Combining simulation-based and analytical traffic models to mitigate urban congestion

Carolina Osorio – Michel Bierlaire

Civil and Environmental Engineering Department
Massachusetts Institute of Technology (MIT)

Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne (EPFL)

INFORMS TSL Workshop
June 28, 2011

Context

- Congested urban networks: complex traffic dynamics.
- To derive and evaluate traffic management schemes for congested networks: microscopic traffic simulators
- These simulators embed numerous detailed and realistic traffic models
- But this realisms leads to functions that are:
  - nonlinear
  - stochastic
  - expensive to evaluate
  - with no closed form available
  - gradient estimations are also more involved (noise, availability of source code)

Integrating these simulators within an optimization context is an intricate task

Context

- **Simulation models**
  - capture complexity of the network: realistic
  - difficult to optimize
- **Analytical models**
  - simple models with sound mathematical properties
  - optimization friendly: analytical gradients are available, less expensive to evaluate
- Objective:
  Use detailed simulation models to efficiently identify appropriate traffic control schemes
  To perform efficient simulation-based optimization (SO)
- Approach: to combine information from: simulation model + *analytical* model

1) Preserve realism.
2) Achieve tractability: capture structural information analytically

How can we combine them?

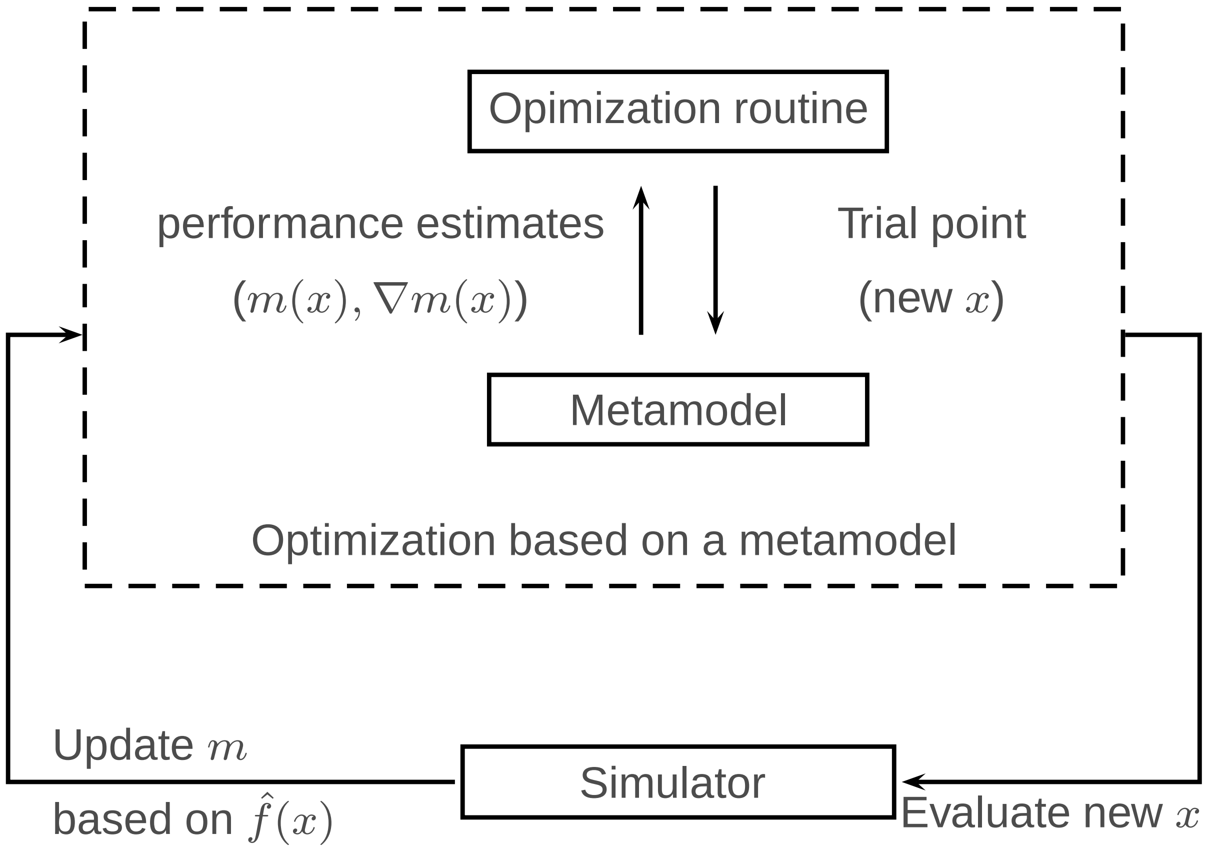
Simulation-based Optimization

- Metamodel methods
  - Simulation models are expensive to evaluate
  - Use simpler approximations within the optimization framework.

The stochastic response of the simulation is replaced by a deterministic metamodel response function, then deterministic optimization techniques are used.

