Modeling opportunities from modern pedestrian data

Michel Bierlaire Marija Nikolić Flurin Hänseler Antonin Danalet
Riccardo Scarinci

Transport and Mobility Laboratory TRANSP-OR
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

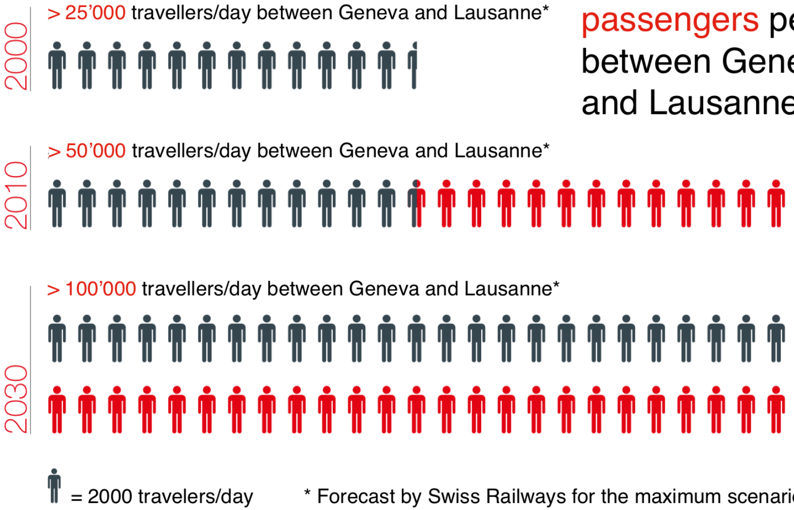
November 18, 2014

Outline

1. Data, index characterization and fundamental diagram
2. Estimation of pedestrian OD flows in railway stations
3. Activity choice in pedestrian facilities
4. Conclusions

Swiss context

By 2030, 100’000 passengers per day between Geneva and Lausanne



* Forecast by Swiss Railways for the maximum scenario

Context and Motivation

- Pedestrian movement analysis in transportation hubs
  - large increase in number of passengers
  - congestion of pedestrian facilities at peak hours
- Pedestrian indexes
  - performance: travel time, timetable stability, level of service
  - comfort: "degree of crowdedness"
  - safety: in case of evacuation
- Models needed for better understanding and prediction of pedestrian flows
  - optimize pedestrian facilities and their operation

Overview


Data
From data to indexes
Activity based models
Modeling of fundamental relationships
Assignment models
Origin-destination estimation

Outline

1. Data, index characterization and fundamental diagram
2. Estimation of pedestrian OD flows in railway stations
3. Activity choice in pedestrian facilities
4. Conclusions

Traditional pedestrian data collection

Pedestrian counting

- Real life data
- Infrared beams and switching mats
- Video surveillance
  - Manual extraction of relevant data

Pedestrian tracking: Video-based technology

- Experimental data
- Controlled environment
- Video analysis

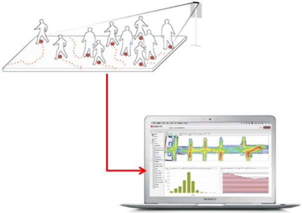
Modern pedestrian data collection

Pedestrian tracking: Pervasive technology

- Bluetooth/WLAN traces
- People equipped with signal emitting devices (e.g. smartphones)
- Intrusive or non-intrusive methods

Visiosafe - new technology

- Spin-off of EPFL
- Anonymous tracking of pedestrians
- Large-scale data collection running on a continuous basis
- Thermal and range sensors



Visiosafe

Visiosafe video

Visiosafe data

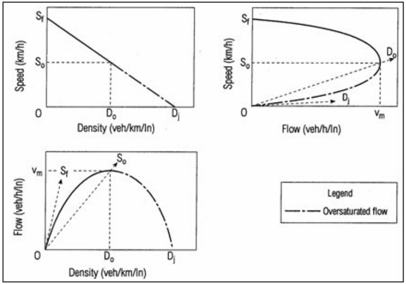
- Detailed pedestrian trajectories
- Position of every single individual over time

$$(t, x(t), y(t), pedestrian_{id})$$

Flow indicators

Traffic flow theory

- Flow
- Density
- Speed
- Fundamental relationships



source: HCM

Pedestrian traffic

Density

- Number of pedestrians per square meter at a given moment

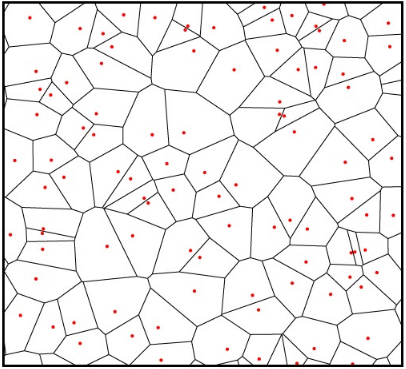
Issues

- Spatial discretization is arbitrary
- Results may be highly sensitive
- Idea: data driven spatial discretization

Density: data-driven discretization

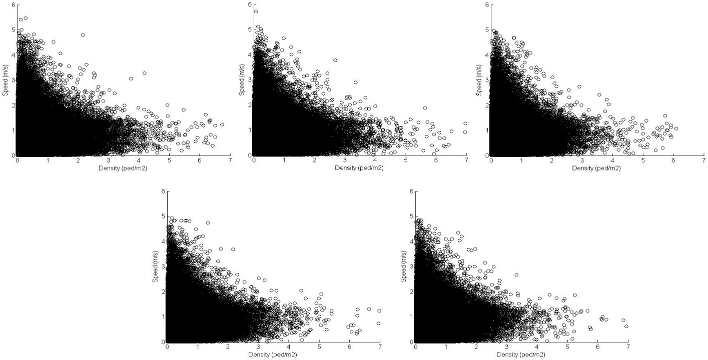
Voronoi tessellations

- $p_1, p_2, ..., p_N$ is a finite set of points
- Voronoi space decomposition assigns a region to each point p_i $V(p_i) = \{p | \|p - p_i\| \leq \|p - p_j\|, i \neq j\}$



Empirical speed-density relationship

Speed-density profiles



February, 2013.: morning peak hour

Probabilistic speed-density model

Theoretical foundation

- Speed is affected by different factors
  - congestion level, trip purpose, age, health condition, etc.
- Congestion level: speed decreases with increasing density
- Pedestrian heterogeneity
  - Slower walkers: elderly people, people unfamiliar with environment, people influenced by static and dynamic objects from the scene, etc.
  - Faster walkers (less sensitive to congestion): business travelers, people in a hurry to catch a train, etc.
- Characterization of the observed phenomena: probabilistic approach

