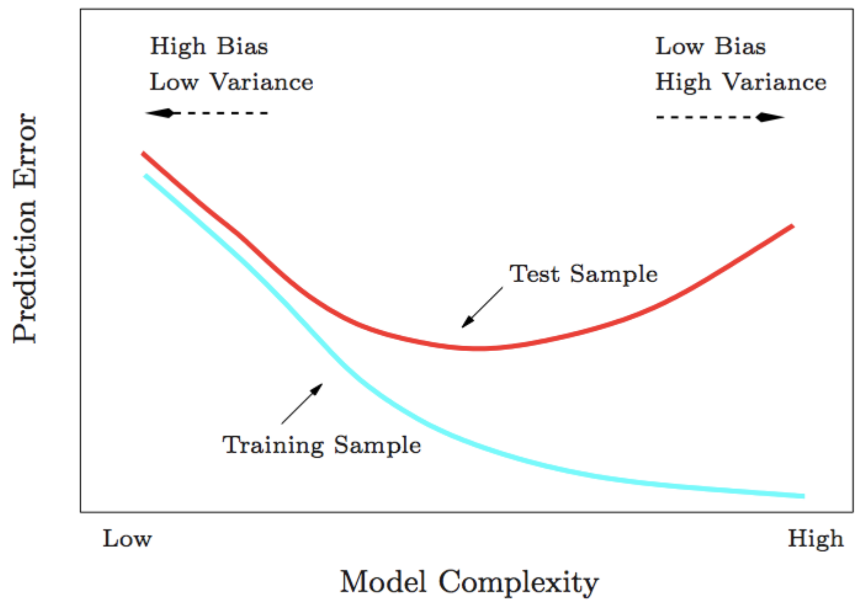
2 - Definitions

Validation

Validation is the act of measuring the generalization error, which means to measures the performance of the model on validation data, different from the calibration data.

There is a tradeoff between « calibration/overfitting» and « generalization », which can be decomposed in variance and biais:





From: Hastie et al. (2013). *The Elements of Statistical Learning; Data Mining, Inference, and Prediction* (2nd ed. 10th printing). Springer, New York, New York, USA

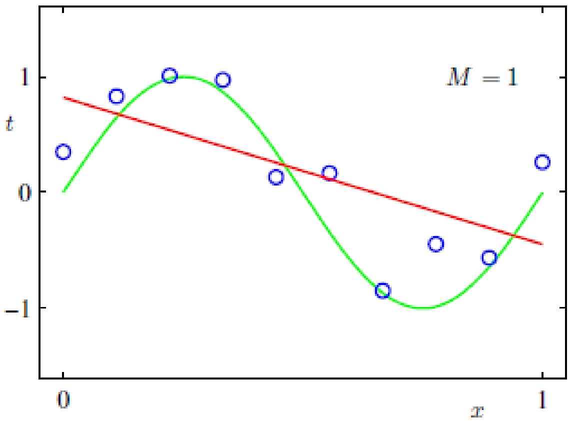
2 - Definitions

Validation

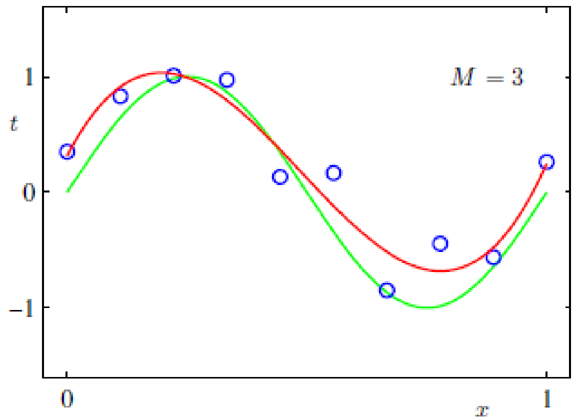
Validation is the act of measuring the generalization error, which means to measures the performance of the model on validation data, different from the calibration data.

