Simulation-based optimization algorithms that enable the efficient use of inefficient traffic simulators for large-scale network optimization

Carolina Osorio
Civil and Environmental Engineering, MIT

ETHZ. Feb. 11, 2016

Work partially support by the US NSF Awards 1351512 & 1334304.

Collaborators

- Graduate Students
  - Nate Bailey
  - Linsen Chong
  - Evan Fields
  - Jing Lu
  - Chao Zhang
  - Kevin Zhang
  - Tianli Zhou
- Faculty
  - António Antunes (Uni. Coimbra)
  - Gunnar Flötteröd (KTH)

Past collaborators

- Students: Xiao Chen, Kanchana Nanduri, Krishna Selvam, Jana Yamani, Carter Wang
- Faculty: Cynthia Barnhart (MIT)

Outline

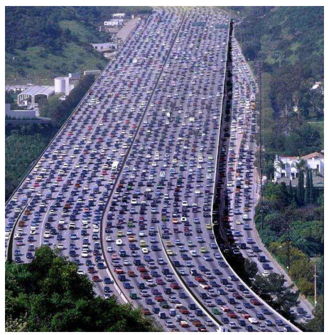
- General research goals and framework
- Simulation-based optimization (SO): main ideas
- Calibration problems
  - Toy network
  - Berlin metropolitan network
- Traffic management
  - New York City: QBB and MTM
  - Lausanne
- Some ongoing projects

Urban mobility research challenges

- Massive amounts and variety of high-resolution mobility data can now be collected
  - Smartphone apps, GPS, taxi data, connected vehicles

1. Understand complex traffic dynamics
   - How individuals make, and revise, travel decisions
2. Use this understanding to inform the design and operations of our networks.

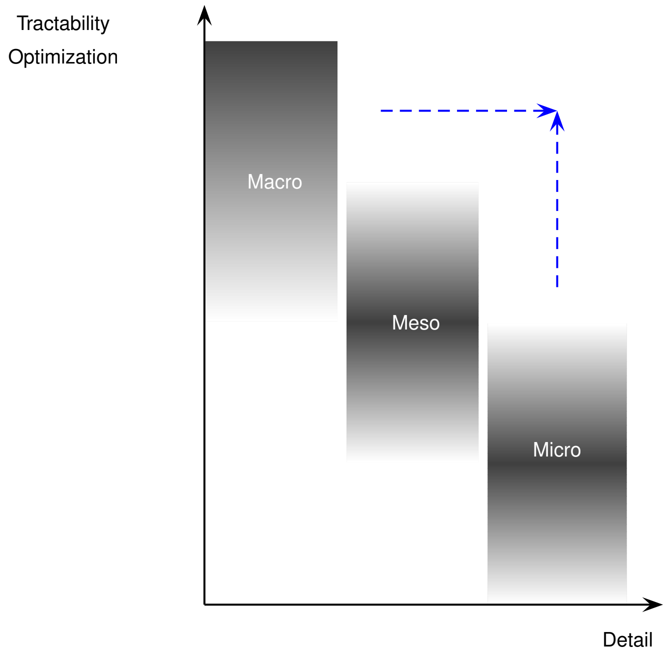
Goal: develop methods to enable the use of high-resolution knowledge/data, at the scale of the individual traveler or vehicle, to optimize urban networks at the scale of full cities or regions.



Research framework

Research goals:

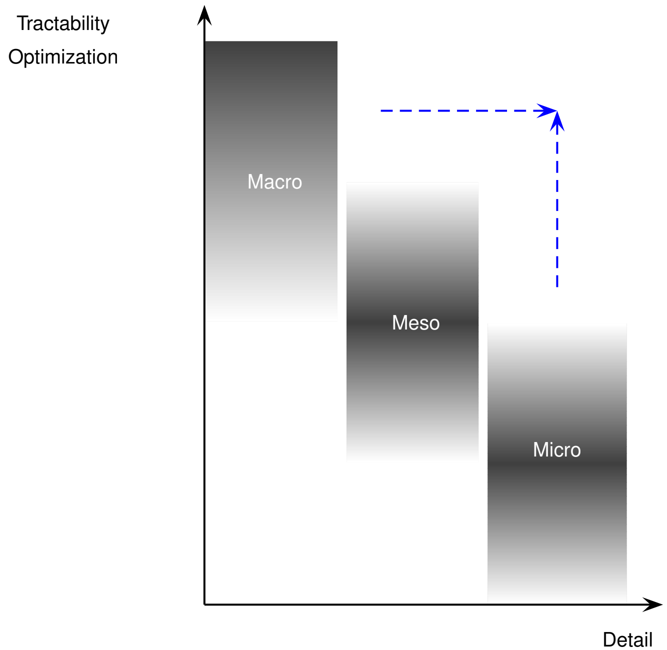
1. Develop efficient optimization methods for high-resolution models
2. Develop probabilistic macro. models
3. Develop optimization methods that enable the combined use of multiple traffic models



Research framework

Research goals:

1. Develop efficient optimization methods for high-resolution models
2. Develop probabilistic macro. models
3. Develop optimization methods that enable the combined use of multiple traffic models



Research framework

- Focus: analytical, macroscopic, probabilistic, scalable, tractable
- Approach: Traffic flow theory $\leftrightarrow$ Queueing network theory
- Stochastic LTM (link transmission model):

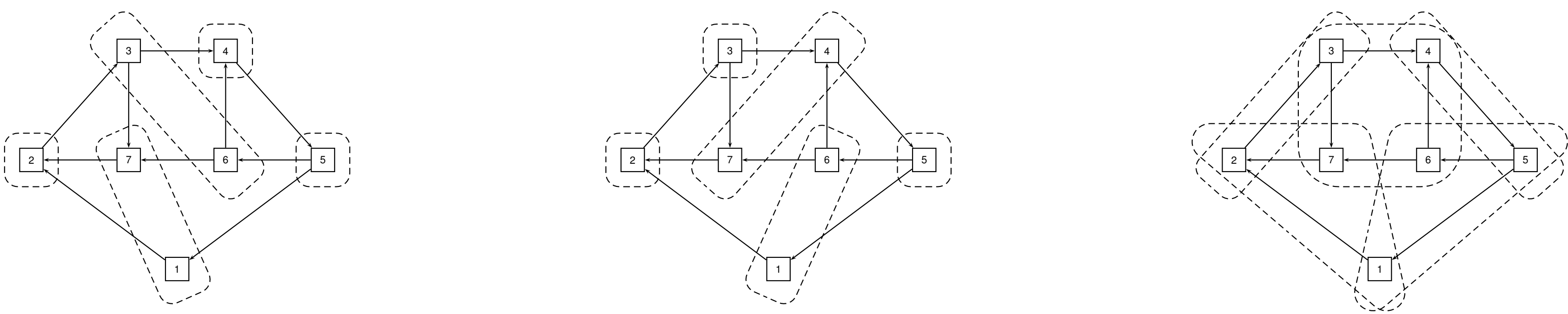
  Osorio and Flötteröd (2015) Transp. Science

  Lu and Osorio (2015) Proc. TRISTAN

  Osorio, Flötteröd and Bierlaire (2011) Transp. Res. Part B
- Markovian network models: *Osorio and Yamani (forthcoming) Transp. Science*

  Flötteröd and Osorio (2013) Proc. DTA

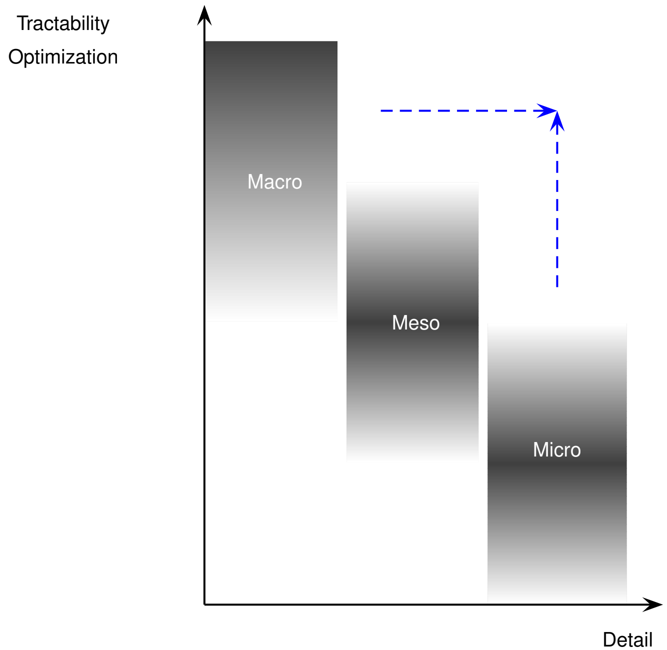
  Osorio and Wang (2012) Proc. EWGT
- Higher-order Little's law for congested networks: *Chen and Osorio (2014) Proc. EWGT*
- Purpose: stand-alone traffic models, auxiliary traffic models



Research framework

Research goals:

1. Develop efficient optimization methods for high-resolution models
2. Develop probabilistic macro. models
3. Develop optimization methods that enable the combined use of multiple traffic models



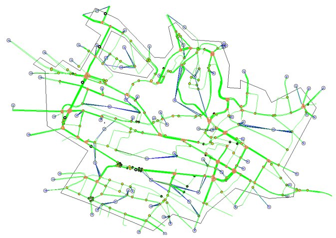
Research framework

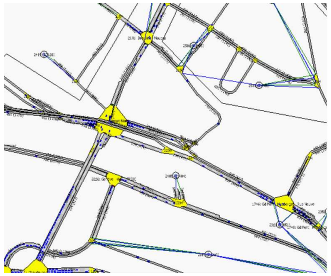
- Develop efficient optimization methods for high-resolution models
  - Model calibration problems
  - Network design and operation problems: traffic management

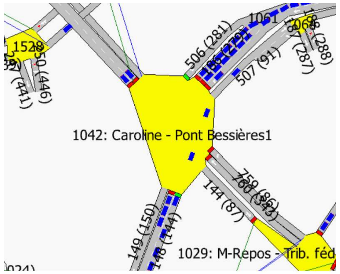
High-resolution traffic models

1. High-resolution: stochastic microscopic traffic simulators

- Probabilistic demand models (departure-time, mode, route, lane-changing)
- Detailed supply models (traffic management strategies)







High-resolution traffic models

2. Efficient optimization: tight computational budgets, of interest to practitioners

- Computationally costly to evaluate, stochastic outputs, no closed-form available for optimization
- Current use: what-if analysis
  - Relies on prior knowledge
  - Complexity of deriving a priori a strategy with good local and network-wide performance

How can complex stochastic simulation systems be used efficiently for optimization ?

