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*

- **2000**
  - > 25'000 travellers/day between Geneva and Lausanne*

- **2010**
  - > 50'000 travellers/day between Geneva and Lausanne*

- **2030**
  - > 100'000 travellers/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

Assignment models

Origin-destination estimation

Modeling of fundamental relationships
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 data

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

\[(t, x(t), y(t), \text{pedestrian}_{id})\]
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
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

![Graph showing empirical observations and model predictions for speed-density relationship. The graph illustrates the frequency of different speeds at various density levels, with key points labeled as α, β, and the critical speed density (v_m).]
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

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

- $\alpha$ - probability density corresponding to small speed values
  \[ \alpha(k) = a_\alpha \cdot k + b_\alpha \]

- $\beta$ - probability density corresponding to the most likely speed values
  \[ \beta(k) = a_\beta \cdot k + b_\beta \]

- $\lambda$ - 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

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

\[ \omega_l(v_i; \theta) = 1 - F_{vm}(v_i; \theta) \]

\[ \omega_e(v_i; \theta) = F_{vm}(v_i; \theta) \]
## Estimation results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Std err</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_\alpha$</td>
<td>0.029</td>
<td>$0.028e^{-04}$</td>
</tr>
<tr>
<td>$b_\alpha$</td>
<td>0.232</td>
<td>$0.069e^{-04}$</td>
</tr>
<tr>
<td>$a_\beta$</td>
<td>0.082</td>
<td>$0.035e^{-04}$</td>
</tr>
<tr>
<td>$b_\beta$</td>
<td>0.805</td>
<td>$0.105e^{-04}$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.988</td>
<td>$0.017e^{-04}$</td>
</tr>
<tr>
<td>$\nu_f$</td>
<td>1.126</td>
<td>$0.078e^{-04}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>4.753</td>
<td>$0.014e^{-04}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.291</td>
<td>$0.193e^{-04}$</td>
</tr>
</tbody>
</table>

$log L \quad -506208.864$

#parameters 8
#observations 756691
Comparison with deterministic models

Exponential specifications

![Graph showing exponential specifications and goodness of fit](Image)

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tregenza</td>
<td>0.406</td>
</tr>
<tr>
<td>Weidmann</td>
<td>0.441</td>
</tr>
<tr>
<td>Rastogi</td>
<td>0.221</td>
</tr>
<tr>
<td>Probabilistic</td>
<td>0.041</td>
</tr>
</tbody>
</table>
Comparison with deterministic models

Linear specifications

Goodness of Fit

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanariboon</td>
<td>0.591</td>
</tr>
<tr>
<td>Fruin</td>
<td>0.948</td>
</tr>
<tr>
<td>Navin and Wheeler</td>
<td>4.751</td>
</tr>
<tr>
<td>Lam</td>
<td>1.244</td>
</tr>
<tr>
<td>Older</td>
<td>1.044</td>
</tr>
<tr>
<td>SFPE</td>
<td>1.170</td>
</tr>
<tr>
<td>Probabilistic model</td>
<td>0.041</td>
</tr>
</tbody>
</table>
Density level: $< 0.1 \text{ped}/m^2$

#observations : 21178
Density level: $0.5 \text{ped}/m^2$

Frequency

Empirical observations
Model

#observations : 40470
Density level: $1 \text{ped}/m^2$

#observations : 10705
Density level: $1.5 \text{ped/m}^2$

Empirical observations:

- Number of observations: 6781

Model:

- Speed (m/s)
- Frequency

Graph shows the comparison between empirical observations and the model for different speeds.
Density level: $2 \text{ped/m}^2$

#observations : 2509
Density level: $2.5 \text{ped/m}^2$

#observations : 898
Density level: $3 \text{ped/m}^2$

# observations : 354
Density level: 3.5\textit{ped}/m$^2$

\#observations : 158
Density level: \(4 \text{ped}/m^2\)

\# observations: 73
Density level: $5 \text{ped/m}^2$

![Graph showing empirical observations and model predictions.](image)

#observations: 22
Density level: 6 \textit{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

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

- Train timetable, frequentation data
- Historical information
- Network layout

- Travel demand estimator
  - Congestion-sensitive estimation of travel demand

- Traffic assignment model
  - Demand-dependent estimation of infrastructural supply

- \( \delta(s) \)
- \( \sigma(d) \)

- Network state

- OD demand

- Link flow counts
- Trajectory recordings

Fixed-point problem
Travel demand estimator $\delta(s)$

Requirements:
- accurate prediction of dynamic OD demand $d^*$
- explicit integration of train timetable (→ ‘micro-peaking’) 
- aggregate model (no socio-economic data)