There is a tradeoff between « calibration/overfitting» and « generalization », which can be decomposed in variance and biais:

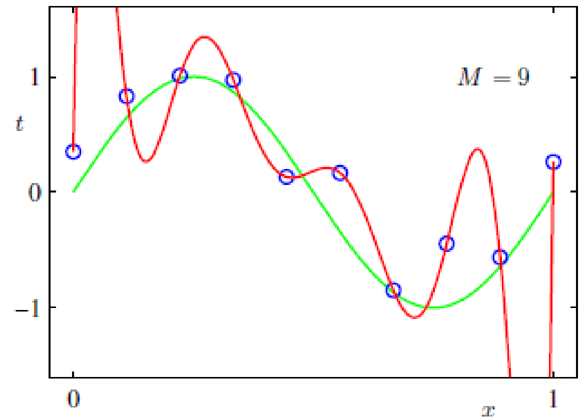
Underfitted



Well adjusted



Overfitted



From: Bishop, Christopher (2006). *Pattern Recognition and Machine Learning.* Springer, New York, New York, USA

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

WSP

State of the practice

3 – State of the practice: FHWA TAT Volume 3

→**Calibration is done in 3 steps : 1) Saturation, 2) Route choice, 3) Global performance (travel time, flows at intersection, etc.)**

→**Step 1 : Saturation**

- Parametres : headway times, stop distance, etc.
- Données : Saturation measurements at intersection
- Objective function:

$$MSE = \frac{1}{R}\sum_{r}(M_{ltpr} - F_{l})^2$$

- An optimization algorithm for 2 parameters (mean headway and mean reaction time) and 2 objectives (average speeds and average flow) is presented in appendix D

- **Steps 2 et 3 are described in a similar fashion**

- **Step 3 : "Since changes made at this step may compromise the prior two steps of calibration, these changes should be made sparingly**" (page 63)

POLYTECHNIQUE MONTRÉAL

LE GÉNIE EN PREMIÈRE CLASSE

WSP

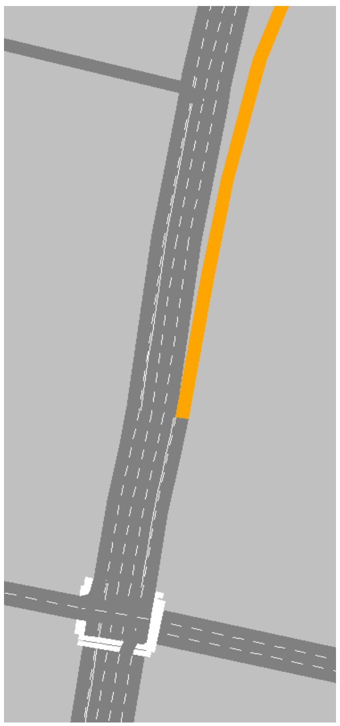
Issues with calibration and validation

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

WSP

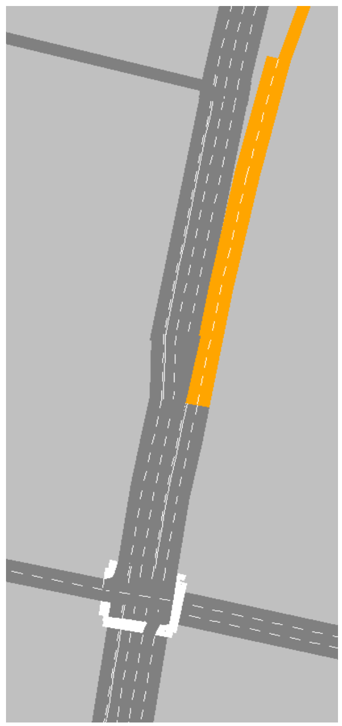
4 – Issues with calibration and validation Example

→ A public entity wants to build an off-ramp, and studies 2 scenarios single lane off-ramp (#1) or a two lane access (#2) : which is better ?



Scenario #1

Single lane



Scenario #2

Two lanes

→ The analysis will be made using two sets of parameters :