Research area: simulation-based optimization

$$\min_{x \in \Omega} E[F(x, z; p)]$$

F: performance measure of interest (travel time, fuel consumption)

x: decision vector (green times)

z: endogenous simulation variables (queue-lengths)

p: exogenous simulation parameters (network topology, total demand)

- Problems of interest:
  - Objective function: simulation-based
  - Constraints: general form (non-convex), analytical, differentiable
- Current methods:
  - Black-box approach: a large number of simulated observations are needed to identify points with improved performance
  - Asymptotic properties
  - Not practical

Research area: simulation-based optimization

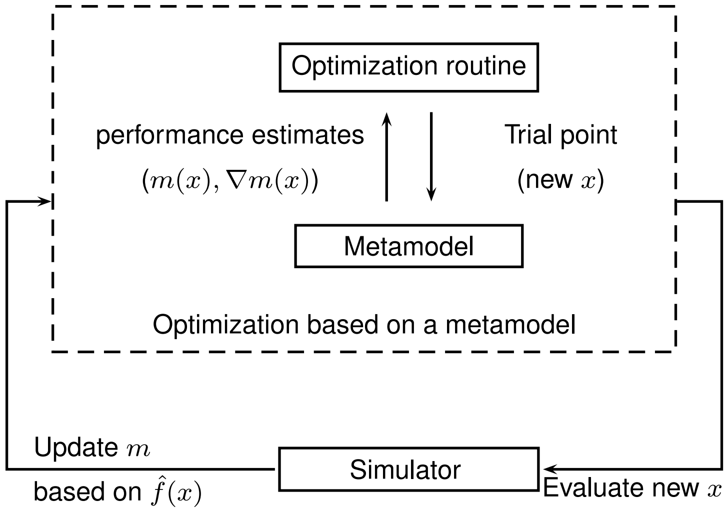
$$\min_{x \in \Omega} E[F(x, z; p)]$$

General objective

- To develop methods that:
  - Have good short-term performance: of interest to practitioners
    Tight computational budget
  - Are suitable for complex problems: high-dimensional, generally constrained, dynamic
  - Do not require a priori solution information
- Challenges: algorithmic, computational, modeling

Approach: metamodel simulation-based optimization

- *Osorio and Bierlaire (2013)* ***Operations Research***



$$\min_{x \in \Omega} E[F(x)] \quad \leftrightarrow \quad \min_{x \in \Omega \cap \Psi_k} m_k(x)$$

- The algorithms combine ideas from the fields of probability theory, simulation, simulation-based optimization, derivative-free optimization, nonlinear optimization, statistics, traffic control and traffic flow theory.

Example

$$\min_x E[F(x)]$$
$$x \geq c$$

$$\min_x \; m = \alpha f_A(x, y; q) + \phi(x; \beta)$$
$$x \geq c$$
$$h(x, y; q) = 0$$

- f_A: global approximation, problem structure, tractable, prior knowledge, derivatives
- ϕ: correction term, asymptotic properties
- Advantage: Achieves an excellent detail-tractability trade-off
- Challenges:
  - Derive problem-specific tractable formulations for f_A
  - Evaluation time of $h \ll$ Simulation run time

Metamodel

$$\min_{x \in \Omega} E[F(x)] \quad \leftrightarrow \quad \min_{x \in \Omega \cap \Psi_k} m_k(x)$$

- Metamodel: functional + physical

$$m(x, y; \alpha, \beta, q) = \alpha f_A(x, y; q) + \phi(x; \beta)$$

- Combination of:
  - f_A: the analytical (macro) traffic model
  - ϕ: a quadratic polynomial

x : decision vector

y : endogenous macro. model variables (e.g., link densities, queue lengths)

q : exogenous macro. model parameters (e.g., network topology, total demand)

α, β : metamodel parameters

$$\phi(x; \beta) = \beta_0 + \sum_{j=1}^{d} \beta_j x_j + \sum_{j=1}^{d} \beta_{d+j} x_j^2$$

Metamodel

$$m(x, y; \alpha, \beta, q) = \alpha f_A(x, y; q) + \phi(x; \beta)$$

- At each iteration k the parameters β and α of the metamodel are fitted using the current sample by solving the least squares problem:

$$\min_{\alpha,\beta} \sum_{i=1}^{n_k} \left\{ w_{ki} \left(\hat{f}(x^i, z^i; p) - m(x^i, y^i; \alpha, \beta, q) \right) \right\}^2 + (w_0.(\alpha - 1))^2 + \sum_{i=1}^{2d+1} (w_0.\beta_i)^2$$

x^i: i^{th} point in the sample

$\hat{f}(x^i, z^i; p)$: corresponding simulated observation

w_{ki}: weight associated to the i^{th} observation

w_0: fixed weight for augmented data

Calibration Problem

$$\min_{\tilde{\theta}\in\Omega} f(\tilde{\theta}) = \sum_{a,k} \left(y_{a,k} - E[F_{a,k}(\tilde{\theta}; z)] \right)^2 \tag{1}$$

- $\tilde{\theta}$: vector of cailbration parameters
- $y_{a,k}$: number of vehicles counted in reality on link a in time step k.
- $E[F_{a,k}(\tilde{\theta}; z)]$: simulation-based expected number of vehicles counted on link a in time step k.
- z: exogenous simulation parameters (e.g., network topology)

Calibration Problem

$$\min_{\tilde{\theta}\in\Omega} f(\tilde{\theta}) = \sum_{a,k} \left(y_{a,k} - E[F_{a,k}(\tilde{\theta}; z)] \right)^2 \qquad (2)$$

Challenging problem:

1. No closed-form expression available for $E[F]$: estimation through stochastic simulation.
2. Each simulation run is computationally costly.
3. $E[F_{a,k}(\tilde{\theta}; z)]$ lacks sound mathematical properties (e.g., convexity)
4. Difficult optimization problem: several local minima, identification of physically plausible solutions is difficult.
5. Large-scale problem: the dimension of $\tilde{\theta}$ is in the order of 100,000.

Calibration Problem

- Extensive research in the field of calibration of (stochastic) traffic simulators.
- For a survey, cf. Balakrishna (2006)
- Most approaches resort to techniques that are:
  - black-box optimization techniques
  - exploit little problem structure
  - are designed to achieve asymptotic properties

- Can we design efficient calibration algorithms?

Calibration Problem

- Can we design efficient calibration algorithms?
- **General idea:** Embed information from efficient (e.g., analytical, differentiable) macroscopic models within the calibration algorithm to to design efficient calibration algorithms.
- Example

$$\min_x E[F(x)]$$
$$x \geq c$$

$$\min_x \; m = \alpha f_A(x, y; q) + \phi(x; \beta)$$
$$x \geq c$$
$$h(x, y; q) = 0$$

Metamodel

- Calibration of one route choice parameter
- Single time interval
- Calibrate parameter such as to fit link flows on a subset of links

$$\min_{\tilde{\theta} \in \Omega} f(\tilde{\theta}) = \sum_a \Big(y_a - E[F_a(\tilde{\theta}; z)] \Big)^2 \quad (3)$$

$$\min_{\theta \in \Omega} \sum_a \Big(y_a - m_a^k(\theta; \beta) \Big)^2 \quad (4)$$

- Metamodel for link a: $m_a(\theta; \beta) = \beta_{i,0}\lambda_a(\theta) + \beta_{i,1} + \beta_{i,2}\theta$

  $\tilde{\theta}$: calibration parameter

  θ: (approximated) calibration parameter

  β: metamodel parameter

  $\lambda_a(\theta)$: flow on link a as approximated by an analytical macroscopic traffic model

Metamodel

- Metamodel for link a: $m_a(\theta; \beta) = \beta_{i,0}\lambda_a(\theta) + \beta_{i,1} + \beta_{i,2}\theta$
- $\lambda_a(\theta)$: problem-specific approximation of $E[F]$.
- Linear term: general-purpose approximation of $E[F]$.

Macroscopic traffic model

$$\begin{cases} \lambda_i & = & \gamma_i + \sum_j p_{ji}\lambda_j \\ \rho_i & = & \frac{\lambda_i}{\mu_i} \\ E[N_i] & = & \frac{\rho_i}{1-\rho_i} - \frac{(\ell_i+1)\rho_i^{\ell_i+1}}{1-\rho_i^{\ell_i+1}} \\ E[T_i] & = & \frac{E[N_i]}{\lambda_i} + \frac{l^{\text{veh}}(\ell_i - E[N_i])}{v^{\text{free-flow}}}. \end{cases}$$

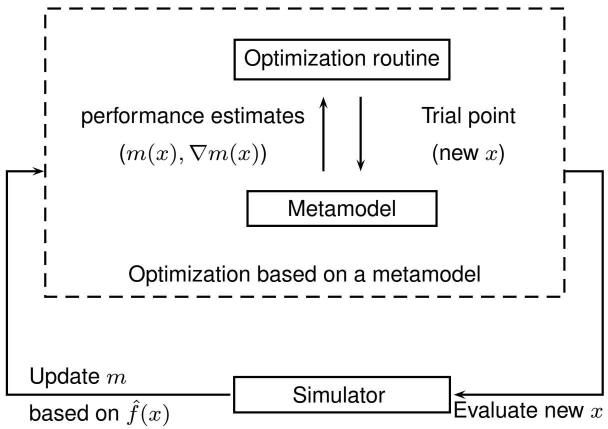
$$\begin{cases} c_p & = & \sum_i r_{pi} E[T_i] \\ l_{sp} & = & \frac{e^{-\theta c_p}}{\sum_u e^{-\theta c_u}} \\ f_p & = & \sum_s d_s l_{sp} \\ \gamma_i & = & \sum_p \hat{r}_{pi} f_p \\ p_{ij} & = & \frac{\sum_{t \in \mathcal{G}_{ij}} f_t}{\sum_{t \in \mathcal{H}_i} f_p} \end{cases}$$