  - Compute the gradient of a surrogate model (or metamodel), as opposed to using the gradient of the simulation response

Simulation-based Optimization



Adapted from Alexandrov et al., 1999

Metamodel Methods

- Metamodels are often a linear combination of basis functions from parametric families
- General purpose metamodels include:
  - Linear or quadratic polynomials
  - Spline models: continuous piecewise polynomials
  - Radial basis functions: sum of radially symmetric functions centered at different points in the search space (*Oeuvray and Bierlaire, 2009*)
- Instead of using a general purpose metamodel we use one that captures the network structure and the interactions between the main network components

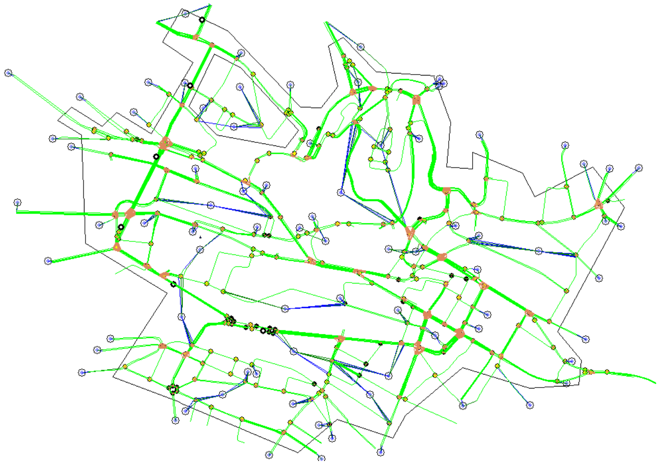
Network models

- We present a metamodel that combines:
  1. Calibrated microscopic traffic simulation model of the Lausanne city center

     Dumont and Bert, 2006
  2. Analytical queueing network model

     Osorio and Bierlaire, 2009

Network model 1



- Calibrated microscopic traffic simulation model : SIMLO
- Developed in LAVOC, EPFL (*Dumont and Bert, 2006*)
- Lausanne city center

Network model 2

- Analytical queueing model
- Finite capacity queueing theory
- Captures the key elements of the underlying network structure, e.g. how upstream and downstream queues interact, and how this interaction is linked to network congestion
- *Blocking* : describes how and where congestion arises, propagates and how it impacts the networks performance
- Novel state space formulation: explicitly models blocking events
- System of nonlinear equations

Network model 2

$$
\begin{cases}
\lambda_i^{\text{eff}} &= \gamma_i(1 - P(N_i = K_i)) + \sum_j p_{ji}\lambda_j^{\text{eff}} \\
\lambda_i &= \lambda_i^{\text{eff}}/(1 - P(N_i = K_i)) \\
\frac{1}{\tilde{\mu}_i} &= \sum_{j \in \mathcal{I}+} \lambda_j^{\text{eff}}/(\lambda_i^{\text{eff}}\hat{\mu}_j) \\
\frac{1}{\hat{\mu}_i} &= \frac{1}{\mu_i} + P_i^f/\tilde{\mu}_i \\
P_i^f &= \sum_j p_{ij} P(N_j = k_j) \\
P(N_i = k_i) &= \frac{1-\rho_i}{1-\rho_i^{k_i+1}}\rho^{k_i} \\
\rho_i &= \frac{\lambda_i}{\hat{\mu}_i}.
\end{cases}
$$

- Endogenous parameters describe congestion in terms of its:
  - sources (conditional transition probabilities)
  - frequency (blocking probabilities)
  - how it propagates/dissipates (unblocking rates)
  - its impact (expected blocked vehicles)
- For more details and a case study see *Osorio and Bierlaire (2009)*
- It has been successfully used in past work to solve a fixed-time traffic signal control problem (*Osorio and Bierlaire, 2008*)

Metamodel

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

- Combination of:
  - T: the analytical queueing model
  - ϕ: a quadratic polynomial

x : decision vector

y : endogenous queueing model variables, such as stationary queue length distributions, congestion indicators

q : exogenous queueing model paramaters, such as network topology, total demand

α, β : metamodel parameters

Metamodel

- Polynomial ϕ:
  - Quadratic with diagonal second derivative matrix
  - This choice is based on existing numerical experiments which show that they are often more efficient than full quadratic models for derivative-free trust region methods (*Powell, 2003*)

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

Metamodel

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

- 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

- Least squares problem solved using the Matlab routine *lsqlin*

Metamodel

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

- Weights:
  - Capture the importance of each point with regards to the current iterate
  - Atkeson (1997) surveys weight functions and analyzes their theoretical properties
  - We use the *inverse distance* weight function, along with the Euclidean distance
  - The weight of a given point is therefore inversely proportional to its distance from the current iterate

$$w_{ki} = \frac{1}{1 + \|x^k - x^i\|_2}$$

Optimization framework

How can we integrate this metamodel within an optimization framework?

- Since most existing metamodel approaches assume a quadratic model
- we resort to **multi-model optimization frameworks**
- also called hybrid methods
- that allow for an arbitrary metamodel

Optimization framework

- **Common motivation**: to combine the use of models with varying evaluation costs
- **Common approach**: Use the low-fidelity (coarse) models, which are faster to evaluate, as the main tool for optimization

- Conn (2009) framework chosen for 3 reasons:
  1. Derivative-free
  2. Trust region method
  3. Makes no assumption on how these metamodels are derived (interpolation or regression)

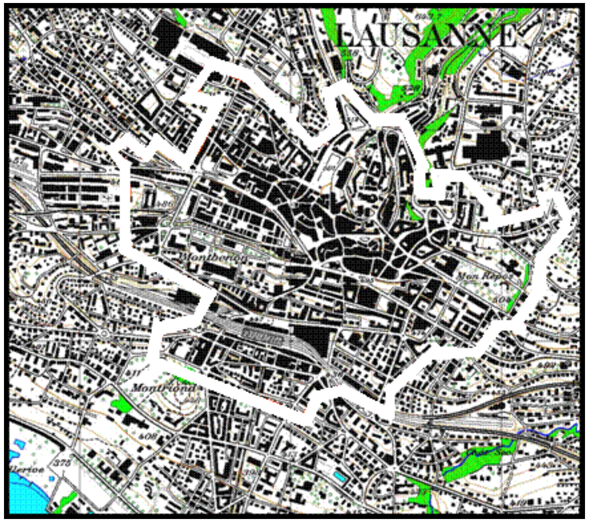
We will integrate our metamodel within the Conn (2009) framework