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

\[
f = A(s) \Delta(s) d + \eta \\
f = f_{\text{arr}} + f_0 \\
f_{\text{arr}} = \varphi_{\text{arr}}(w, s) + \varepsilon_{\text{arr}} \\
f_0 = \varphi_0(d, s) + \varepsilon_0
\]

where

- \(A(s):\) link-paths matrix
- \(\Delta(s):\) route choice matrix
- \(f_{\text{arr}}, f_0:\) train-induced arrival and ‘base’ flow
- \(\eta, \varepsilon_{\{\text{arr,0}\}}:\) error terms
\( \delta(s) \): Measurement model and problem formulation

Measurement model:

\[
\hat{d} = d + \omega_d
\]

\[
\hat{f}' = R_f f + \omega_f'
\]

Estimation problem (base vs. full estimator):

\[
d^* = \arg \min_{d \geq 0} \text{dist}_1 \left\langle \begin{pmatrix} \hat{f}' \\ \varphi'_\text{arr} + \varphi'_0 \end{pmatrix}, \begin{pmatrix} R_f \\ R_\varphi \end{pmatrix} f \right\rangle + \text{dist}_2 \left\langle \hat{d}, d \right\rangle
\]

where

\[
R_{f,\varphi} : \text{reduction matrices}
\]

\[
\omega_d, \omega_f' : \text{error terms}
\]

\[
\text{dist} \langle \cdot \rangle \{1,2\} : \text{weighted distance measures}
\]
$\delta(s)$: Specification & Case study

Model specification:
- demand-invariant supply $\sigma \neq f(d)$
  - 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

\[
\text{time} \quad \text{cumulative arrivals} \\
\text{observation} \quad \text{model}
\]

**Figure**: Model for train-induced arrival flows (\( m \): train, \( w_{m,\lambda}^{\text{off}} \): alighting volume, \( \lambda \): link, \( t_{m}^{\text{arr}} \): arrival time, \( s_{m,\lambda} \): dead time, \( C_{m,\lambda} \): flow capacity)
δ(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
Estimation of pedestrian OD flows in railway stations

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

![Graph showing demand in pedestrian underpasses, 10-day reference set, 2013](image)

**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

(a) 7:40–7:41
(b) 7:41–7:42
(c) 7:42–7:43
(d) 7:43–7:44
(e) 7:44–7:45
(f) 7:45–7:46
(g) 7:46–7:47
(h) 7:47–7:48

10 ped/min  100 ped/min
0  25  50  75  $\geq$ 100 ped/min
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
σ(d): Hughes’ theory for pedestrian flow (2002)

\[ \frac{\partial k_i(x, t)}{\partial t} + \nabla (k_i(x, t)v_i(x, t)) = 0 \]

\[ v_i(x, t) = -F(k(x, t))v_f \frac{\nabla \phi_i(x, t)}{\|\nabla \phi_i(x, t)\|} \]

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

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

- \( x \): 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, \( v_i \): velocity, \( v_f \): free-flow speed
- \( \phi_i(x, t) \): scalar potential field
σ(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

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

Cumulated number of students
in class by week based on
class schedules
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
- \( P(\hat{m}_{1:J} | a_{1:K}) = \prod_{k=1}^{K} \prod_{j=1}^{J} P(\hat{x}_{j} \mid x_{k}) \), the measurement likelihood
- \( P(a_{1:K}) \), the prior based on attractiveness of the POI
Intermediary measurements
Eliminate intermediary measurement if
\[ E(t^+) - E(t^-) < T_{min} \]
since we generate an activity episode at each measurement.
WiFi traces: No stop, no semantics
True activity sequence

Legend
Pedestrian network
Destinations
Shortest path

7: Metro stop
2, 4, 6: Author's office
5: Cafeteria
1: Classroom
Restaurant

M. Bierlaire et al. (EPFL)  Modeling opportunities  November 18, 2014  57 / 68
Output of the model: 1 candidate
Output of the model: 100 candidates
Aggregate results
Activity types

- Waiting for the train (on platform 9)
- Having a tea (in Starbucks)
- Buying a ticket (at the machine)

Observations: activity patterns in a transport hub
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 types

\[ A_1 \]
\[ A_2 \]
\[ \vdots \]
\[ A_k \]

Activity network

1 2 \[ \cdots \] T Time
Activity network

Convenience store
Fast food
Cafe
Service
Shop
No activity
Modeling framework

- **Choice set generation**
  Metropolis-Hastings sampling of paths

- **Utility**
  - time-of-day preferences
  - satiation effects: marginal utility decreases with increasing duration
  - scheduling constraints: 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

Activity types
- Activity type 1
- Activity type 2
- Activity type 3

Activity network

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

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

Conclusions
Impacts

- Passengers
- Operators
- Society
- Fundamental research SNSF
- Applied research SBB

Pedflux: Pedestrian flow modeling in train stations

Pedestrian dynamics: flows and behavior

Léman 2030: Flux piétons Gare de Lausanne