Algorithm

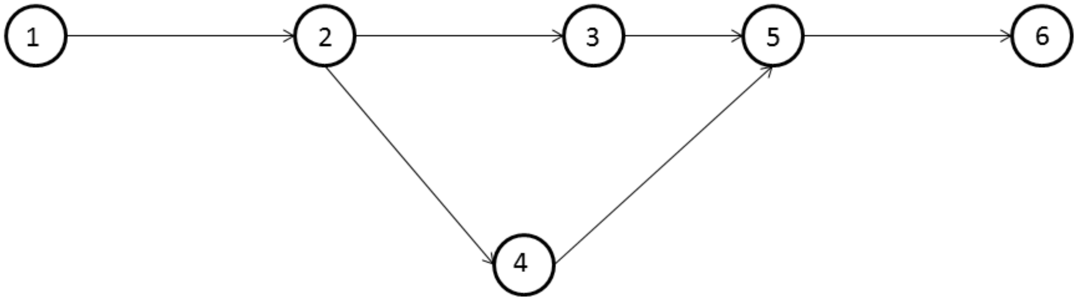
$$\min_{\tilde{\theta}\in\Omega} f(\tilde{\theta}) = \sum_a \Big(y_a - E[F_a(\tilde{\theta}; z)]\Big)^2 \quad (5)$$

$$\min_{\theta\in\Omega} \sum_a \Big(y_a - m_a^k(\theta, y; \beta)\Big)^2 \quad (6)$$

$$\text{s.t.} \quad h(\theta, y; q) = 0 \quad (7)$$



Simple case study



- Calibrate the travel time coefficient of an MNL route-choice model
- “True” flows: simulated
- Demand: 1400 veh/hr
- Comparison of two metamodels: with and without macroscopic traffic model information:
  1. $m_i(\theta;\beta) = \beta_{i,0}\lambda_a(\theta) + \beta_{i,1} + \beta_{i,2}\theta$
  2. $\phi_i(\theta;\beta) = \beta_{i,1} + \beta_{i,2}\theta$

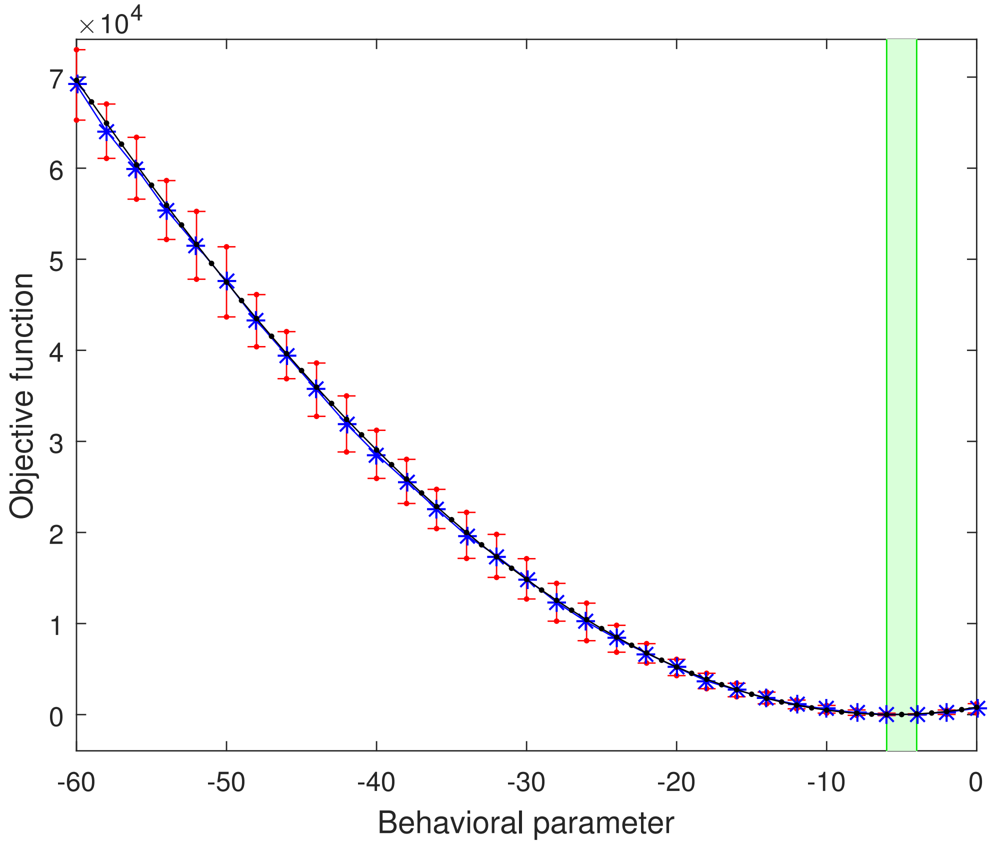
Simple case study

- True values [1/hr]: $\{-55, -20, -5\}$
- Three initial values [1/hr]: $\{0, -30, -60\}$
- Bounds: $[-60, 0]$

- m: with macro. model information (proposed approach)
- ϕ: without macro. model information

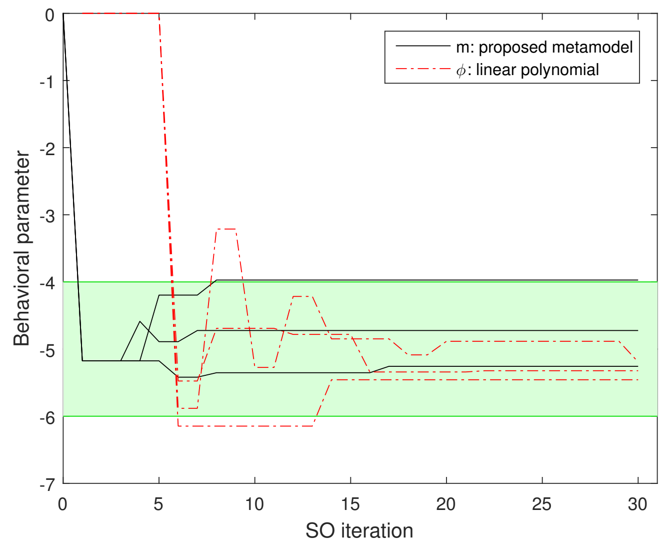
- For each experiment, the algorithm is run 3 times
- Each run allows for the evaluation of 30 θ values.
- For each value:
  - Assignment iterations: 50
  - Simulation replications: 5
- I.e., each algorithmic run allows for a total of 30*(50+5) = 1,650 simulation runs

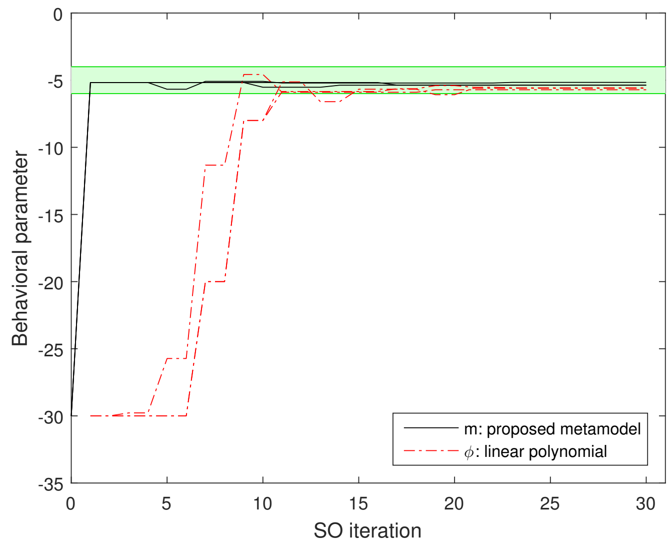
Simple case study

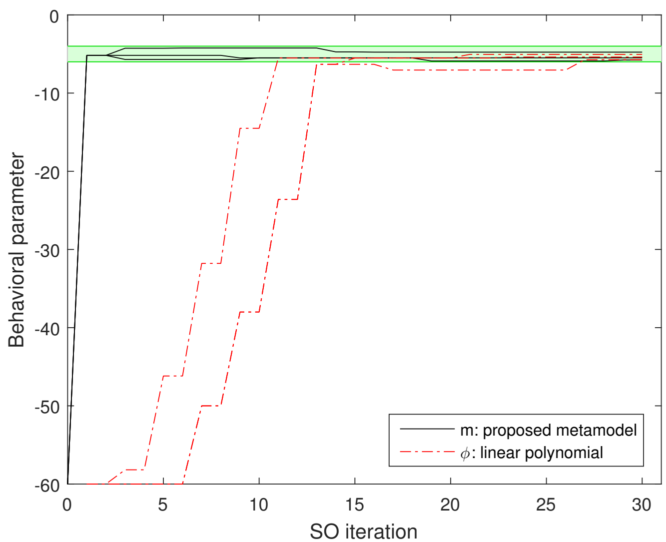


- True value: -5
- Objective functions: simulated and analytical

Simple case study

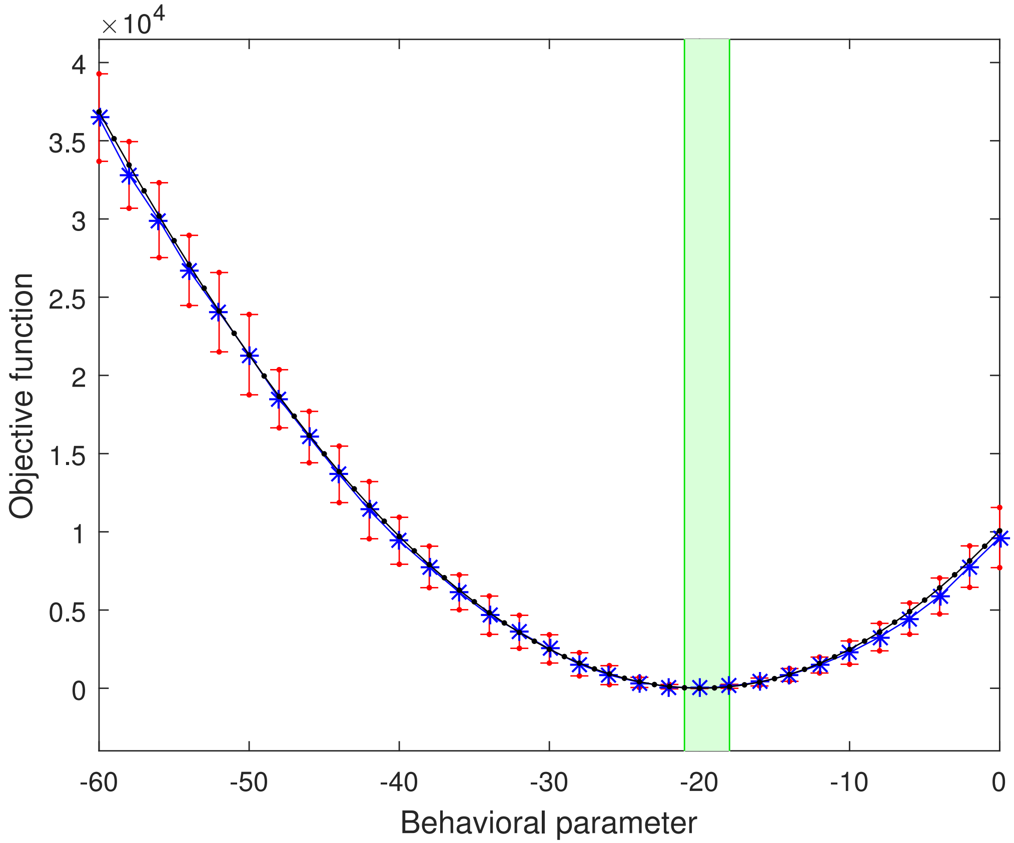






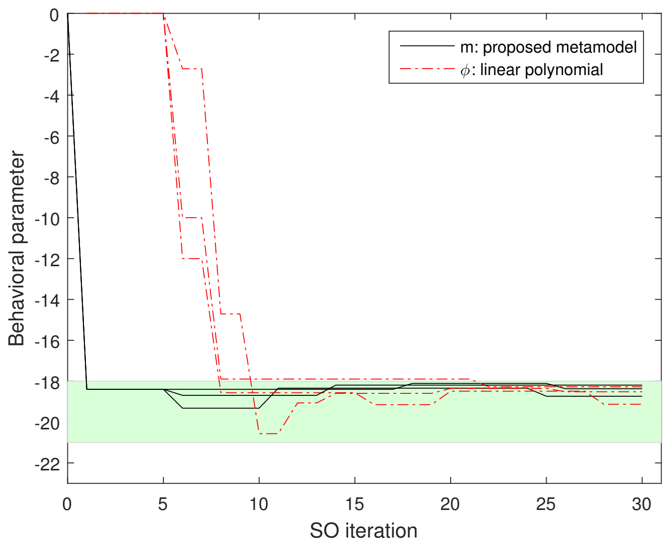
- True value: -5
- Three initial values: $\{0, -30, -60\}$