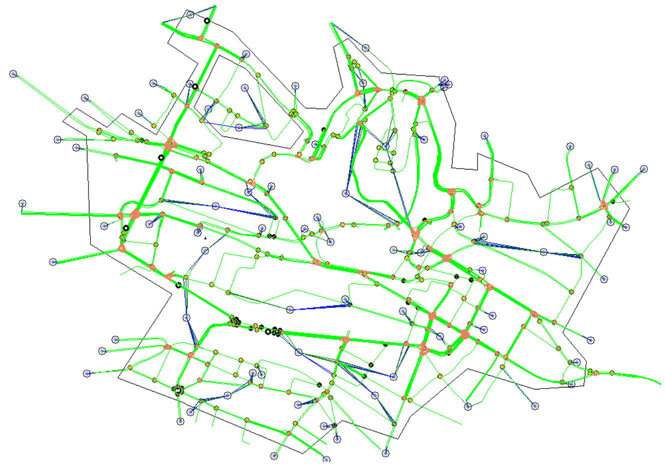
DF TR algorithm

- Builds upon the *Basic trust-region algorithm* (*Conn et. al, 2000*)
- For a given iteration k the algorithm considers a metamodel m_k, an iterate x_k and a TR radius Δ_k.
  Each iteration consists of 5 main steps.

- **Criticality step.** This step may modify m_k and Δ_k if the measure of stationarity is close to zero.
- **Step calculation.** Approximately solve the TR subproblem to yield a trial point
- **Acceptance of the trial point.** The actual reduction of the objective function is compared to the reduction predicted by the model, this determines whether the trial point is accepted or rejected
- **Model improvement.** Either certify that m_k is *fully linear* in the TR or carry out improvement steps
- **TR radius update.**

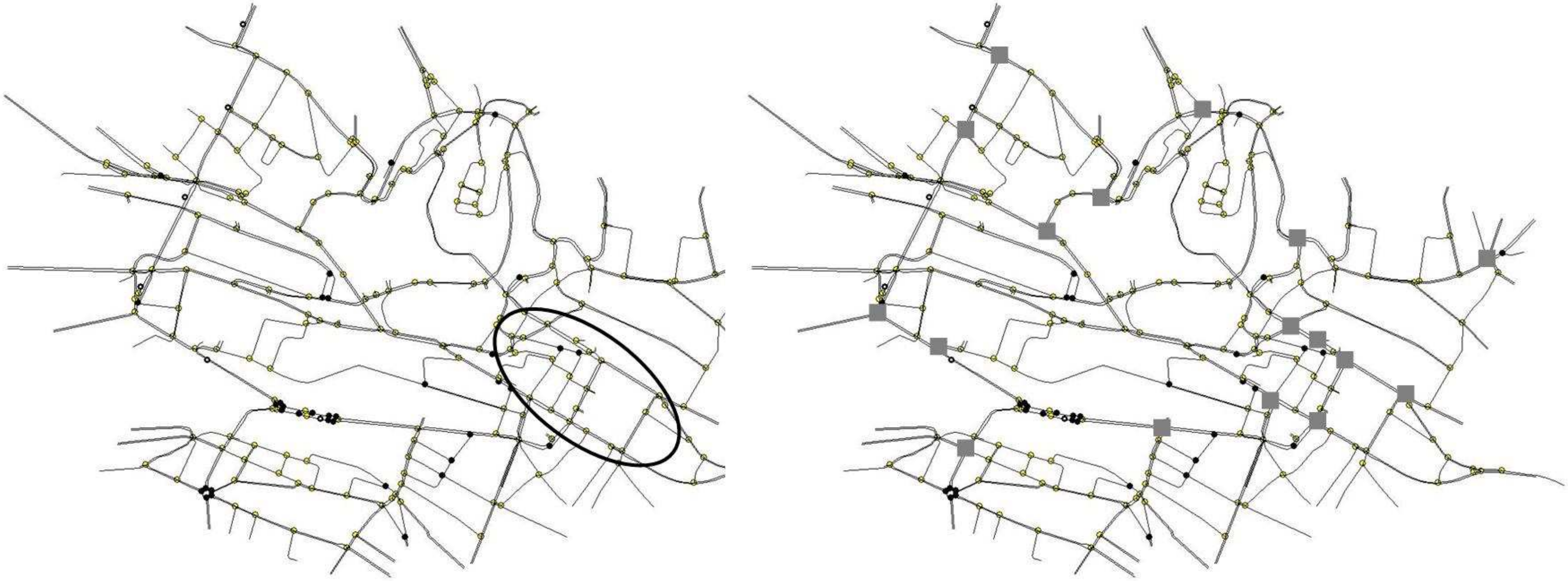
Optimization Problem





- Traffic signal optimization
  - Fixed-time signal control problem
  - Offsets, cycle time, all-red durations and stage structure are given
  - Determine green times for each phase

Optimization Problem



- Comparison versus existing methods:
  - The proposed plans delay the propagation of congestion
  - Importance of grasping the between-queue interactions
- More details: *Osorio and Bierlaire (2008)*.