Probabilistic speed-density relationship

Mixture model specification

$$f(v,k) = f_l(v,k) \cdot P(v_m \geq v) + f_e(v,k) \cdot P(v_m \leq v)$$

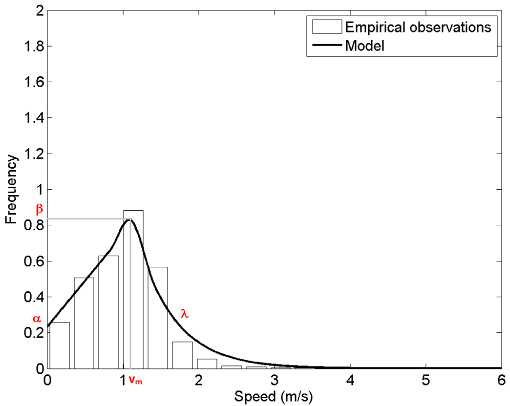
Mixture components

$$f_l(v,k) = \frac{\beta(k)-\alpha(k)}{v_m(k)} \cdot v + \alpha(k))$$

$$f_e(v,k) = exp(-\lambda \cdot v + log(\beta(k)) + \lambda \cdot v_m(k))$$

Probabilistic speed-density relationship

Illustration - one density level



Model parameters

- v_m - mode of the distribution
- Assumed to follow symmetric triangular distribution

$$f_{v_m}(\bar{v}_m(k), \sigma) = \begin{cases} \frac{v_m(k) - \bar{v}_m(k) + \sigma}{\sigma^2}, & \bar{v}_m(k) - \sigma \leq v_m \leq \bar{v}_m(k) \\ \frac{\bar{v}_m(k) + \sigma - v_m(k)}{\sigma^2}, & \bar{v}_m(k) < v_m \leq \bar{v}_m(k) + \sigma \end{cases}$$

- The mean value is assumed to correspond to the Underwood's model

$$\bar{v}_m(k) = v_f \cdot exp(-\frac{k}{\gamma})$$

Model parameters

- α - probability density corresponding to small speed values

$$\alpha(k) = a_\alpha \cdot k + b_\alpha$$

- β - probability density corresponding to the most likely speed values

$$\beta(k) = a_\beta \cdot k + b_\beta$$

- λ -the rate of the exponential mixture component

Model estimation

Maximum likelihood

$$\hat{\theta} = \max_{\theta} \sum_{i=1}^{n} \ln \left(f_l(v_i, k_i; \theta) \cdot \omega_l(v_i; \theta) + f_e(v_i, k_i\theta) \cdot \omega_e(v_i; \theta) \right)$$

Notation

$$\theta = \{a_\alpha, b_\alpha, a_\beta, b_\beta, \lambda, v_f, \gamma, \sigma\}$$

$$\omega_l(v_i; \theta) = 1 - F_{v_m}(v_i; \theta)$$

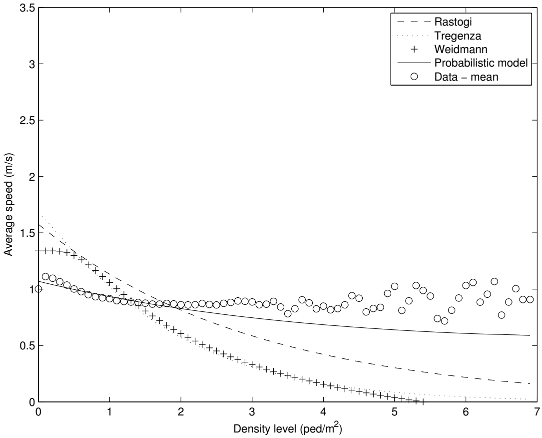
$$\omega_e(v_i; \theta) = F_{v_m}(v_i; \theta)$$

Estimation results

Parameter	Value	Std err
a_α	0.029	$0.028e^{-04}$
b_α	0.232	$0.069e^{-04}$
a_β	0.082	$0.035e^{-04}$
b_β	0.805	$0.105e^{-04}$
λ	1.988	$0.017e^{-04}$
v_f	1.126	$0.078e^{-04}$
γ	4.753	$0.014e^{-04}$
σ	0.291	$0.193e^{-04}$
$log \mathcal{L}$	-506208.864	
$\#parameters$	8	
$\#observations$	756691	

Comparison with deterministic models

Exponential specifications

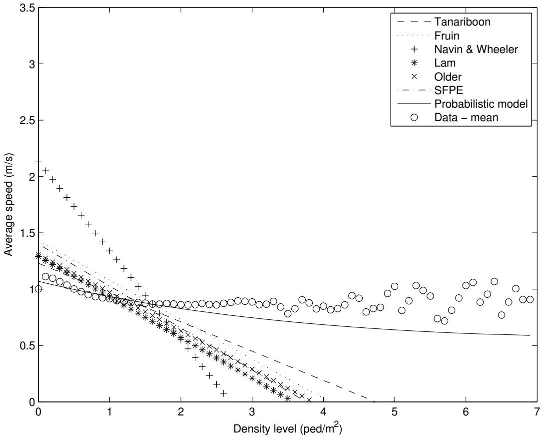


Goodness of Fit

Model	MSE
Tregenza	0.406
Weidmann	0.441
Rastogi	0.221
Probabilistic	0.041

Comparison with deterministic models

Linear specifications

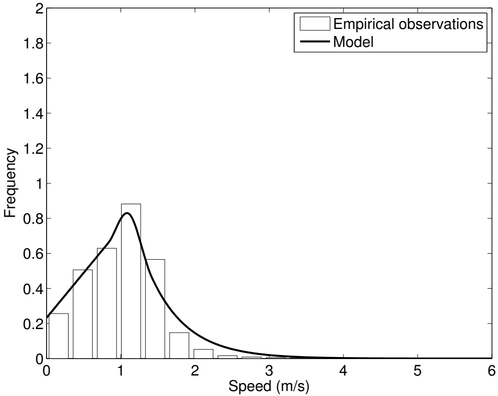


Goodness of Fit

Model	MSE
Tanariboon	0.591
Fruin	0.948
Navin and Wheeler	4.751
Lam	1.244
Older	1.044
SFPE	1.170
Probabilistic	0.041