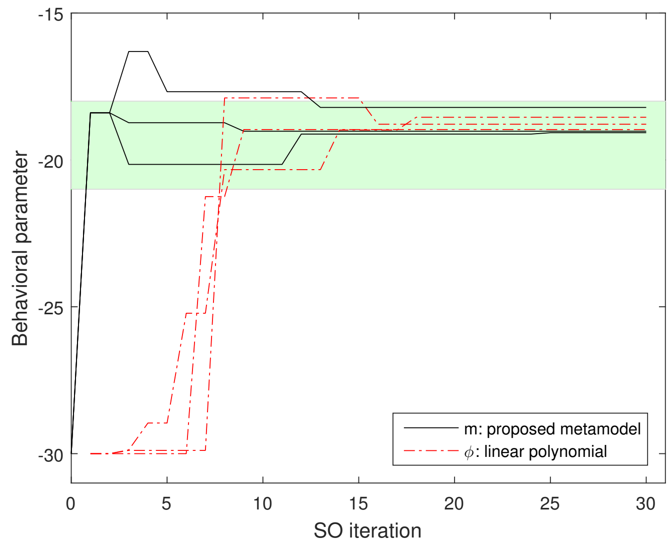
Simple case study

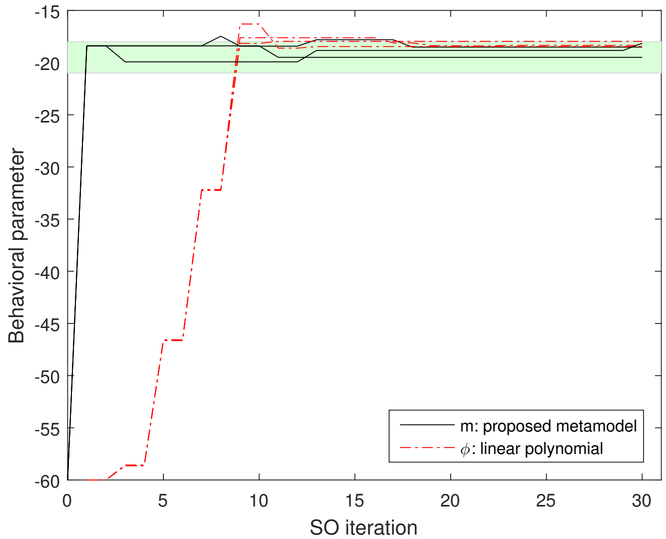


- True value: -20
- Objective functions: simulated and analytical

Simple case study

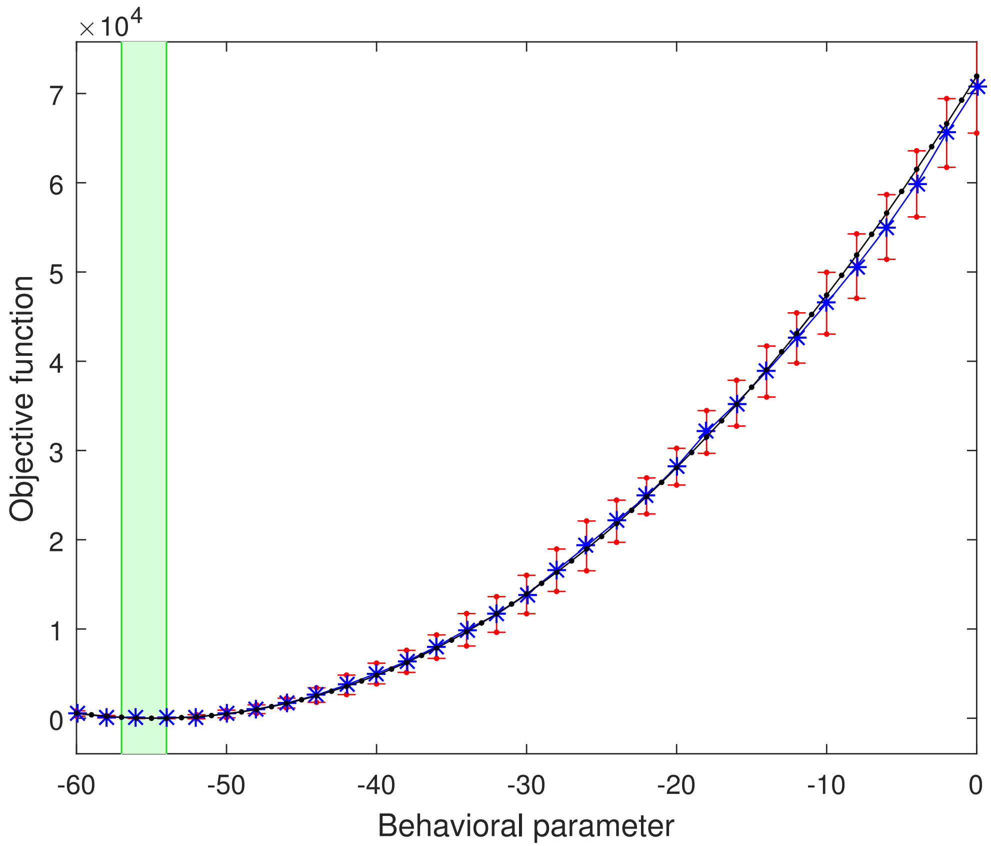






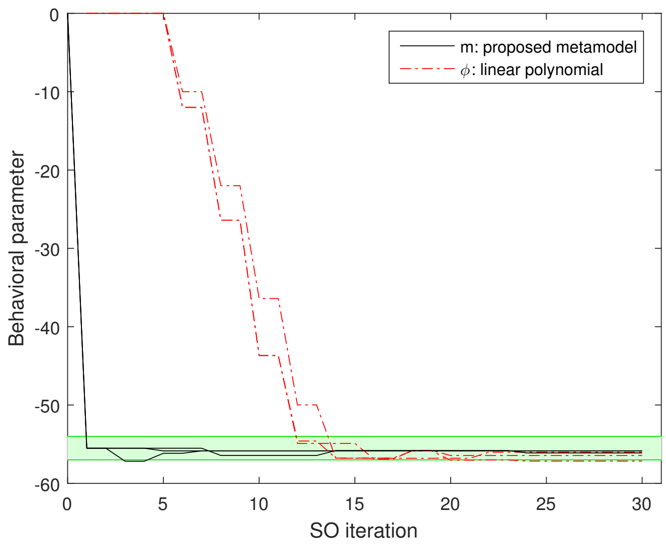
- True value: -20
- Three initial values: $\{0, -30, -60\}$

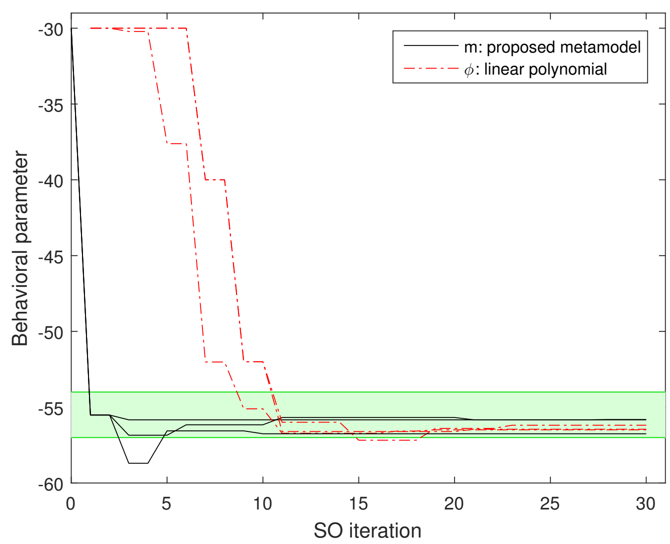
Simple case study

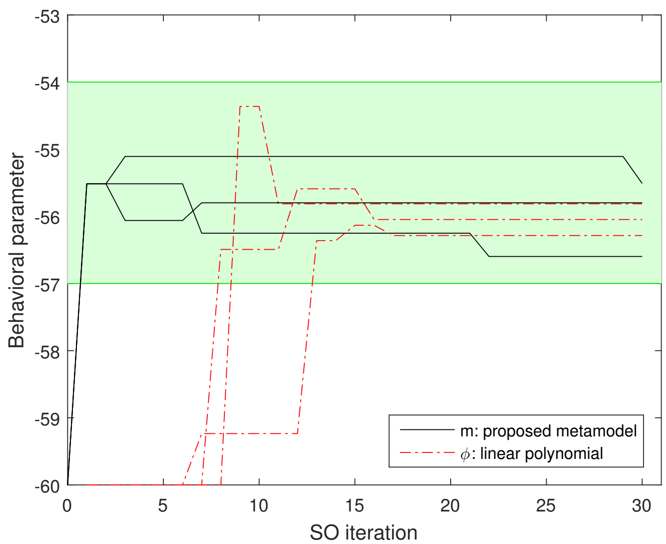


- True value: -55
- Objective functions: simulated and analytical

Simple case study

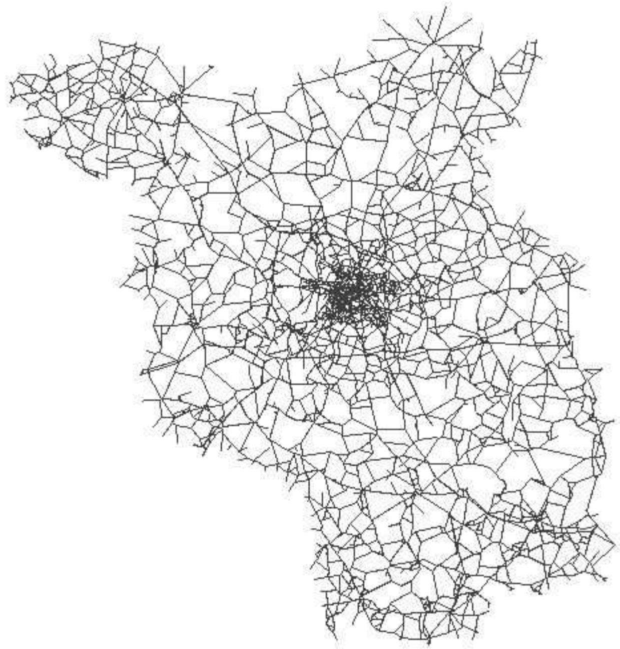






- True value: -55
- Three initial values: $\{0, -30, -60\}$

Berlin metropolitan area



- 24,335 links, 11,345 nodes,
- Morning peak: 8-9am
- Demand: 172,900 vehicles [veh/hr]
- Main challenge: scalability of the approach