Optimization Problem

$$\min_{x,z} E[f(x,z;p)]$$

subject to:

$$\sum_{j \in \mathcal{P}_{\mathcal{I}}(i)} x(j) = b_i, \; \forall i \in \mathcal{I}$$

$$x \geq x_L$$

- $x(j)$: green ratio of phase j (green time of phase j divided by the cycle time of its corresponding intersection)
- f: simulated travel time
- z: endogenous simulation variables
- p: exogenous simulation parameters

TR subproblem

At a given iteration k the TR subproblem is formulated as follows:

$$\min_{x,y} \; m_k(x, y; q, \alpha_k, \beta_k) \tag{1}$$

subject to

$$\sum_{j \in \mathcal{P}_{\mathcal{I}}(i)} x(j) = b_i, \; \forall i \in \mathcal{I} \tag{2}$$

$$\ell(x, y; q) = 0 \tag{3}$$

$$\|x - x_k\|_2 \leq \Delta_k \tag{4}$$

$$y \geq 0 \tag{5}$$

$$x \geq x_L \tag{6}$$

- Includes 2 more constraints than the previous problem:
  1. The TR constraint, which uses the Euclidean norm
  2. The system of nonlinear equations that define the queueing network model
- See implementation notes in *Osorio (2010)* for more details

Empirical analysis

- Lausanne subnetwork with:
  1. Simplified demand distribution
  2. Evening peak hour demand

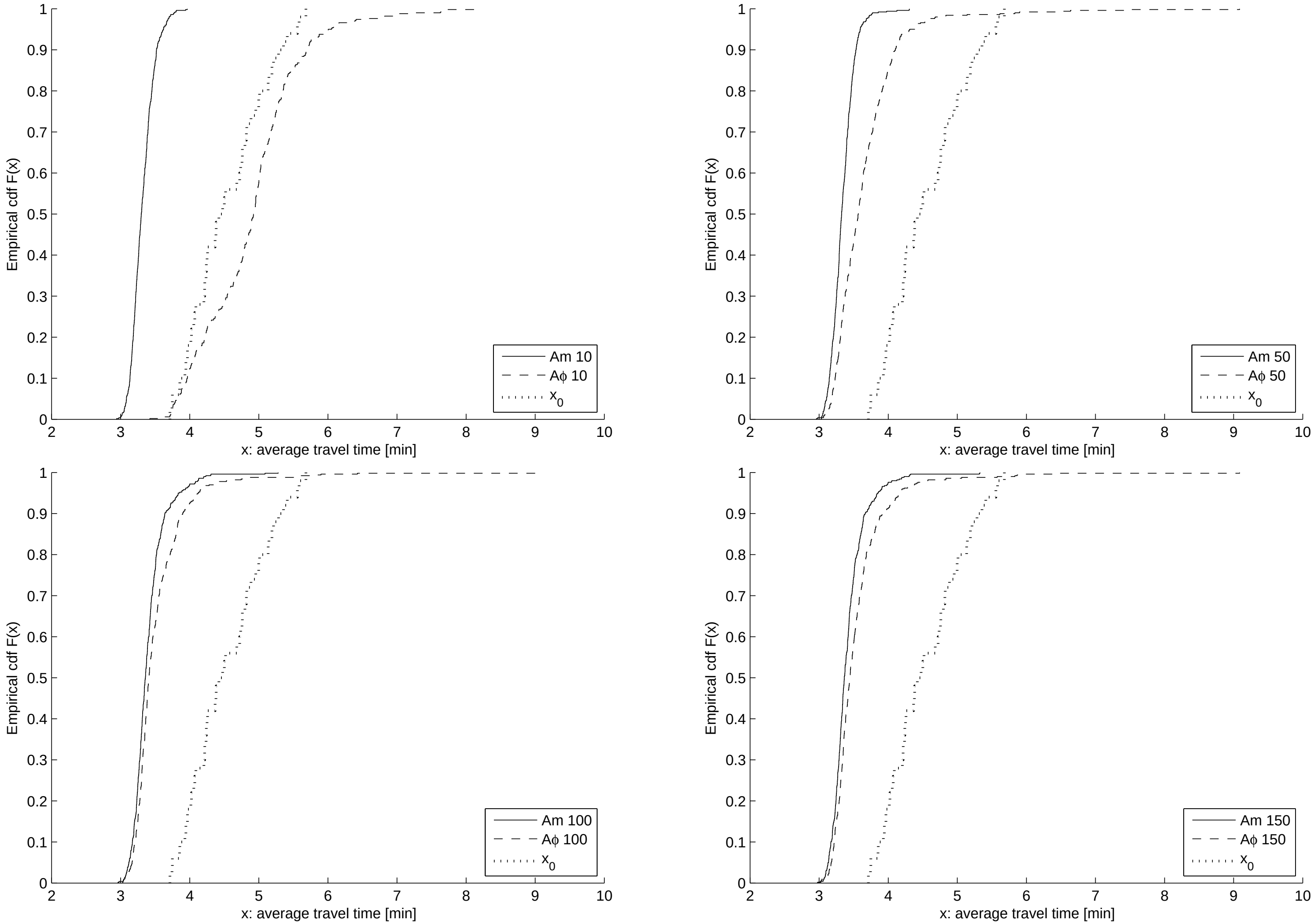
Empirical analysis

- Tight computational budget:
  - No initial observations available
  - 150 simulation evaluations
- Uniformly drawn initial signal plan
- We run the corresponding algorithm 10 times, deriving 10 signal plans
- Simulation model is used to evaluate the effect of the signal plans upon the entire Lausanne network.
- For each signal plan:
  - 50 replications
  - Empirical cumulative distribution function of the average travel times