External validation - PDF

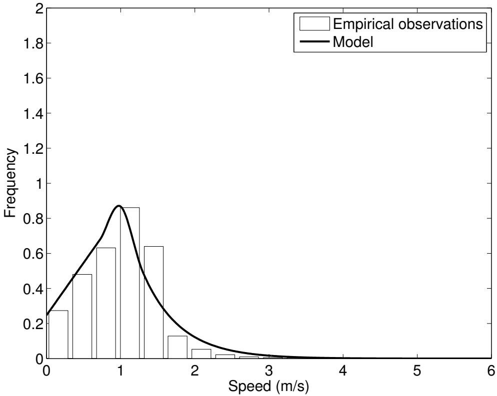
Density level: $< 0.1 ped/m^2$



$\#observations : 21178$

External validation - PDF

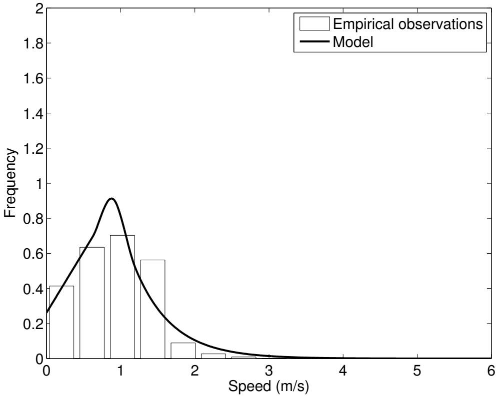
Density level: $0.5 ped/m^2$



$\#observations : 40470$

External validation - PDF

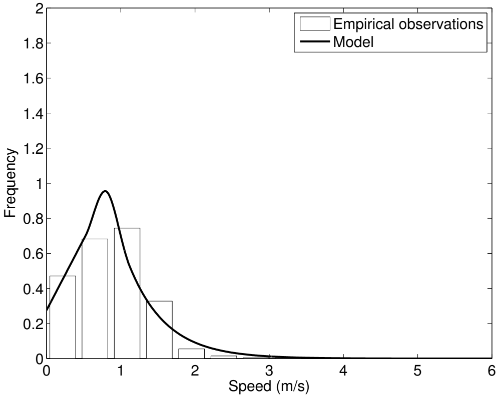
Density level: $1 ped/m^2$



$\#observations : 10705$

External validation - PDF

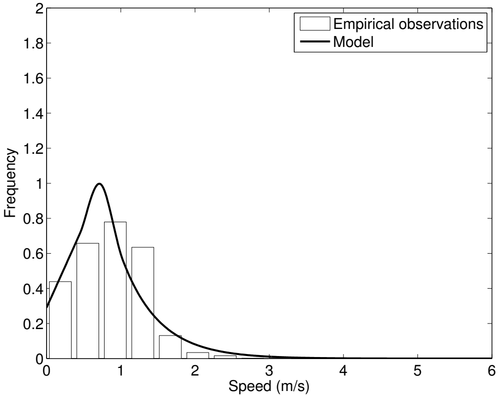
Density level: $1.5 ped/m^2$



$\#observations$: 6781

External validation - PDF

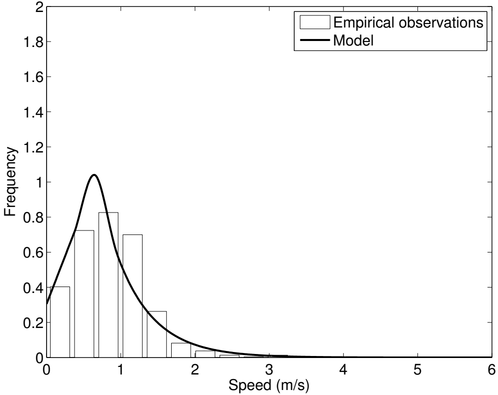
Density level: $2ped/m^2$



$\#observations : 2509$

External validation - PDF

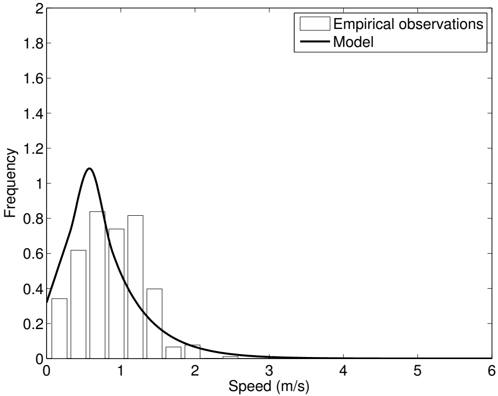
Density level: $2.5 ped/m^2$



$\#observations : 898$

External validation - PDF

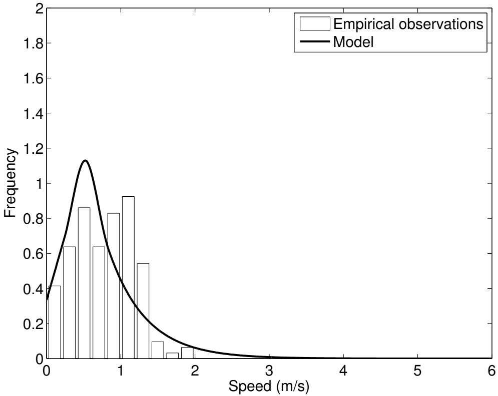
Density level: $3ped/m^2$



$\#observations$: 354

External validation - PDF

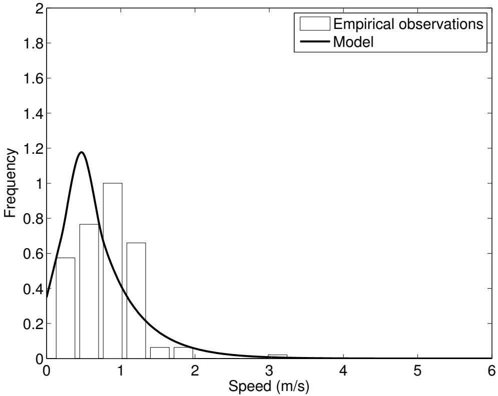
Density level: $3.5 ped/m^2$



$\#observations : 158$

External validation - PDF

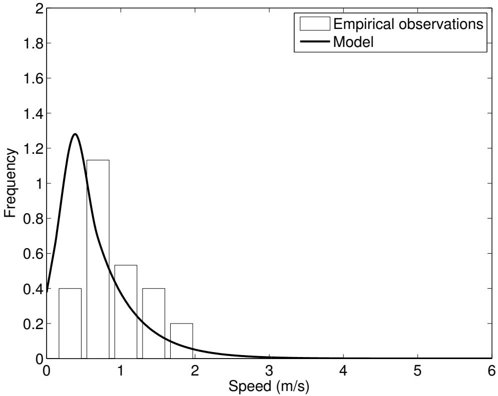
Density level: $4 ped/m^2$



$\#observations : 73$

External validation - PDF

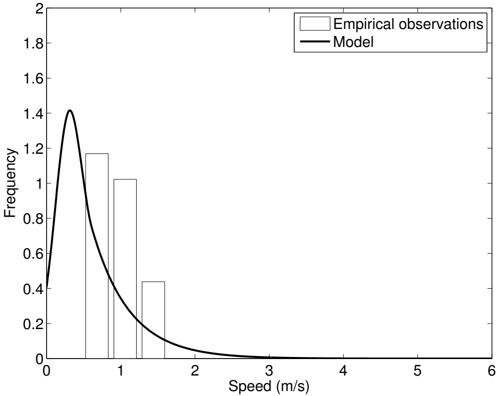
Density level: $5 ped/m^2$



$\#observations : 22$

External validation - PDF

Density level: $6 ped/m^2$



$\#observations : 3$

Other research topics

Pedestrian-oriented flow characterization

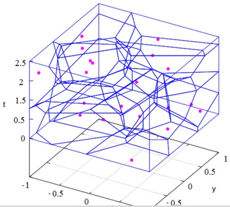
- Definitions of flow characteristics by adapting Edie's definitions

Stream-based approach

- Pedestrian traffic composed of different streams
- A stream definition: exogenous and direction-based
- Trajectories are assumed to contribute to the streams

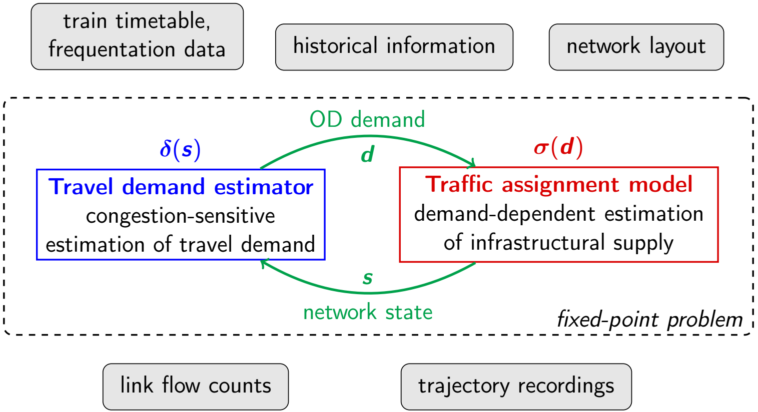
Data-driven discretization framework - 3D Voronoi