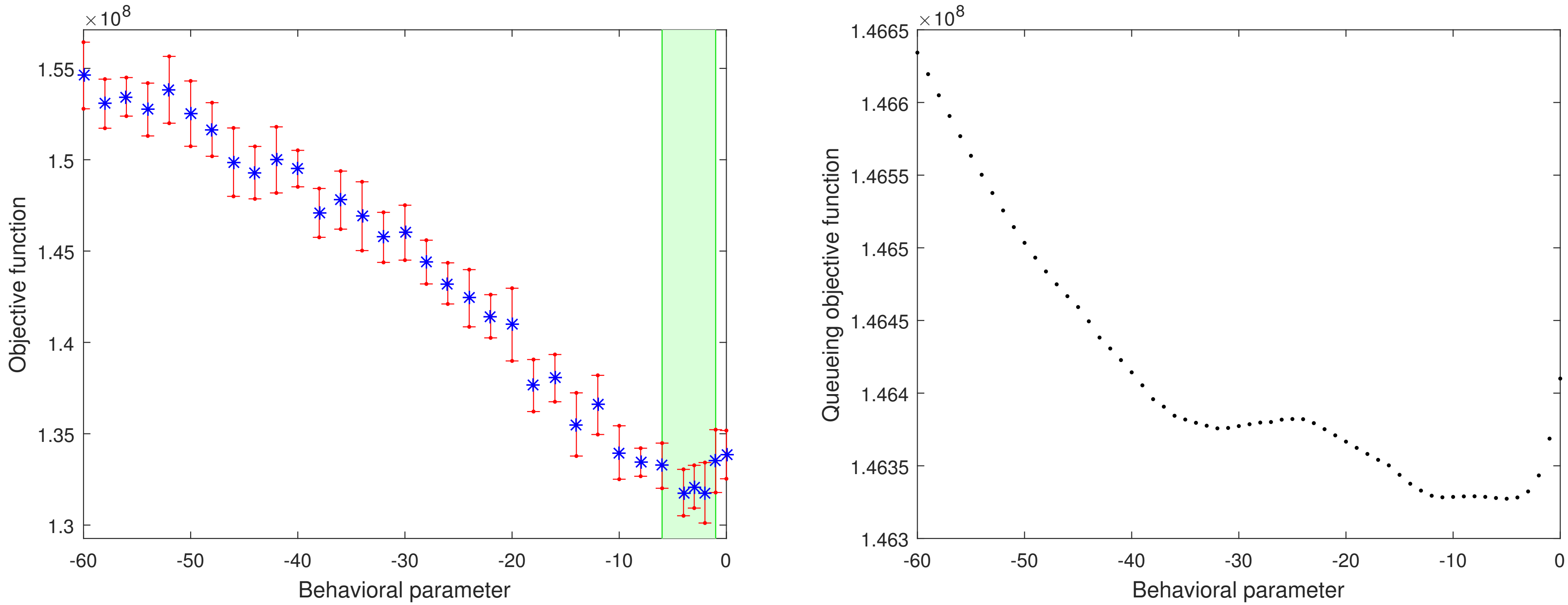
Berlin metropolitan area

- Three initial values [1/hr]: $\{0, -40, -60\}$
- Bounds: $[-60, 0]$

- m: with macro. model information (proposed approach)
- ϕ: without macro. model information

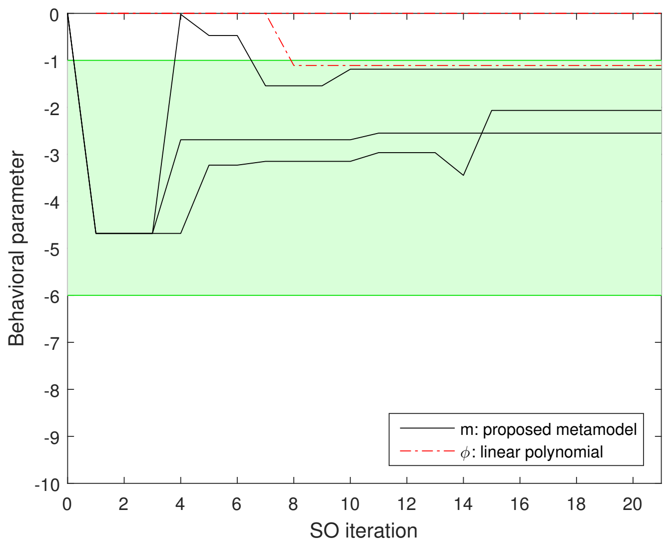
- For each experiment, the algorithm is run 3 times
- Each run allows for the evaluation of 20 θ values.
- For each value:
  - Assignment iterations: 100
  - Simulation replications: 10
- I.e., each algorithmic run allows for a total of 20*(100+10) = 2,200 simulation runs

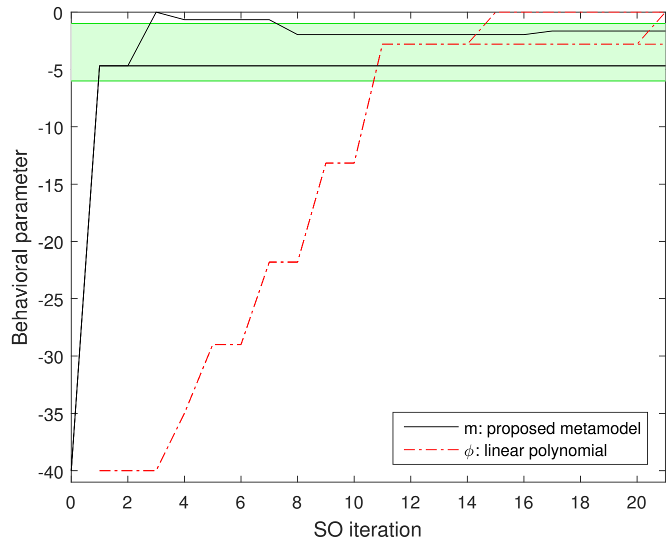
Berlin metropolitan area

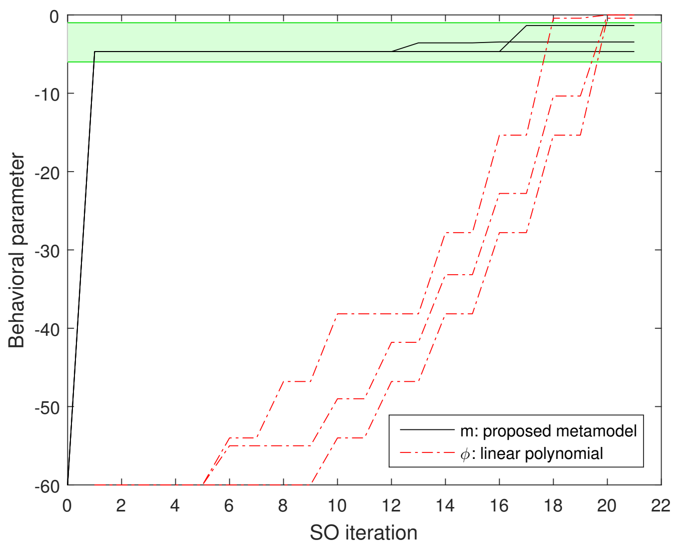


- Objective functions: simulated and analytical

Berlin metropolitan area







- Three initial values: $\{0, -40, -60\}$

Compuational savings

When averaging over all 9 experiments:

- Toy network: an average of 83% savings in runtime (35 minutes)
  Convergence on average at:

	iterations	run time [min.]
m	2.4	7.1
ϕ	14.1	42.2

- Berlin network: an average of 86% savings in runtime (30 hours)
  Convergence on average at:

	iterations	run time [min.]
m	2.4	293.3
ϕ	17.7	2120.0

Calibration Conclusions

- Design of computationally efficient calibration algorithms
- Efficiency is achieved through the use of a problem-specific efficient macroscopic traffic model
- Preliminary results on the toy and the Berlin metropolitan networks are promising
- Ongoing work:
  - Supply calibration
  - Combination of data-driven and metamodel methods
- Future work:
  - Higher-dimensional problems
  - Use of a scalable time-dependent macroscopic model
    Chong and Osorio (Submitted)
  - Fitting higher-order distributional metrics: e.g., higher-order moments of link flow

Research framework

- Develop efficient optimization methods for high-resolution models
  - Model calibration problems
  - Network design and operation problems: traffic management

Transportation problems of interest

1. High-dimensional problems, large-scale micro. models

   Osorio and Chong (2015) Transp. Science

2. Use of instantaneous vehicle performance

   Osorio and Nanduri (2015) Transp. Science

   Osorio and Nanduri (2015) Transp. Part B

3. Use of higher-order distributional information: reliable and robust problems

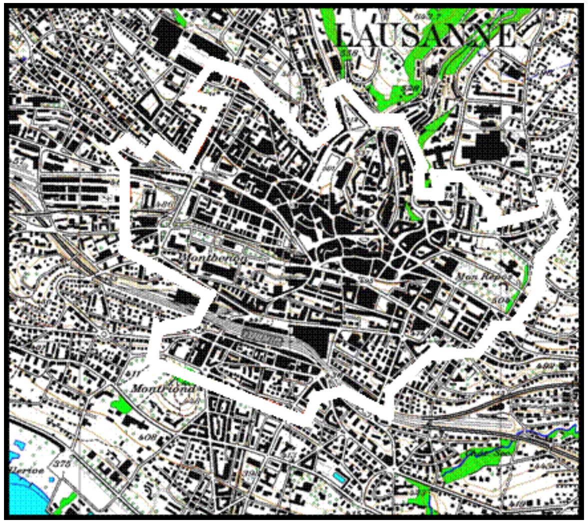
   Osorio, Chen and Santos (2012) Proc. INSTR