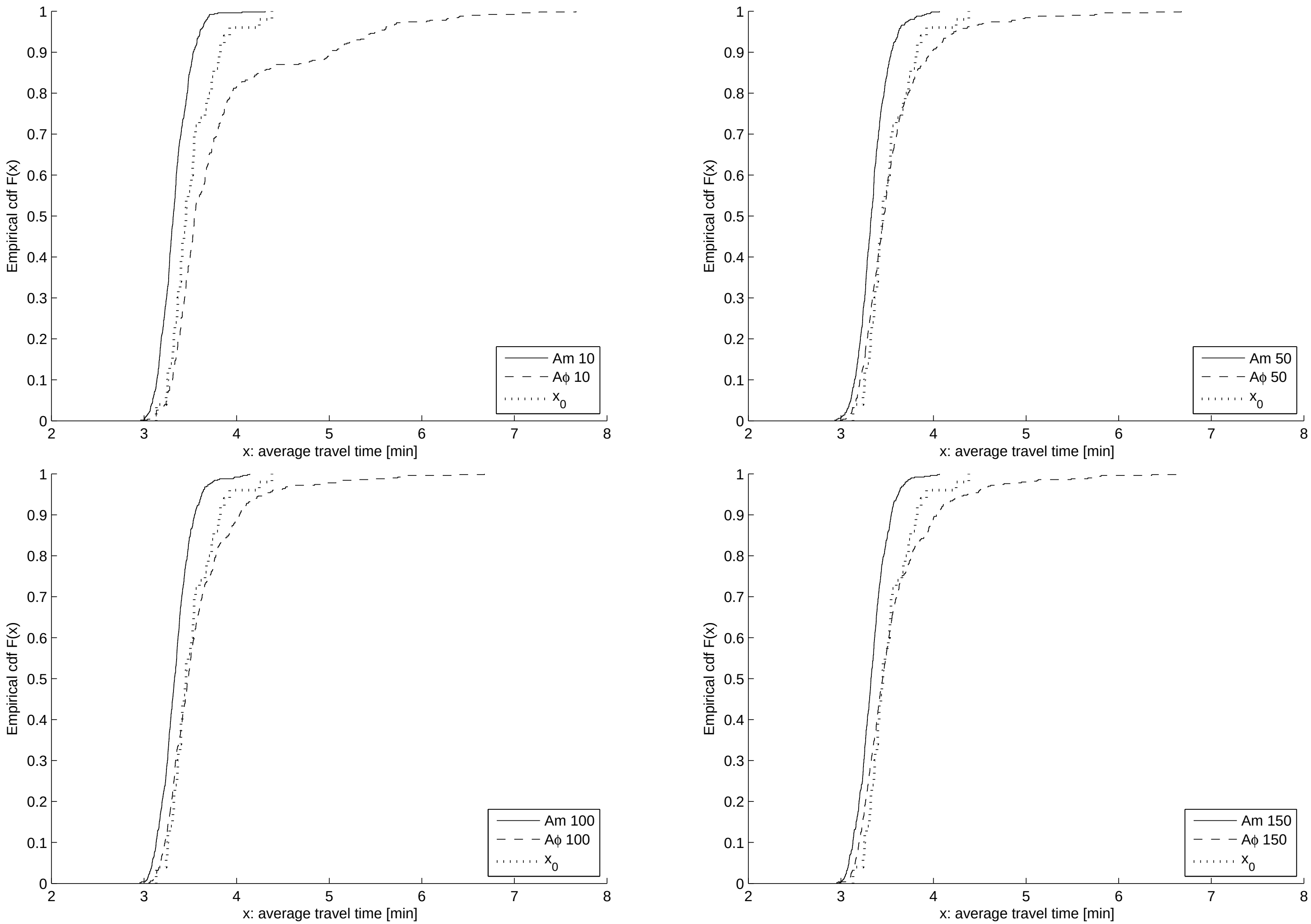
Empirical analysis 1

- Lausanne subnetwork with simplified demand distribution
- Control two adjacent signalized intersections (13 phases)