- Set of all points in a cell corresponding to a given location is a time interval
- Set of all points in a cell corresponding to a specific time is a physical area



Outline

Pedestrian demand and supply



Travel demand estimator $\delta(s)$

Requirements:

- accurate prediction of dynamic OD demand $\boldsymbol{d}^{\star}$
- explicit integration of train timetable ($\rightarrow$ 'micro-peaking')
- aggregate model (no socio-economic data)

Input:

- a priori OD demand estimate $\hat{\boldsymbol{d}}$
- train timetable, alighting volumes per train $\hat{\boldsymbol{w}}$
- network state $\boldsymbol{s}$ (e.g. $\boldsymbol{s} = \boldsymbol{\sigma}(\hat{\boldsymbol{d}})$)
- link flow measurements $\hat{\boldsymbol{f}}'$

$\delta(s)$: Space topology I

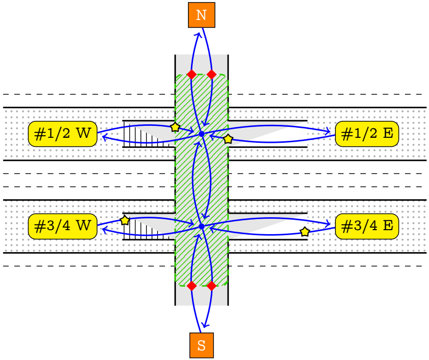


Figure: Sample railway station (legend on next slide)

$\delta(s)$: Space topology II

	pedestrian walking network
entrance	centroid with historical information
platform	centroid without historical information
☆	link with a priori flow estimate based on timetable
◆	link equipped with directed flow counter
	area covered by pedestrian tracking system

$\delta(s)$: Structural model

Link flows:

$$\begin{aligned} \boldsymbol{f} &= \boldsymbol{A}(\boldsymbol{s})\boldsymbol{\Delta}(\boldsymbol{s})\boldsymbol{d} + \boldsymbol{\eta} \\ \boldsymbol{f} &= \boldsymbol{f}_{\text{arr}} + \boldsymbol{f}_0 \\ \boldsymbol{f}_{\text{arr}} &= \varphi_{\text{arr}}(\boldsymbol{w}, \boldsymbol{s}) + \varepsilon_{\text{arr}} \\ \boldsymbol{f}_0 &= \varphi_0(\boldsymbol{d}, \boldsymbol{s}) + \varepsilon_0 \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{s}) &: \text{ link-paths matrix} \\ \boldsymbol{\Delta}(\boldsymbol{s}) &: \text{ route choice matrix} \\ \boldsymbol{f}_{\text{arr}}, \boldsymbol{f}_0 &: \text{ train-induced arrival and `base' flow} \\ \boldsymbol{\eta}, \varepsilon_{\{\text{arr},0\}} &: \text{ error terms} \end{aligned}$$

$\delta(s)$: Measurement model and problem formulation

Measurement model:

$$\hat{\boldsymbol{d}} = \boldsymbol{d} + \boldsymbol{\omega}_d$$
$$\hat{\boldsymbol{f}}' = \boldsymbol{R}_f \boldsymbol{f} + \boldsymbol{\omega}'_f$$

Estimation problem (**base** vs. **full** estimator):

$$\boldsymbol{d}^\star = \arg\min_{\boldsymbol{d}\geq 0} \text{dist}_1 \left\langle \begin{pmatrix} \hat{\boldsymbol{f}}' \\ \boldsymbol{\varphi}'_{\text{arr}} + \boldsymbol{\varphi}'_0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{R}_f \\ \boldsymbol{R}_\varphi \end{pmatrix} \boldsymbol{f} \right\rangle + \text{dist}_2 \langle \hat{\boldsymbol{d}}, \boldsymbol{d} \rangle$$

where

$\boldsymbol{R}_{f,\varphi}$: reduction matrices

$\boldsymbol{\omega}_d, \boldsymbol{\omega}'_f$: error terms

$dist\langle\cdot\rangle_{\{1,2\}}$: weighted distance measures

$\delta(s)$: Specification & Case study

Model specification:

- demand-invariant supply $\boldsymbol{\sigma} \neq f(\boldsymbol{d})$
  - idea: $v \sim \mathcal{N}(1.34 \text{ m/s}, 0.34 \text{ m/s})$ (Weidmann, 1993)
- route choice: shortest path
  - unique path, all-or-nothing assignment
- empirical model for train-induced arrival flow
  - calibrated on Lausanne data (next slide)

Case study:

- Lausanne railway station, Switzerland
- morning peak period (07:30 – 08:00)

$\delta(s)$: Model for train-induced arrival flows

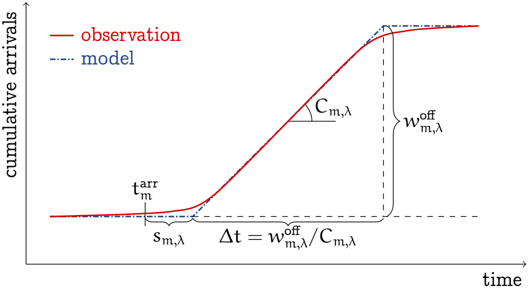


Figure: Model for train-induced arrival flows (m: train, $w^{\text{off}}_{m,\lambda}$: alighting volume, λ: link, t^{arr}_{m}: arrival time, $s_{m,\lambda}$: dead time, $C_{m,\lambda}$: flow capacity)

$\delta(s)$: Illustration of train-induced flow model

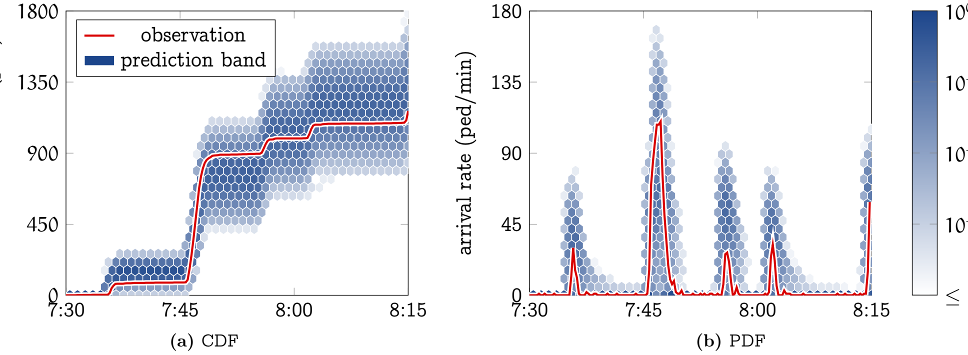
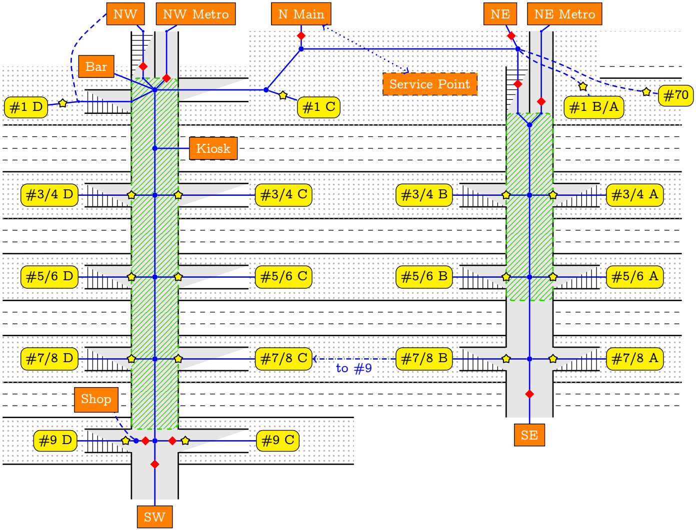