1) Default values

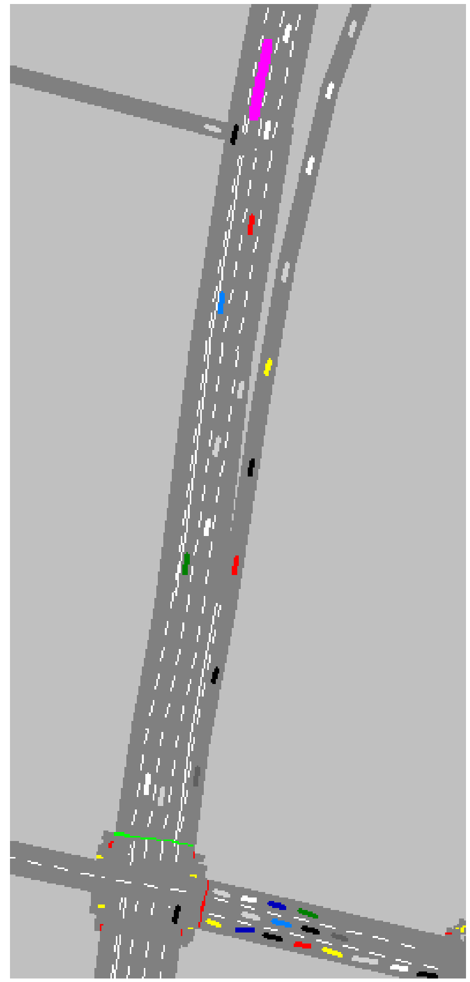
2) Two small “conservative” changes to perception distance and acceleration