Empirical analysis 1

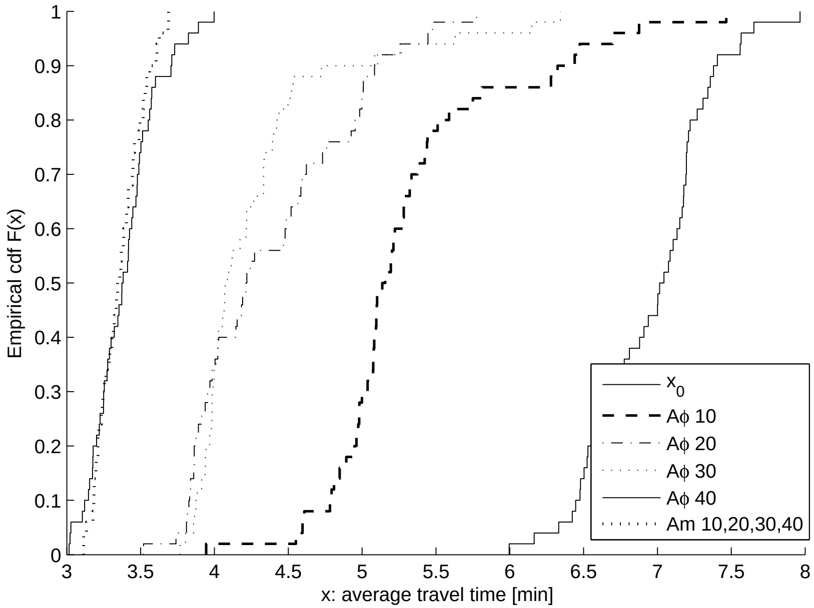


Empirical analysis 1



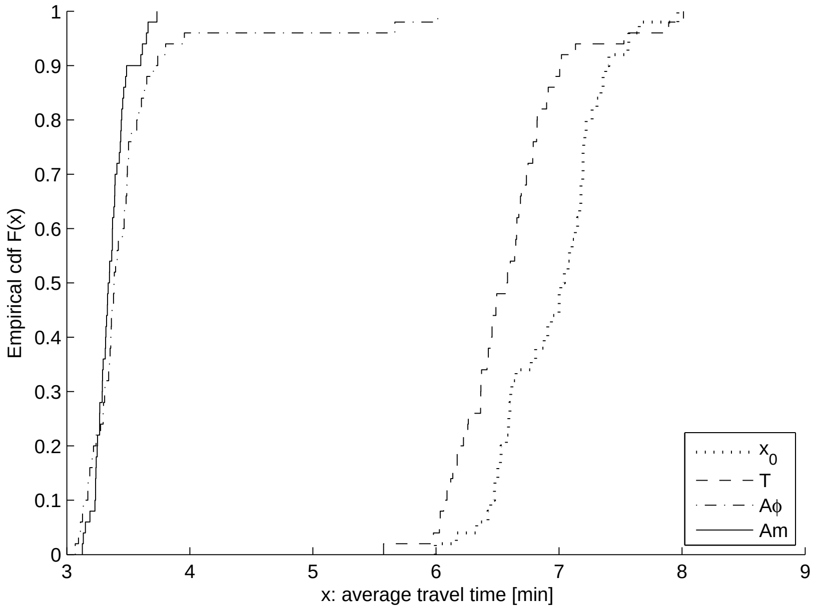
Empirical cdf’s of the average travel times

Empirical analysis 1



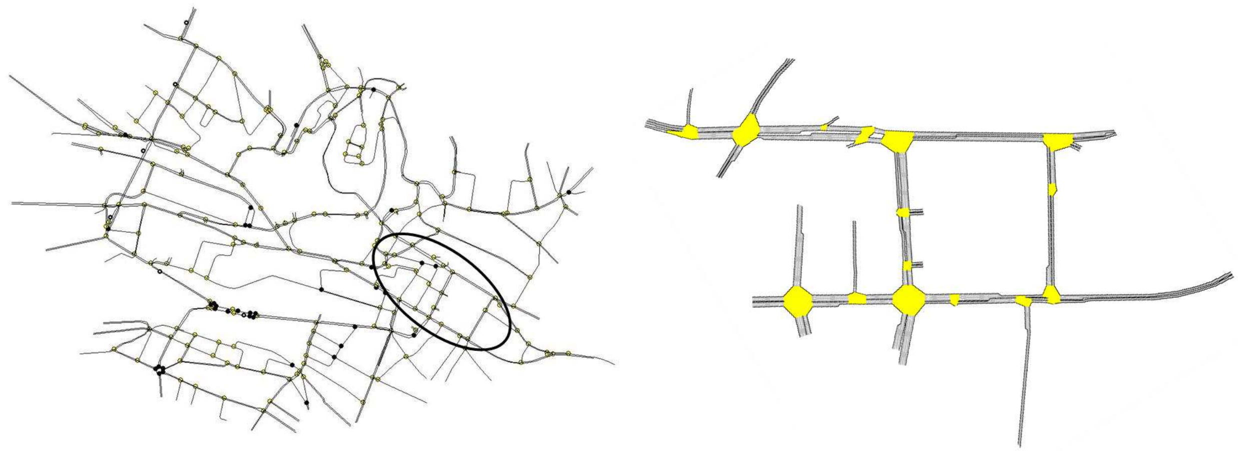
- With a tight computational budget, the proposed method is able to identify signal plans that improve the distribution of the average travel time.
- Outperforms the general-purpose metamodel for tight computational budgets (main motivation to resort to DF algorithms)