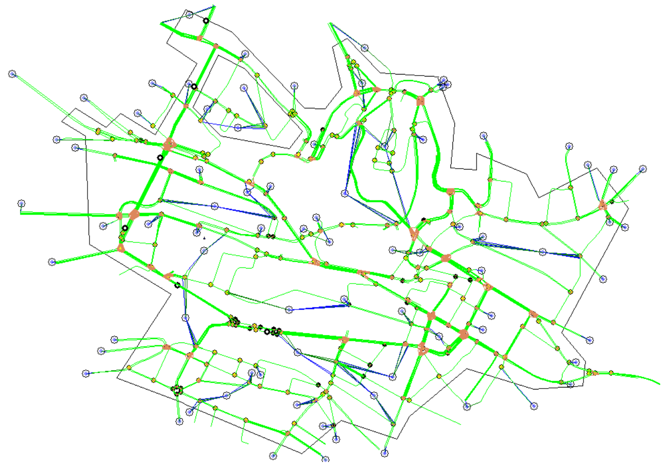
4. Dynamic problems

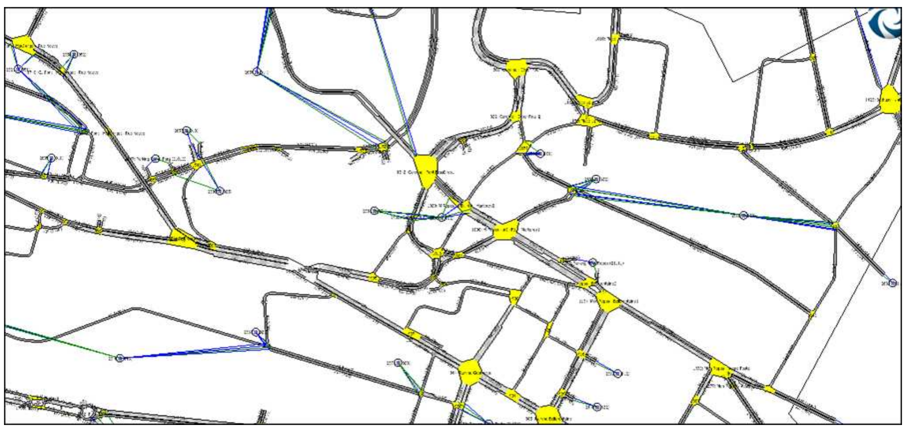
   Chong and Osorio (Submitted)

   Osorio, Chen and Santos (2015) Proc. TRB

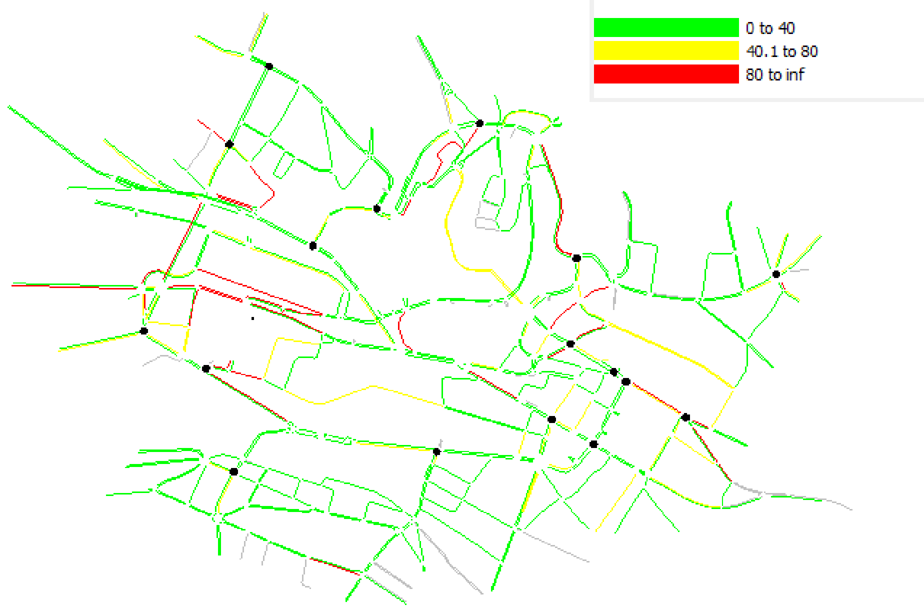
Large-scale simulation-based optimization







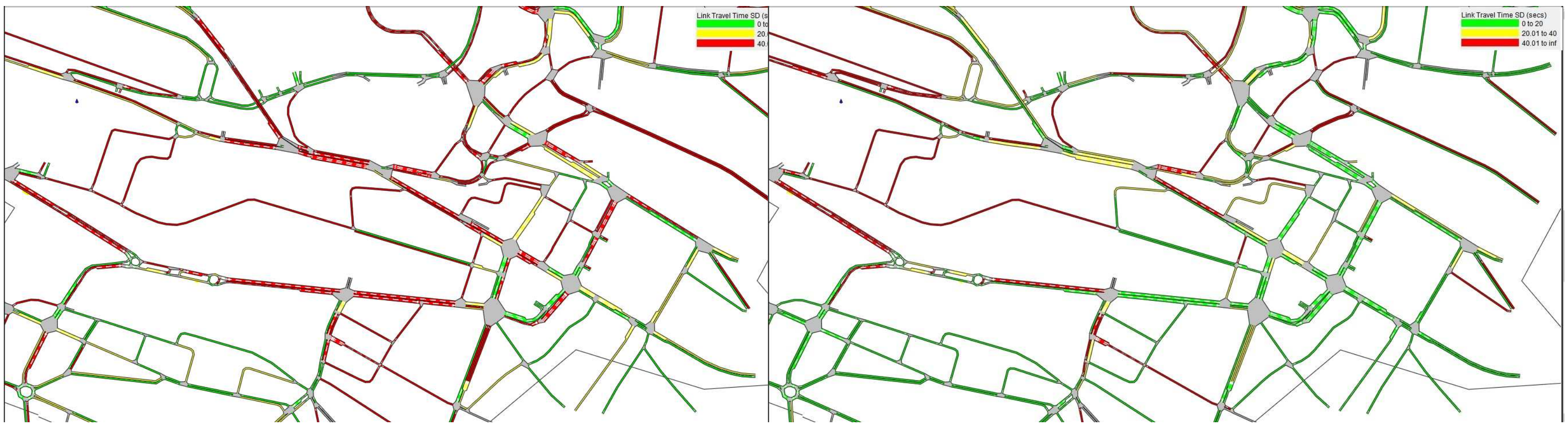
Large-scale signal control



- 603 links and 231 intersections
- Travel demand: 12400 trips, 2076 origin-destination pairs
- Signal control: city-wide performance, 17 intersections, 99 endogenous phases
- What can be done with only 150 simulation runs?
- Osorio and Chong (2015) *Transp. Science*
- Media: MIT News, IEEE Spectrum, Smithsonian Magazine

Reliable and robust operations

- Rethink how we measure network performance
- Go beyond the use of first-order moment information
- Enhancing network reliability is a critical goal of major transportation agencies
- Travel time variability: important attribute in route choice and mode choice



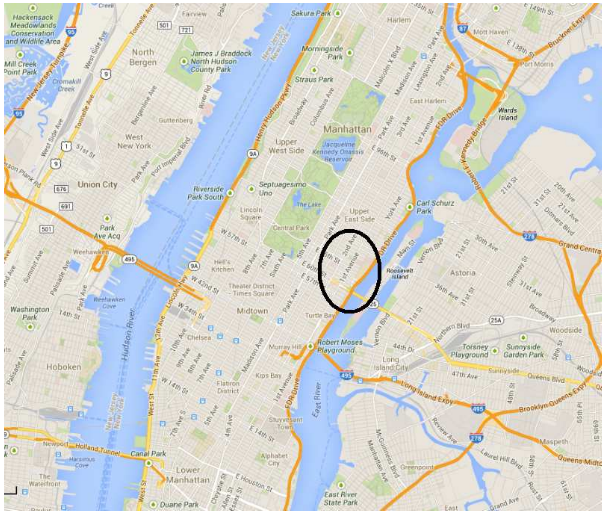
- Use of instantaneous vehicle performance

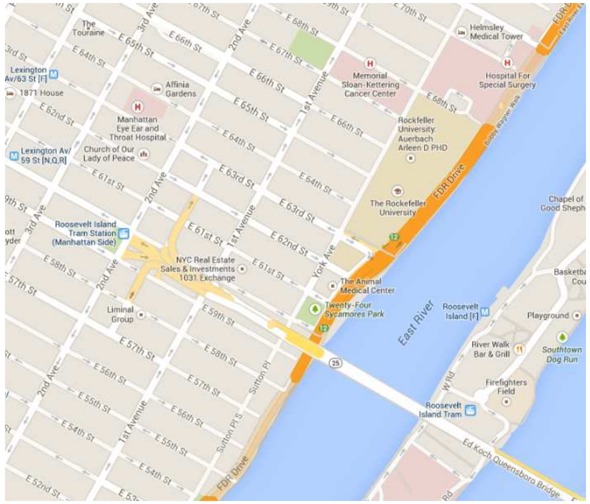
  Osorio and Nanduri (2015) Transp. Science

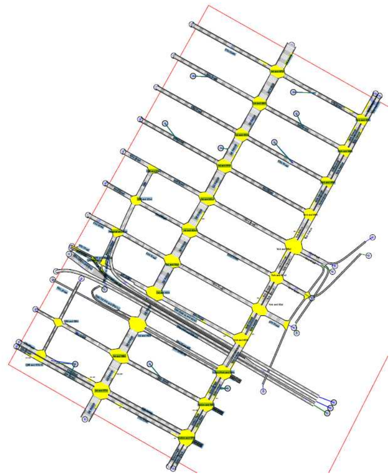
  Osorio and Nanduri (2015) Transp. Res. Part B