4 – Issues with calibration and validation

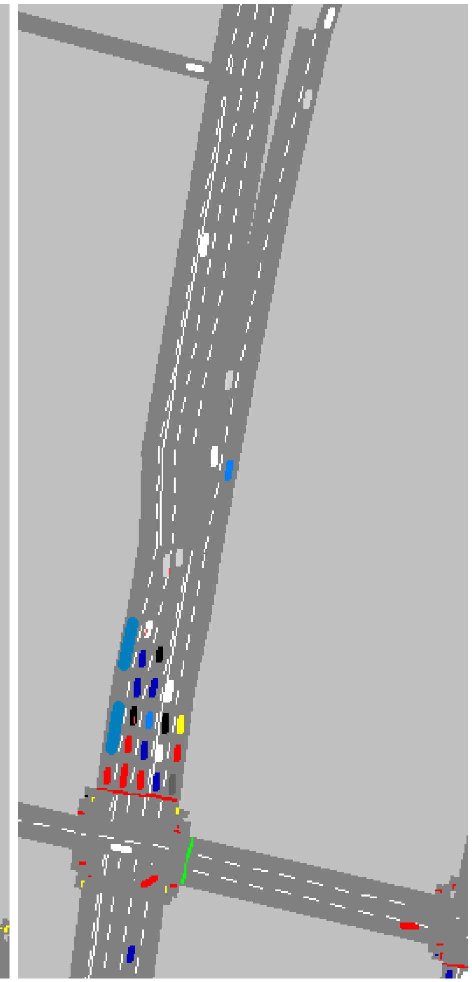
Example : Default values results

→ Best scenario = **2 lanes (scenario #2)**

→ Queues lengths are both shorter on the off-ramp and on the local street



Scenario #1

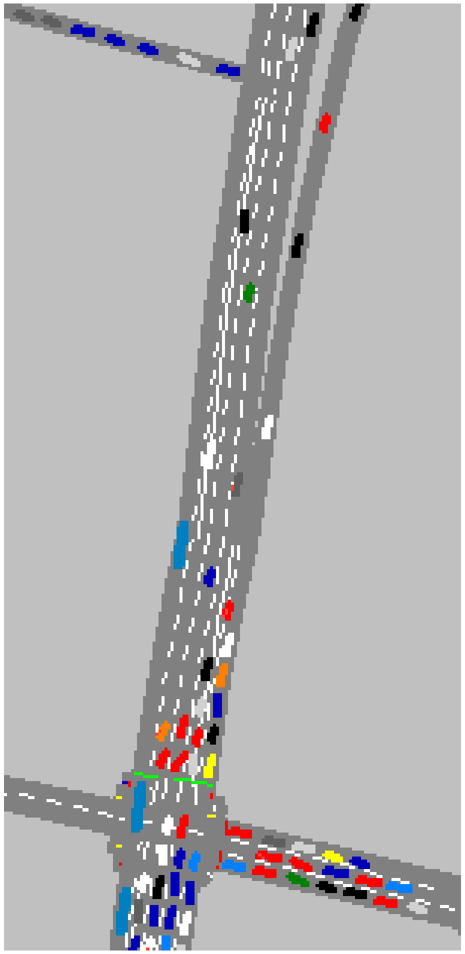


Scenario #2

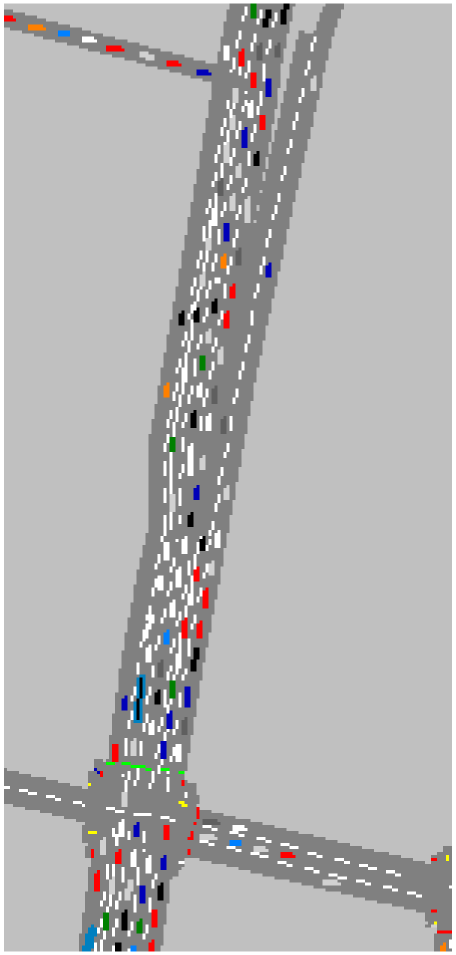
4 – Issues with calibration and validation

Exemple : Conservative values results

- → Best scenario = **1 lane (scenario #1)**
- → Queues on the off-ramp seem better in scenario 2, however:
- → Queues on the local network are longuer in scenario #2



Scénario 1



Scénario 2

4 – Issues with calibration and validation

Example : Conclusions

→ Which scenario should be recommanded ?

Both scenarios show possible simulation results, but it’s also possible that it doesn’t show how drivers would react at all.

Without calibration, it’s impossible to say which scenario should be recommanded or if both should be rejected

A calibrated and validated model cannot be exported to a different outset. A fortiori, default parameters (calibrated and validated for a specific setting) cannot be ported without adjustments.

Unmet needs

5 – Unmet needs

The European **MULTITUDE project** (Methods and tools for supporting the Use, caLibration and validaTIon of Traffic simUlations moDEls) (Antoniou et al. 2014) has identified five gaps

- lack of data
- lack of standardisation and definitions in basic methodology
- the need for illustration and comparison of case studies
- the variability of simulation results
- the need for assisted calibration, especially for automated sensitivity and batch analyses

5 – Unmet needs

Data accessibility

→ Models are over-fitted because datasets are limited

→ Validation is rarely conducted because data access is costly

→ Submodels are rarely calibrated because they lack specific data (lane-changing or car-following models)

→ To ensure statistical precision and robustness, practitionners need more data

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

WSP

5 – Unmet needs

Standardized guidelines

→ Guidelines represent a commun understanding of the required methodology for data collection and analysis. They help build trust in the models and their results.

→ Guidelines enable us to judge the worth of project, but also judge the calibration and validation methodology used.

Minimally, standardized guidelines should include

- Requirements for data collection
- Guidelines to chose the best software for the problem at hand
- A detailed procedure for calibration, validation and sensibility analysis
- A review of statistical tools and targets to achieve

5 – Unmet needs

Automated tools

→ Calibration and validation is a computional intensive task

→ Certain aspects required the engineer’s judgment, while other aspects require brute force

→ Automated tools gives the practitioner time to concentrate on tasks that require is full attention while taking care of the boring bulk of computing combinations of parameters.