Empirical analysis 1



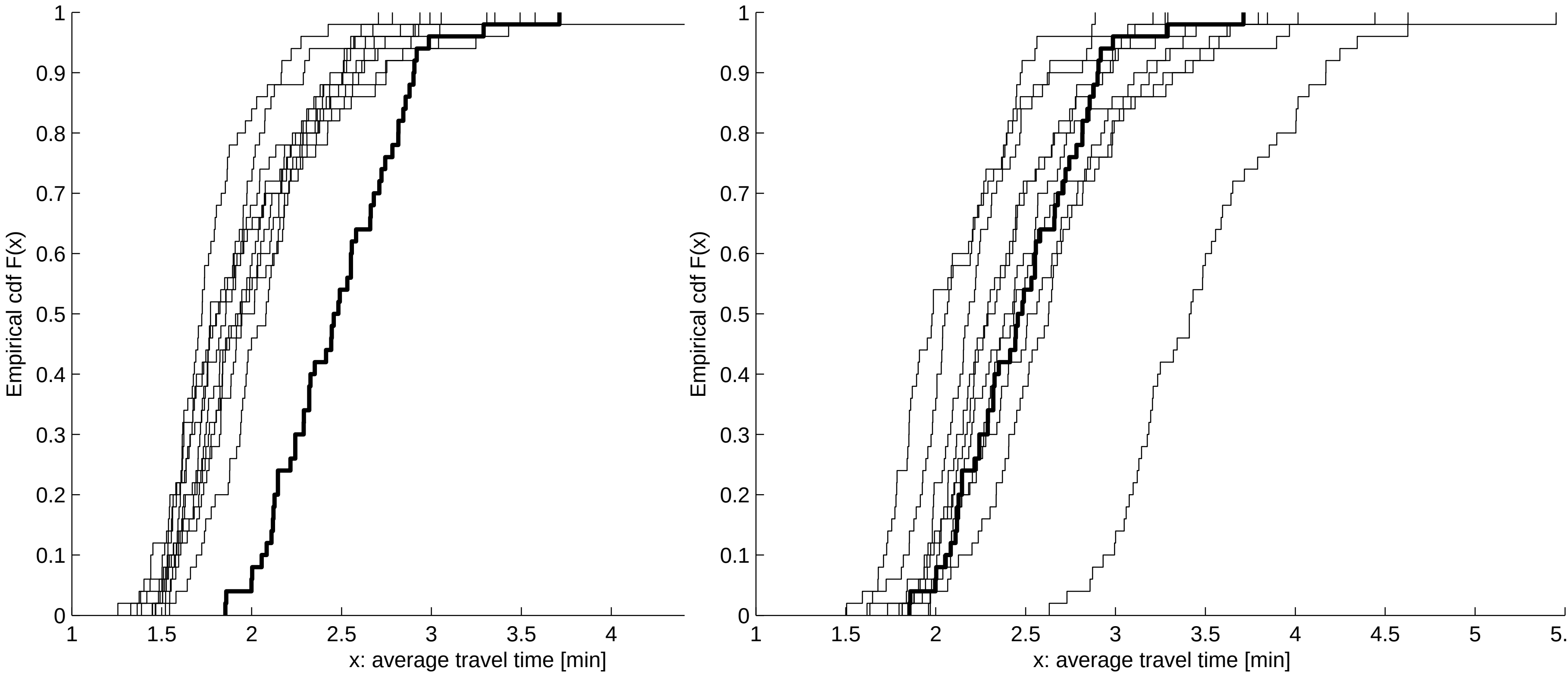
- Simulation budget: 750 runs

Empirical analysis 2



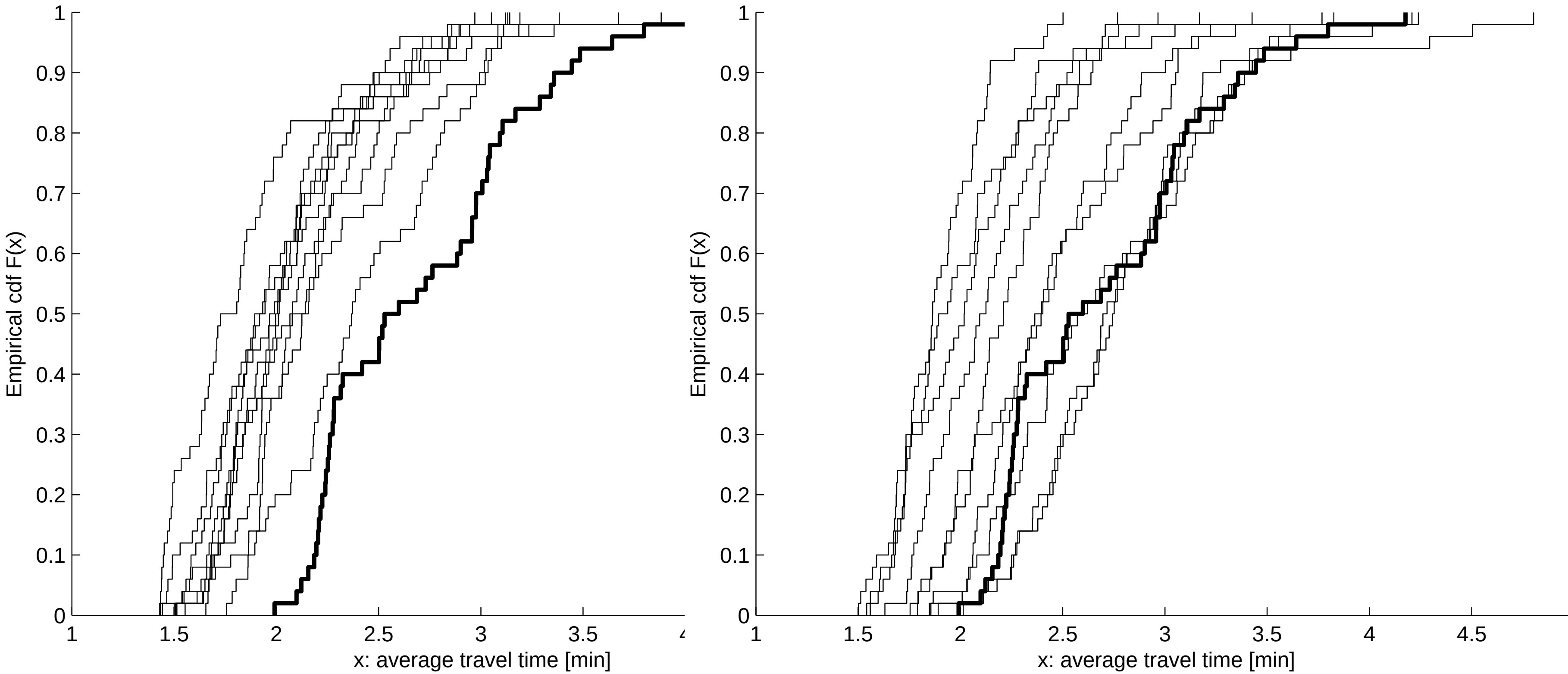
- Subnetwork of the Lausanne city center.
- Demand for the evening peak hour (17h-18h)
- Subnetwork: 48 roads, 15 intersections (9 of which are signalized).
- 51 phases are variable
- cycle times: 90 or 100 seconds.
- minimal green times: 4 seconds.