Figure: Simulation of train-induced arrival flow (Lausanne, platform #5/6, Apr 10, 2013)

$\delta(s)$: Aerial view of Lausanne railway station

$\delta(s)$: Layout of Lausanne railway station



$\delta(s)$: Temporal evolution of demand

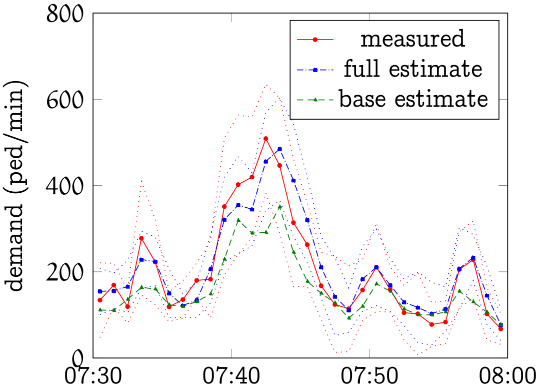
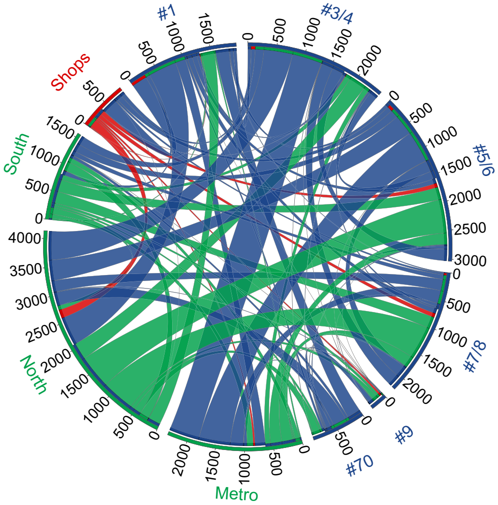
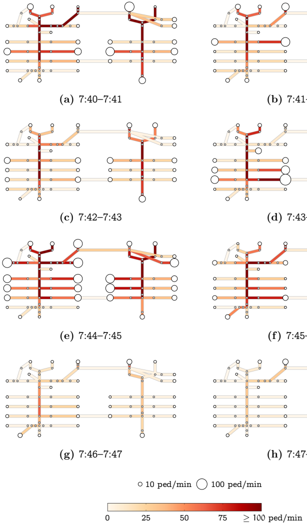


Figure: Demand in pedestrian underpasses, 10-day reference set, 2013

$\delta(s)$: Circos diagram of OD demand in Lausanne



$\delta(s)$: Flow map of Lausanne railway station



Network supply model $\sigma(d)$

Requirements:

- accurate prediction of travel time and density
- low computational cost, 'easy' calibration
- aggregate model (input and output at aggregate level)

Input:

- 'pedestrian groups': route, departure time interval, size
- network topology

$\sigma(d)$: Hughes' theory for pedestrian flow (2002)

$$\frac{\partial k_i(\boldsymbol{x},t)}{\partial t} + \nabla \left(k_i(\boldsymbol{x},t) \boldsymbol{v}_i(\boldsymbol{x},t) \right) = 0$$

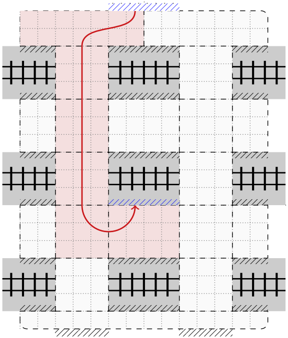
$$\boldsymbol{v}_i(\boldsymbol{x},t) = -F\left(k(\boldsymbol{x},t)\right) v_f \frac{\nabla \phi_i(\boldsymbol{x},t)}{\|\nabla \phi_i(\boldsymbol{x},t)\|}$$

$$F_i(0) \leq 1, \quad F_i(k_c) = 0, \quad \frac{dF_i}{dk} \leq 0$$

$$\|\nabla \phi_i(\boldsymbol{x},t)\| = \frac{1}{F(k(\boldsymbol{x},t))}$$

- $\boldsymbol{x}$: space vector, t: time, i: pedestrian class (defined by route)
- k_i: density, $k = \sum k_i$, k_c: critical density
- $F_i(k)$: fundamental diagram, $\boldsymbol{v}_i$: velocity, v_f: free-flow speed
- $\phi_i(\boldsymbol{x},t)$: scalar potential field

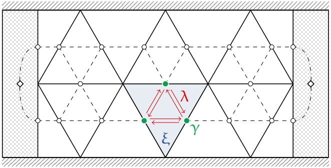
$\sigma(d)$: Isotropic approximation



- heuristic approach inspired by
  - Daganzo (1994)
  - Asano et al. (2007)
- cell-based space representation
  - scalar density k
  - scalar velocity $F(k)$
  - route-specific potential
- case studies
  - Lausanne railway station
  - Dutch bottleneck
- details: Hänseler et al. (2014)

$\sigma(d)$: Anisotropic approximation

- cell- *and* link-based space representation
  - pedestrian groups travel on links
  - areas represent range of interaction of links
- stream-based fundamental diagram
  - subsets of links form streams
  - each stream has different speed
  - example of stream-based FD: SC Wong et al. (2010)
- ongoing collaboration with WHK Lam, PolyU Hong Kong



- links: pedestrian network
- cells: area of interaction
- nodes: potentials for route choice

Outline

Activities in pedestrian infrastructure

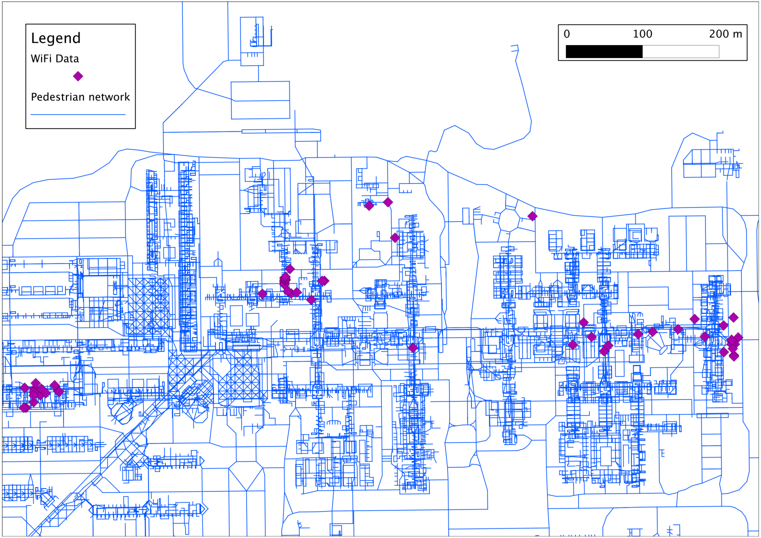
Activity modeling in pedestrian infrastructure (vs in a city)