Case study

6 – Case study: a work in progress

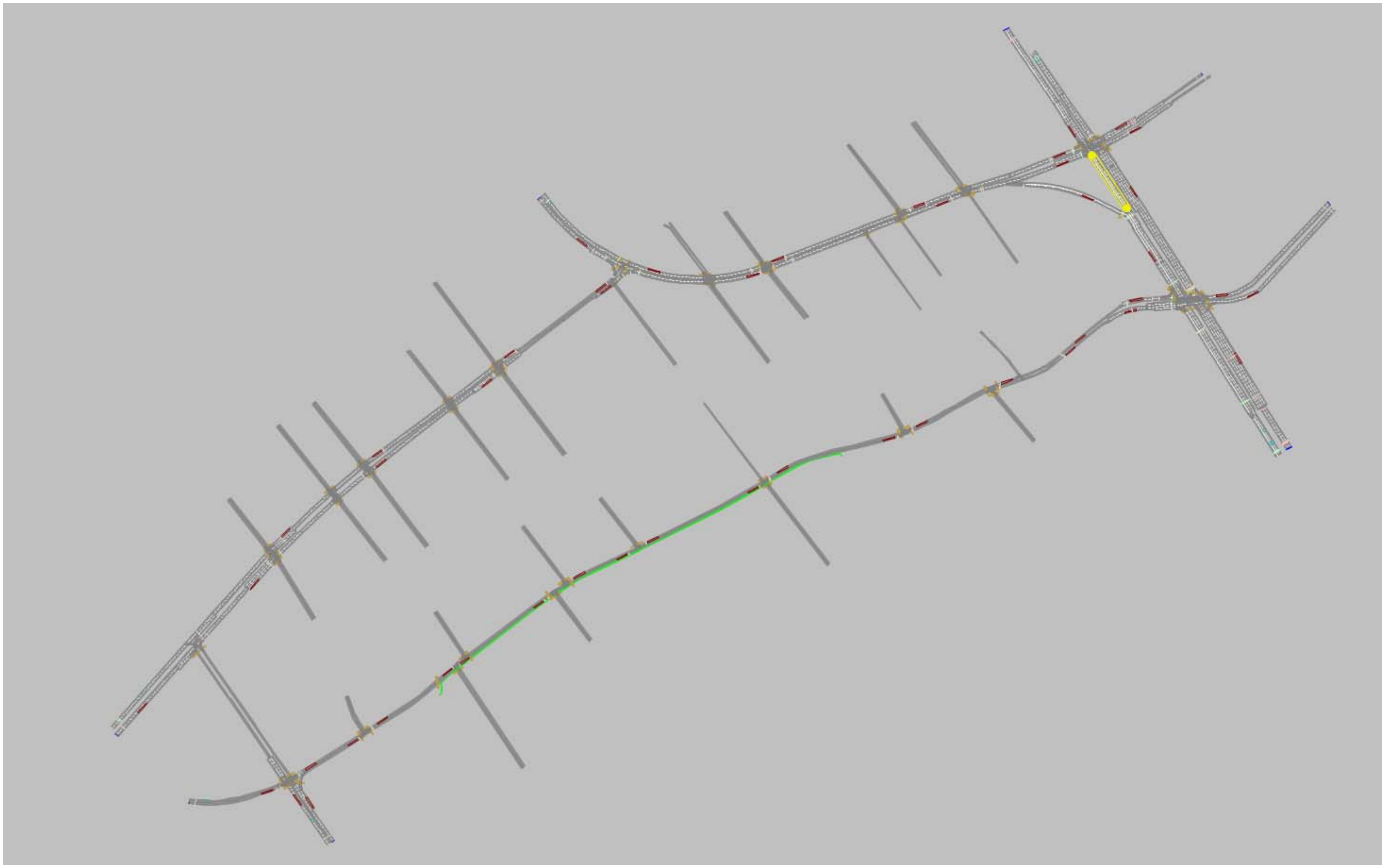
→ WSP et Polytechnique Montréal are collaborating on the calibration and validation process

→ A case study has, using an « industry-sized » model, is used to :

- Develop a methodology for calibration, validation and sensitivity analysis
- Develop automated tools for this methodology
- That can be used as a benchmark for further research
- That will serve as a discussion starter with practitionners

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

6 – Case study: calibration and validation



Conclusion

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

WSP

7 - Conclusion : we want you !

→ There are issues with calibration and validation :

- Lack of confidence if the process and its results

- Lack of data and standardized methodology

→ A case study is being tested by Polytechnique Montréal and WSP

→ More importantly, **we want practitioners to join us** in the thought process, to help make models more trust-worthy

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

WSP

→**francois.belisle@wspgroup.com**
→**laurent-2.gauthier@polymtl.ca**
→**nicolas.saunier@polymtl.ca**