  MIT News, March 2015

New York City: signal control

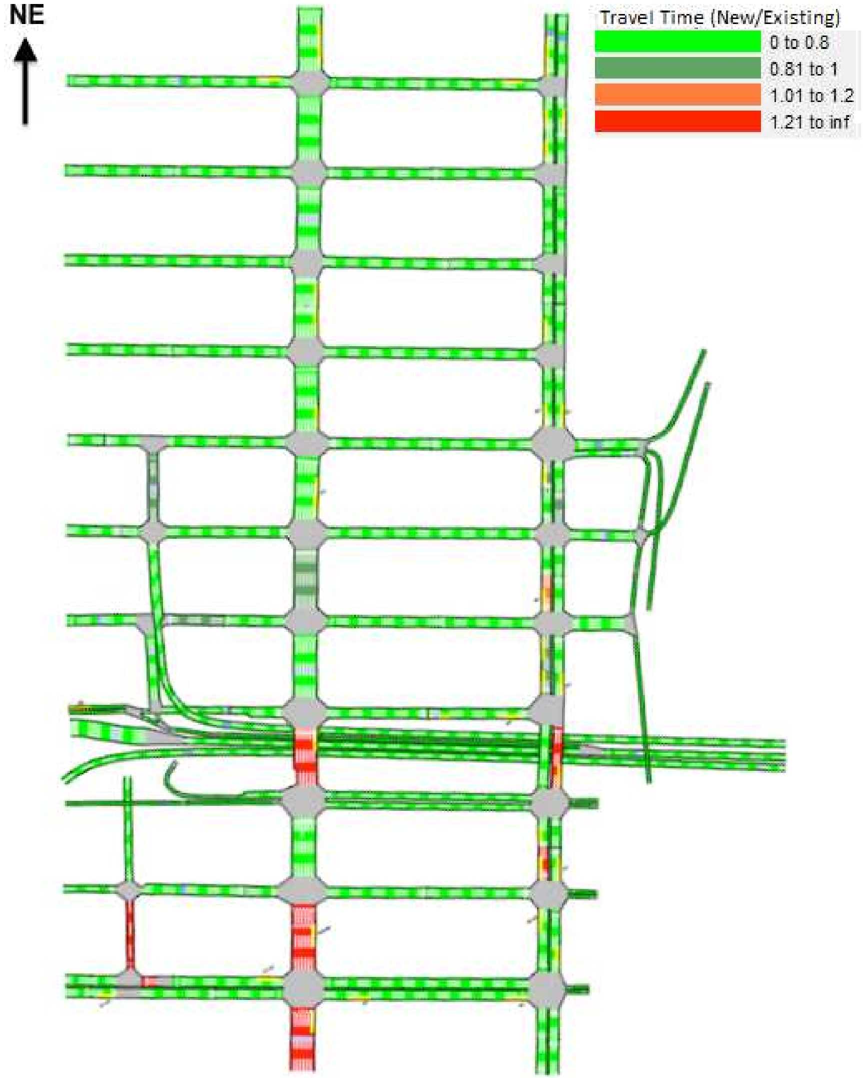






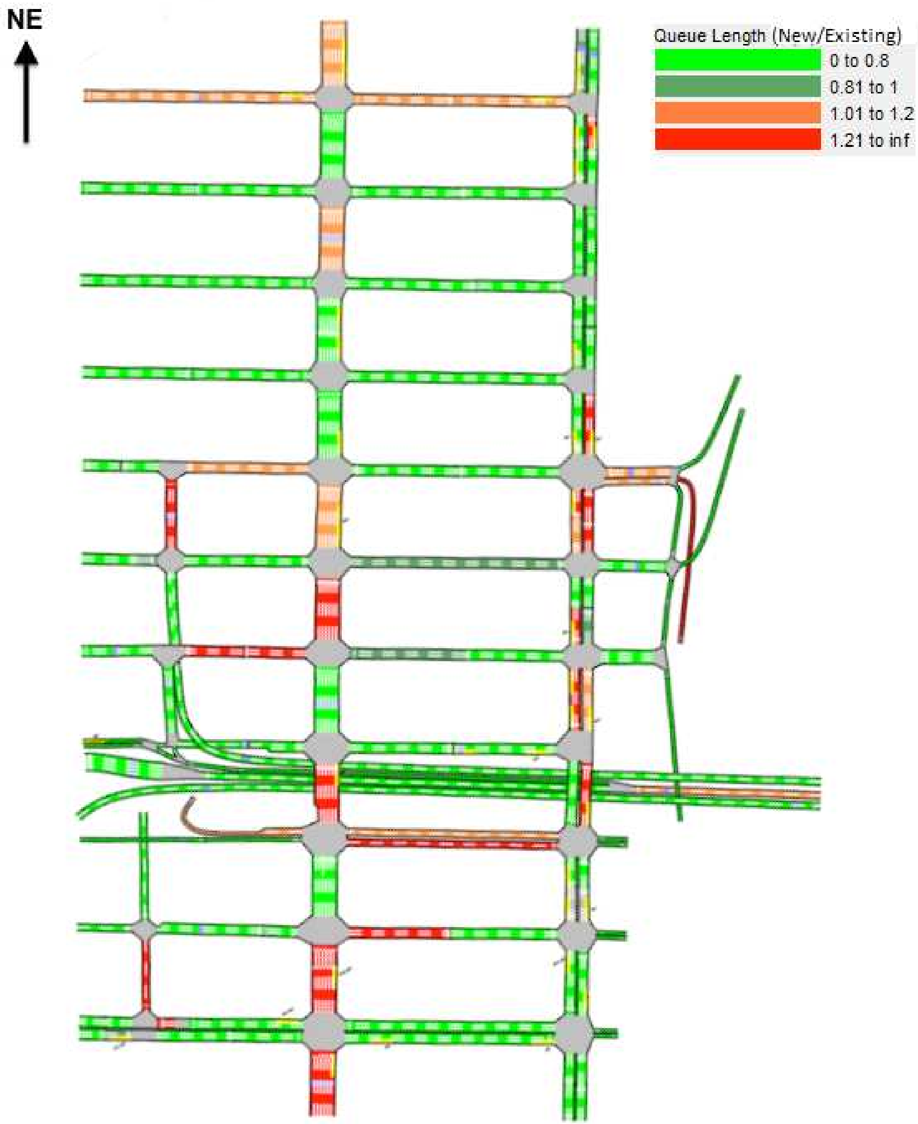
- New York City Department of Transportation (NYCDOT)
- Morning peak: 8-9am, an average of over 11,000 vehicle trips
- 134 Roads, 41 intersections, 26 controlled.
- Intricate traffic dynamics: highly congested, multi-modal, high pedestrian traffic, grid topology, short links, complex travel behavior (e.g., high dimensional route alternatives)
- Critical area: connects Queens to Manhattan through the Queensboro Bridge.
- Spillbacks along access/egress links can have significant large-scale impacts

Average link travel time



- Average travel time per link: $\frac{\text{TT(proposed)}}{\text{TT(existing)}}$
- The smaller the ratio, the larger the improvement
- Legend:
  green: reduction of more than 20%
  dark green: reduction within 0 to 20%
  orange: increase within 0 to 20%
  red: increase within 0 to 20%

Average link queue-length



- Average queue-length per link: $\frac{\text{QL(proposed)}}{\text{QL(existing)}}$
- The smaller the ratio, the larger the improvement
- legend:
  green: reduction of more than 20%
  dark green: reduction within 0 to 20%
  orange: increase within 0 to 20%
  red: increase within 0 to 20%

QBB network results

- Compared to the existing signal plan, the proposed fixed-time plan:
  - Reduces average trip travel time by 10%
  - Reduces average queue-length by 28%
  - Reduces average spillback probability by 23%
  - Increases average throughput by 2%

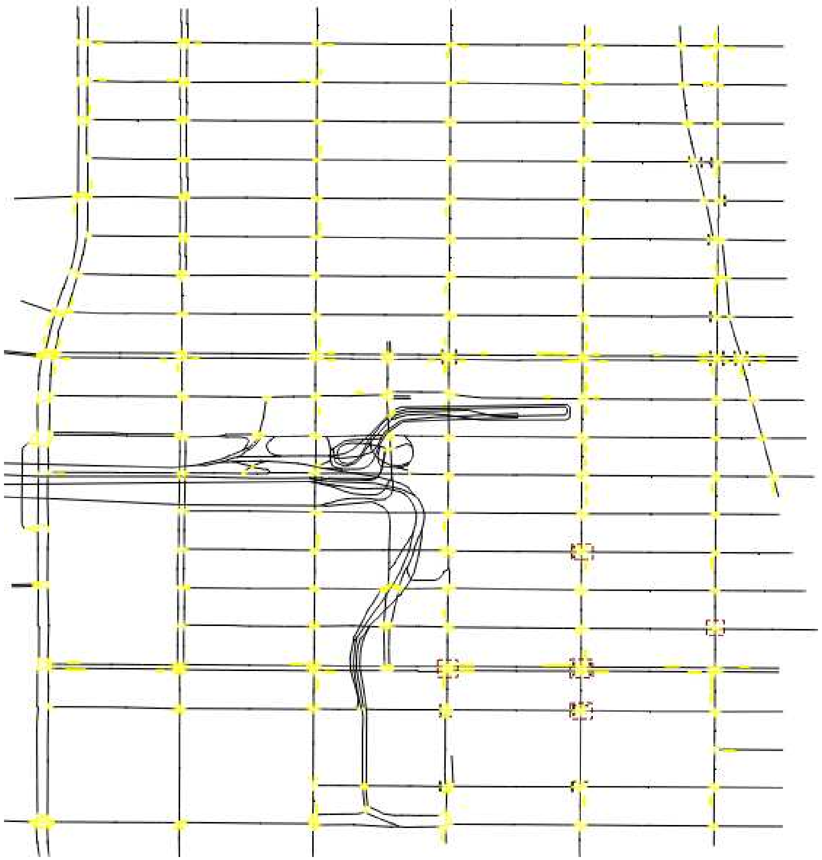
- Traffic-responsive signal control
  - Reduces average trip travel time by 7%
  - Reduces average queue-length by 27%
  - Reduces average spillback probability by 44%
  - Increases average throughput by 8%

QBB: methodological lessons

- Detailed modeling or estimation of between-link interactions is critical for the control of such complex networks
- Work showed the importance of providing the algorithm with an analytical description of between-link interactions
- Need of formulating analytical and differentiable macroscopic models that both quantify these interactions and can be used for large-scale optimization.