Empirical analysis 2

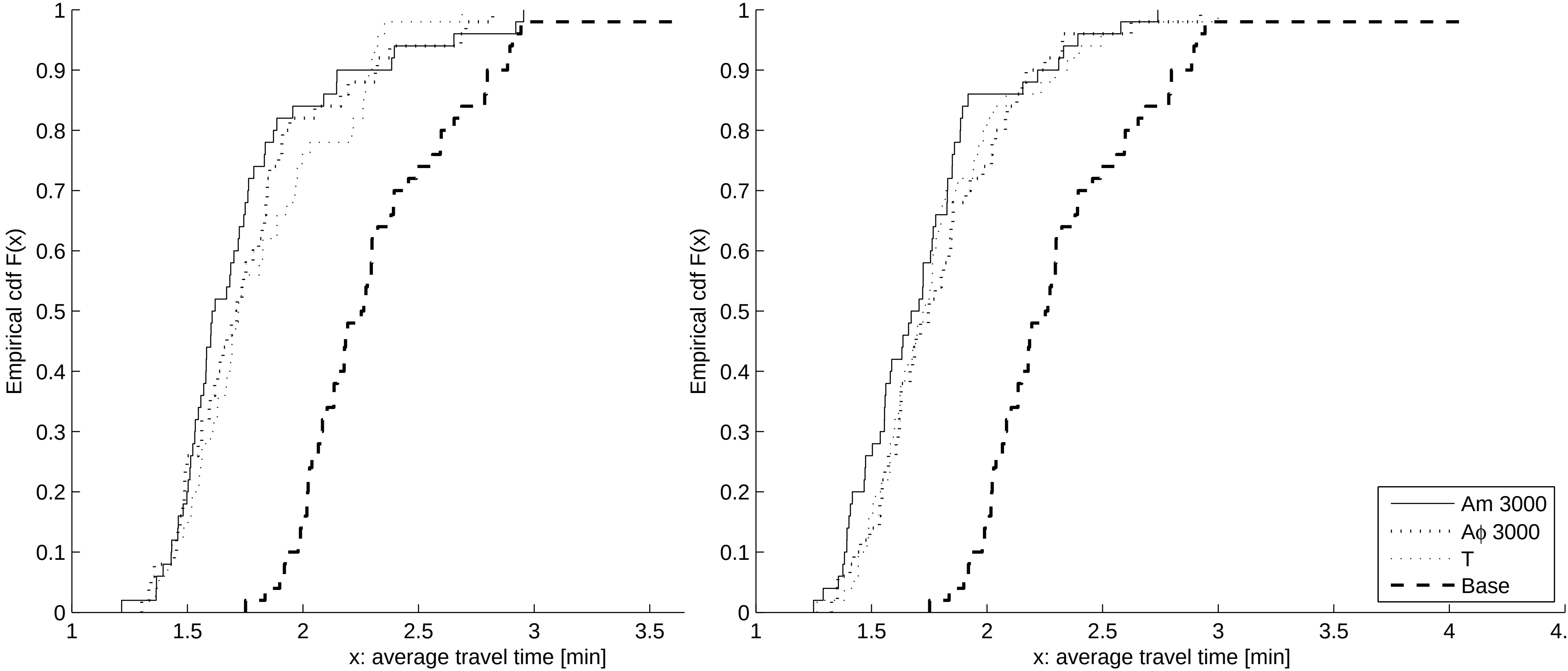


- All 10 signal plans derived by the proposed method yield improved average travel times compared to the initial plan.

Empirical analysis 2



Empirical analysis 2



- Simulation budget: 3000 runs

Conclusions

- Framework for simulation-based optimization of congested networks
  - Metamodel that combines information from a simulator and an analytical network model
  - Derivative-free trust region algorithm
- Empirical analysis
  - Traffic signal control problem
  - Tight computational budget
  - Provides meaningful trial points since the first iterations
  - Outperforms the general-purpose metamodel for tight computational budgets (main motivation to resort to DF algorithms)
- Current work
  - Investigating other metamodel formulations
  - Dynamic analytical model: upcoming ISTTT conference

References

- **Alexandrov, Lewis, Gumbert, Green and Newman (1999).** Optimization with variable-fidelity models applied to wing design, *Technical Report CR-1999-209826, NASA Langley Research Center*
- **Alexandrov, Lewis, Gumbert, Green and Newman (2001).** Approximation and Model Management in Aerodynamic Optimization with Variable-Fidelity Models. *Journal of Aircraft* 38(6):1093-1101
- **Bandler, Cheng, Dakroury, Mohamed, Bakr, Madsen and Sondergaard (2004).** Space mapping: the state of the art. *IEEE Transactions on Microwave Theory and Techniques* 52(1):337-361
- **Booker, Dennis, Frank, Serafini, Torczon and Trosset (1999).** A rigorous framework for optimization of expensive functions by surrogates *Structural and Multidisciplinary Optimization* 17(1):1-13
- **Carter (1986).** *Multi-model algorithms for optimization*. PhD Thesis. Rice University, Houston, TX, USA.
- **Conn, Scheinberg and Vicente (2009).** Global convergence of general derivative-free trust-region algorithms to first- and second-order critical points *SIAM Journal on Optimization* 20(1):387âĂŞ415.
- **Dumont and Bert (2006).** Simulation de l'agglomération Lausannoise SIMLO. Technical Report. LAVOC, ENAC, EPFL.
- **Osorio and Bierlaire (2008).** Network performance optimization using a queueing model. *Proceedings of the European Transport Conference (ETC)*, The Netherlands.
- **Osorio and Bierlaire (2009).** An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. *European Journal Of Operational Research* 196(3):996-1007
- **Osorio (2010).** *Mitigating network congestion: analytical models, optimization methods and their applications* Ph.D. thesis, Ecole Polytechnique Fédérale de Lausanne.