- Small scale of activity episodes
  - Spatially: different activity types can be performed close to each other
  - Temporally: activity episode duration in a train station is $\sim$ 5'
- Not home-based nor tour-based
  - No obvious or natural priorities of activity types (home, work)
  - Tours: a way to decompose time in manageable units with duration

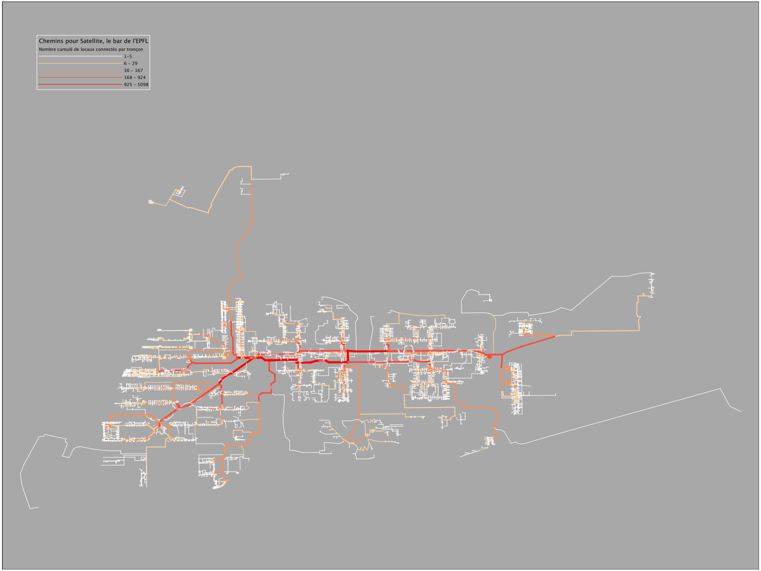
Available data for pedestrian activity modeling

- WiFi traces
  - Data from existing access points
  - Localization, timestamp, estimation of the precision
- Map
  - Localization of points of interest
  - Pedestrian network (allowing for shortest path computation)
- Schedules
  - Class schedule on campuses, train schedule in stations, concert schedule in music festivals
  - Allowing for schedule delay computation
- Attractivity
  - Model of aggregated occupation per point of interest
  - Data sources: checkouts in supermarkets, metro card swapping data, concert tickets data, number of seats in a restaurant, number of employees per office, number of students in class, ...

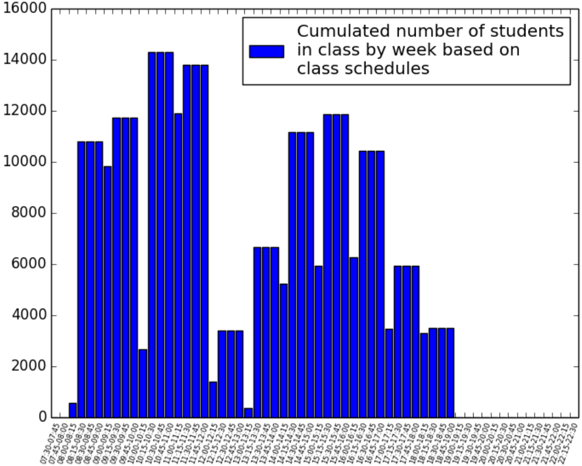
WiFi traces: No stop, no semantics



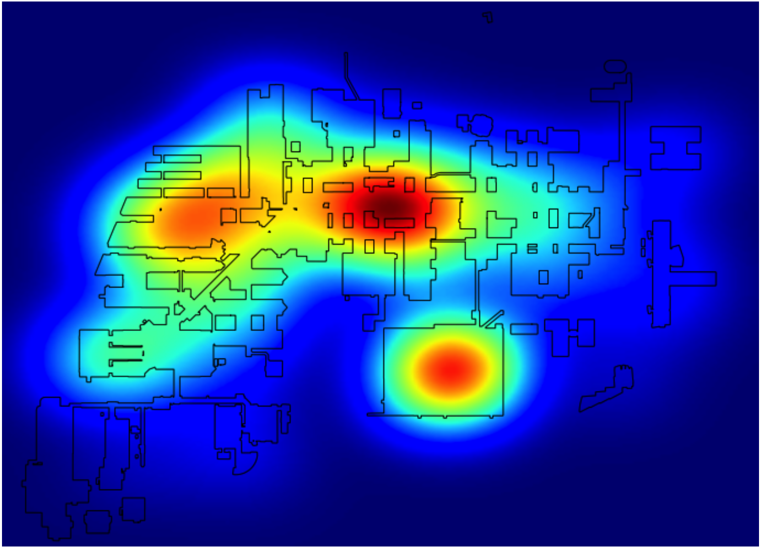
EPFL map: shortest paths to the bar



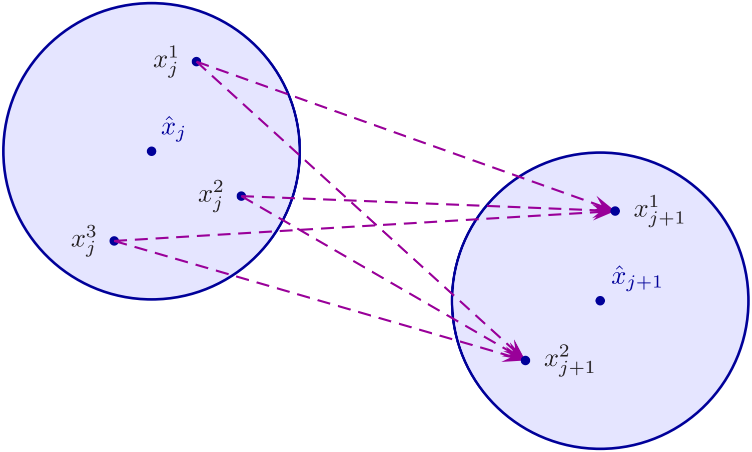
EPFL class schedules (bachelor/master students)



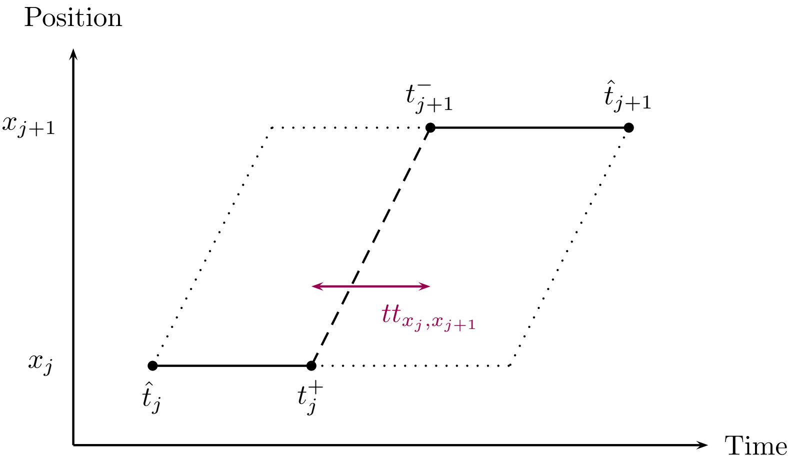
Attractivity on campus for students in civil engineering