Larger scale control

- Based on the QBB insights: simplify the analytical traffic models, while preserving the description of between-link interactions



- 924 links
- 2600 lanes
- 444 nodes
- Over 28000 vehicular trips
- Optimization: 96 controlled intersections
- Simulation budget of 50 runs

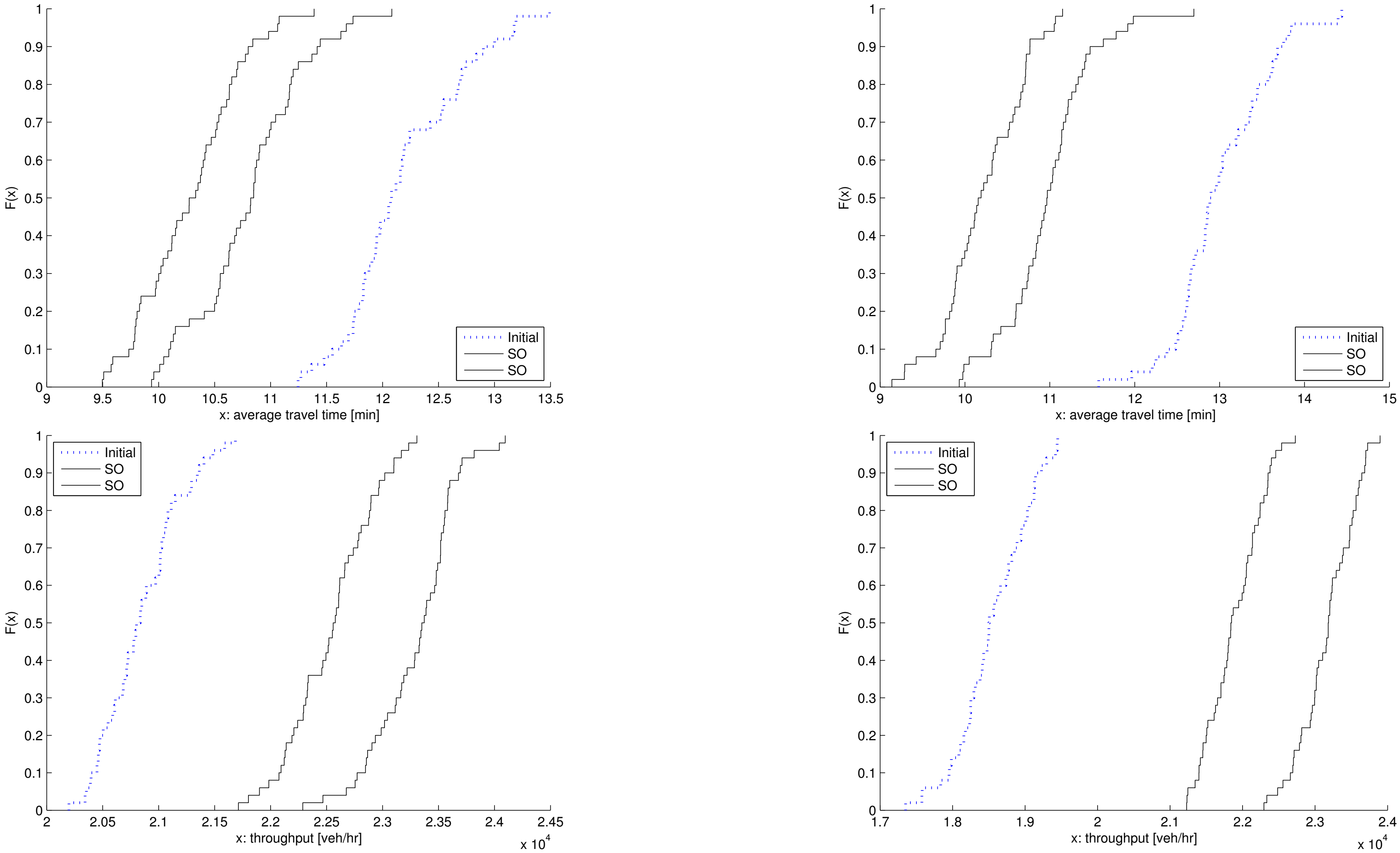
Larger scale control

- NYCDOT signal plan: average link density

Larger scale control

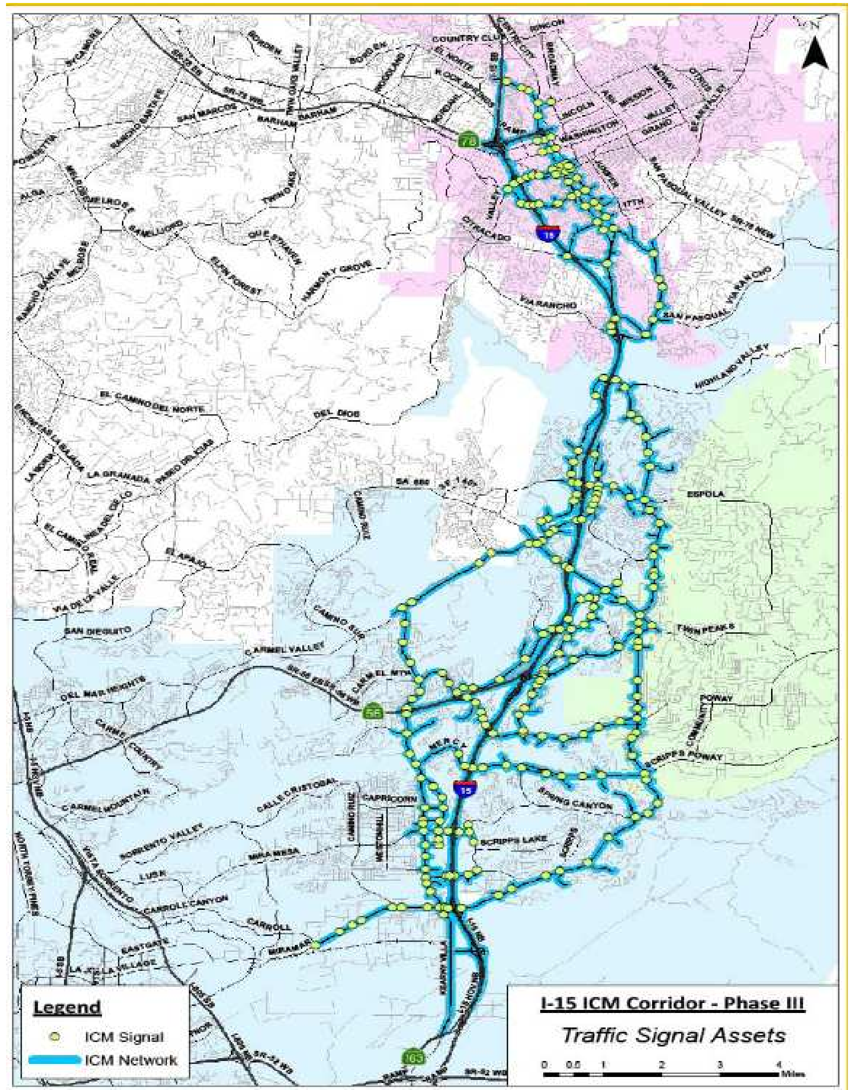
- NYCDOT signal plan: average link travel time

Larger scale control



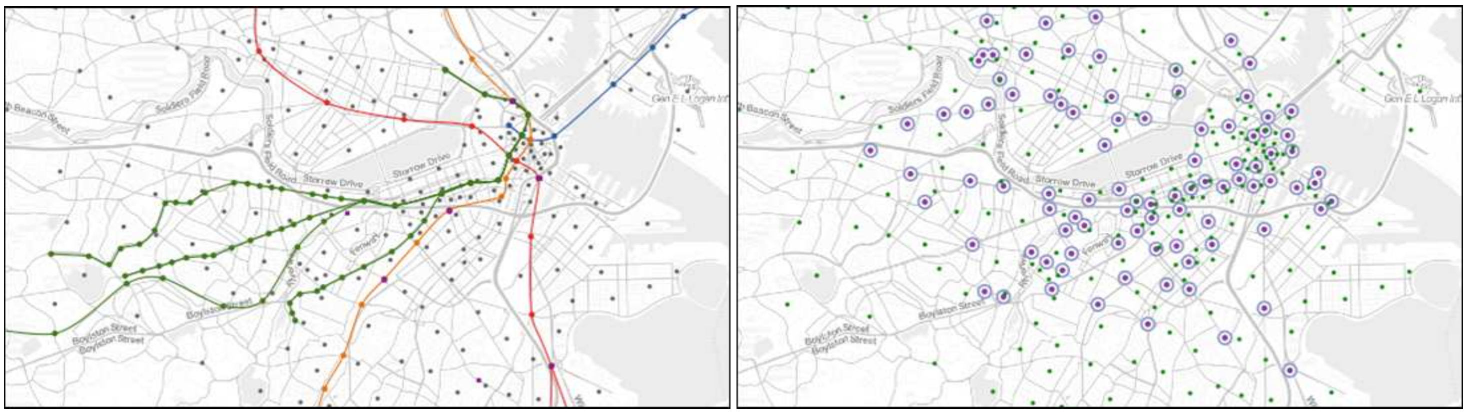
Towards real-time - San Diego Region

- Real-time large-scale control
- Encompasses three cities: Escondido, Poway, San Diego
- 60 minute forecasts based on state-of-the-art traffic models



Other ongoing projects

- Algorithms for autonomous and mixed vehicle fleets
- Traffic control: San Diego I-15 corridor; SANDAG, TSS, City of Escondido.
- Vehicle-sharing network design
  - How can we design on-demand mobility services such as to complement existing services, such as transit?
  - Ford, ZipCar, City of Boston



References

- Osorio and Selvam (forthcoming). Simulation-based optimization: achieving computational efficiency through the use of multiple simulators. **Transp. Science**
- Osorio and Nanduri (2015). Urban transportation emissions mitigation: coupling high-resolution vehicular emissions and traffic models for traffic signal optimization. **Transp. Research Part B**
- Osorio and Nanduri (2015). Energy-efficient urban traffic management: a microscopic simulation-based approach. **Transp. Science** 49(3):637-651.
- Osorio and Chong (2015). A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems. **Transp. Science** 49(3):623-636.
- Osorio and Flötteröd (2015) Capturing dependency among link boundaries in a stochastic network loading model. **Transp. Science** 49(2):420-431.
- Chen, Osorio, Marsico, Talas, Gao and Zhang (2015) Simulation-based adaptive traffic signal control algorithm. **Proc. TRB**
- Osorio, Chen, Marsico, Talas, Gao and Zhang (2014) Reducing gridlock probabilities via simulation-based signal control. **Proc. ISTS**
- Osorio and Chong (2014). A simulation-based optimization algorithm for dynamic traffic control problems. **Proc. TSL Workshop**
- Osorio and Bierlaire (2013). A simulation-based optimization framework for urban transportation problems, **Operations Research**, 61(6):1333-1345
- Chen, Osorio and Santos (2012). A simulation-based approach to reliable signal control. **Proc. INSTR**
- Osorio, Flötteröd and Bierlaire (2011). Dynamic network loading: a differentiable model that derives link state distributions, **Transp. Research Part B**, 45(9):1410-1423