Generation of activity-episode sequences



Generation of activity-episode sequences



Probabilistic measurement model

$$P(a_{1:K}|\hat{m}_{1:J}) \propto P(\hat{m}_{1:J}|a_{1:K}) \cdot P(a_{1:K})$$

where

- $P(a_{1:K}|\hat{m}_{1:J})$, the activity probability of an activity-episode sequence
- $P(\hat{m}_{1:J}|a_{1:K}) = \prod_{k=1}^{K} \prod_{j=1}^{J} P(\hat{x}_j^k|x_k)$, the measurement likelihood
- $P(a_{1:K})$, the prior based on attractivity of the POI

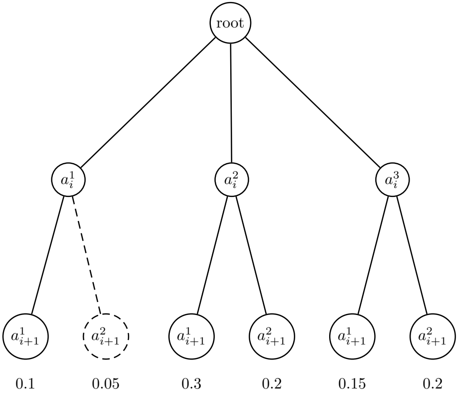
Intermediary measurements

Eliminate intermediary measurement if

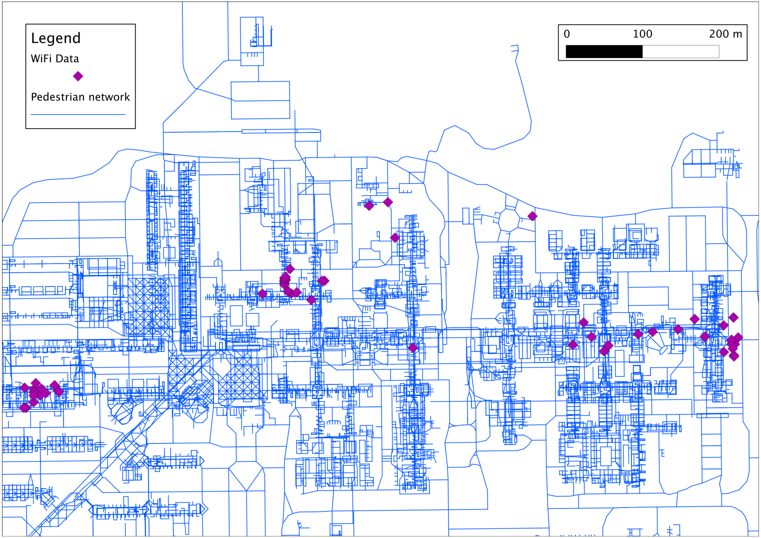
$$E(t^+) - E(t^-) < T_{min}$$

since we generate an activity episode at each measurement

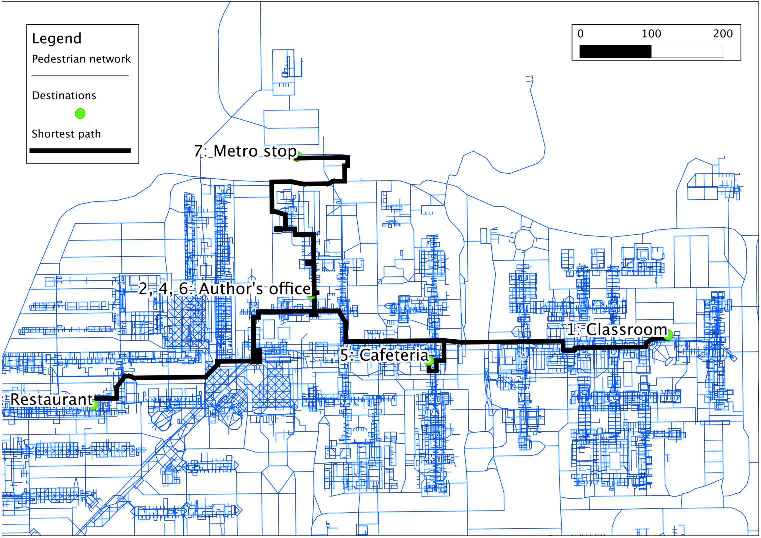
Sequence elimination procedure



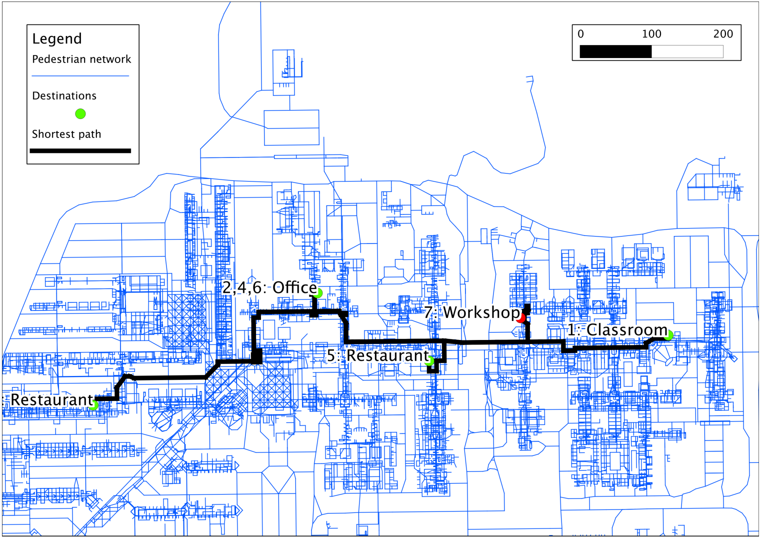
WiFi traces: No stop, no semantics



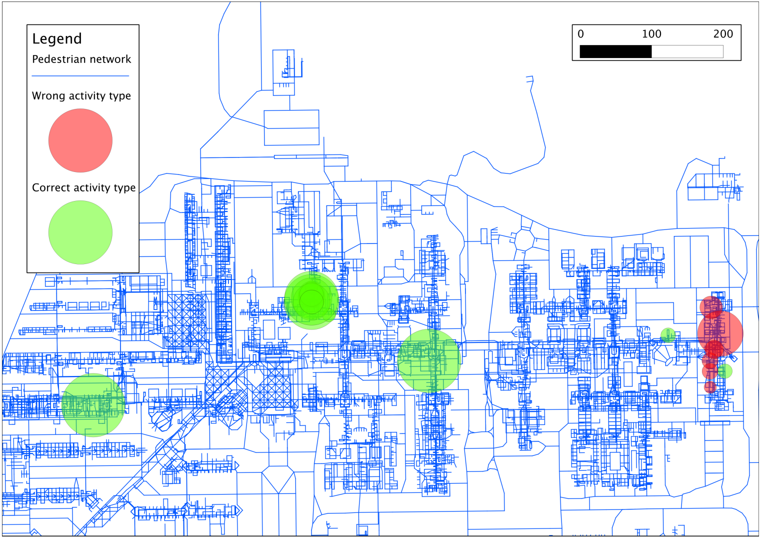
True activity sequence



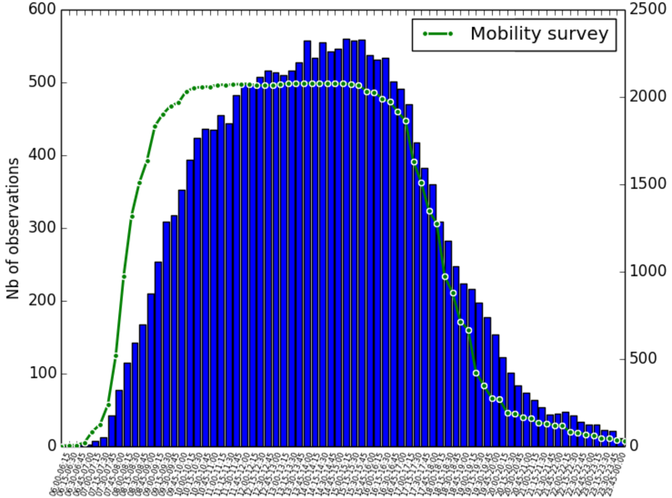
Output of the model: 1 candidate



Output of the model: 100 candidates

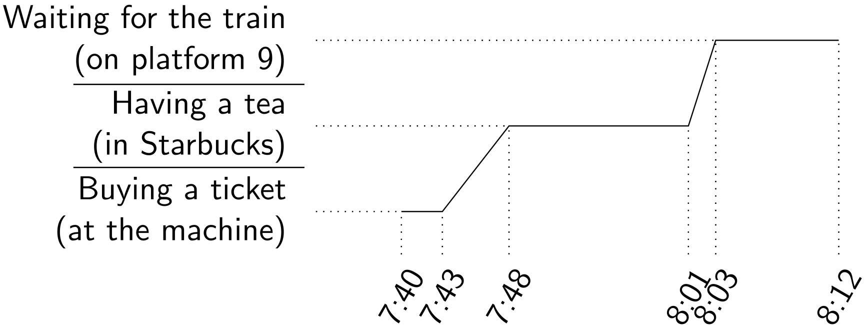


Aggregate results



Observations: activity patterns in a transport hub

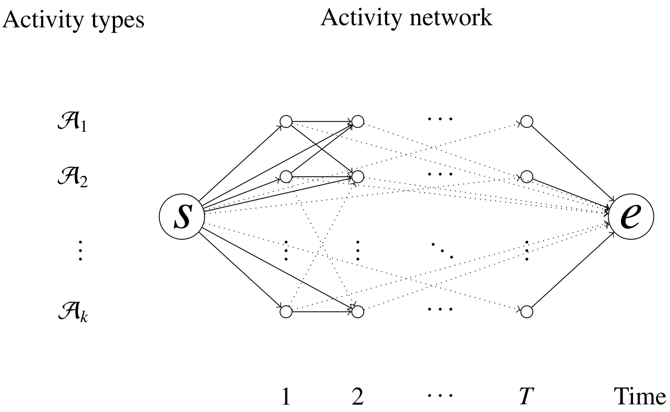
Activity types



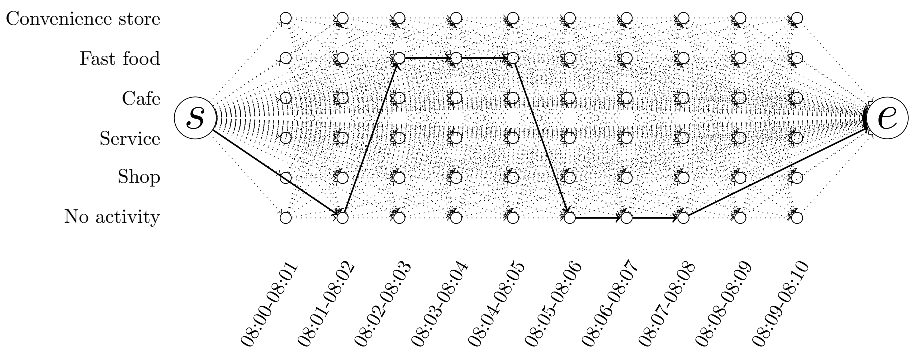
Modeling assumption

- Sequential choice:
  1. activity type, sequence, time of day and duration
  2. destination choice conditional on point (1.)
- Motivations:
  - Behavior: precedence of activity choice over destination choice
  - Dimensional: destinations $\times$ time $\times$ position in the sequence is not tractable

Activity network



Activity network

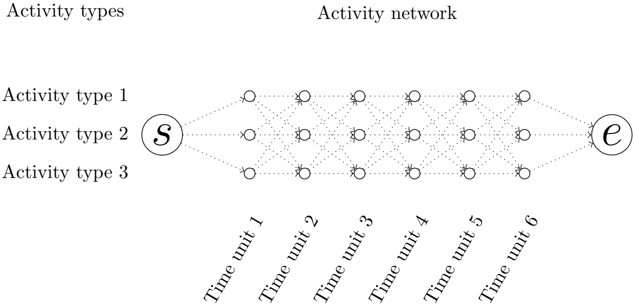


Modeling framework

- **Choice set generation**
  Metropolis-Hastings sampling of paths
- **Utility**
  - **time-of-day preferences**
  - **satiation effects**: marginal utility decreases with increasing duration
  - **scheduling contraints**: schedule delay approach
  - sampling correction
- **Correlation structure**
  CNL model with sampling of alternatives

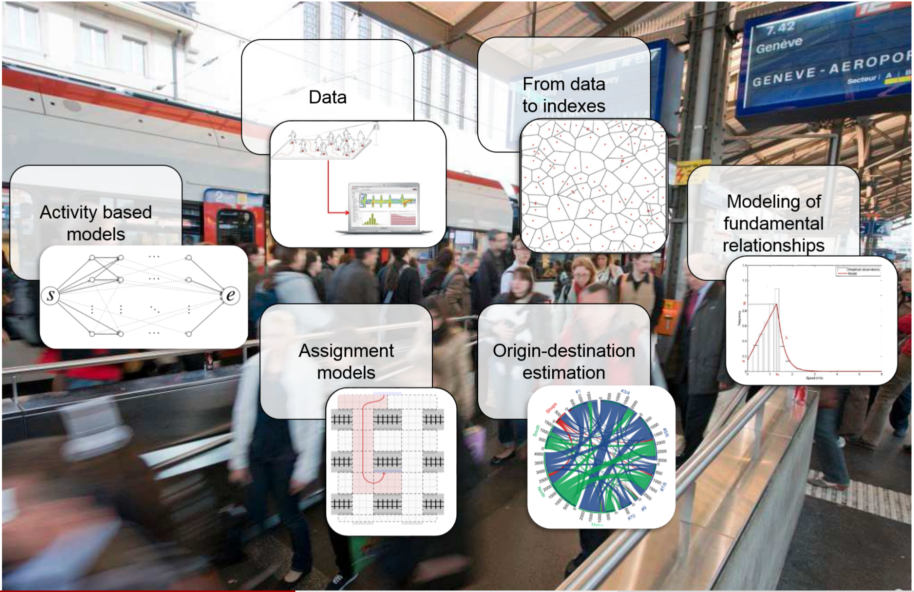
Validation with synthetic data

- 4000 synthetic observations
- 3 activity types, 6 time units: 729 alternatives



- Sampling 4 elements of the choice set per observation
- All parameters are recovered (t-tests with 95% confidence)

Summary


Data
From data to indexes
Activity based models
Modeling of fundamental relationships
Assignment models
Origin-destination estimation
7.42
Genève

Impacts

- Passengers
- Operators
- Society

- Fundamental research SNSF
- Applied research SBB

Pedflux: Pedestrian flow modeling in train stations

SBB CFF FFS

Pedestrian dynamics: flows and behavior

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Léman 2030: Flux piétons Gare de Lausanne

SBB CFF FFS