MATSIM-T: Aims, approach and implementation

KW Axhausen

IVT
ETH
Zürich

July 2007
Overview

- Structure and team
- Task and solution methods
- MATSIM aims

- Description of the scenario, population and its travel demand
- Progress on shortest path calculations
- Traffic flow model
- Scheduling and its utility function
- Improving the convergence
- System architecture

- Outlook and next steps
Structure

Software:

• Open-source project under GNU public licence

Coordination:

• Kai Nagel, TU Berlin

Data:

• Public sources, where available
• Private sources, when needed or as occasion arises
Current team

Strategy:

• Kai Nagel, TU Berlin
• Kay Axhausen, ETH Zürich
• Fabrice Marchal, LET, Lyon

Coordination of the implementation and project management:

• Michael Balmer, ETH Zürich
  • Marcel Rieser, TU Berlin
Current team: Implementation

- Michael Balmer, ETH
- David Charypar, ETH
- Francesco Ciari, ETH
- Jeremy Hackney, ETH
- Andreas Horni, ETH
- Nicolas Lefebvre, ETH
- Michael Löchl, ETH
- Fabrice Marchal, LET
- Konrad Meister, ETH
- Kai Nagel, TU Berlin
- Marcel Rieser, TU Berlin
- Nadine Schüssler, ETH
- David Stripgen, TU Berlin
Current funding sources

- Basic research support for the chairs
- (competitive) ETH research fund
- Swiss National Fund
- German Research Society
- EU Framework funding
- VW Foundation
- Swiss Commission for Technology and Information (KTI) (datapuls, Lucerne)
Task and solution methods
Understanding scheduling

- Budget constraints
- Capability constraints

- Generalised costs of the schedule
  - Generalised cost of travel
  - Generalised cost of activity participation
  - Risk and comfort-adjusted weighted sums of time, expenditure and social content
Degrees of freedom of activity scheduling

- Number and type of activities
- Sequence of activities
  - Start and duration of activity
  - Composition of the group undertaking the activity
  - Location of the activity
- Connection between sequential locations
  - Location of access and egress from the mean of transport
  - Vehicle/means of transport
  - Route/service
  - Group travelling together
Understanding supply

Slot: A path in the time-space environment, which allows moving or activity performance

- Regulated slots (e.g. table in a restaurant, reserved seat in a theatre, gate position of a plane, green light at a junction)

- Emergent slots (e.g. trajectory of a car on a motorway, players in a pub-soccer tournament)

Waiting time \sim Reserve capacity = Capacity – Demand for slots
What we would like to do?

Personal worlds of others

Social capital: stock of joint abilities, shared histories and commitments

Personal world

Household locations
Social network geography
Mobility tools

Projects

Biography

Learning
What would we like to do? Personal daily dynamics

- Activity repertoire (t)
- Activity repertoire (t+1)
- Activity calendar (t)
- Scheduling
- Activity schedule (t)
- Rescheduling, Execution
- Networks, Opportunities
- Unexecuted activities
- Updates, Innovations
- Mental map (t)
- Mental map (t+1)

Physiological needs
Commitments
Desires
Pending activities
What would we like to do? Personal long-term dynamics

(Life) goals (t) → Projects (t) [commitments] → Project sequence (t) → Replanning, Execution → Replanning, Execution → Unexecuted projects → Updates, Innovations, Reflection

Projects (t) [commitments] → Personal world (t) → Markets and networks → Projects (t) [commitments]

Definition of „Self“ Desires Pending projects

Planning, Negotiation

Planning, Negotiation

Planning, Negotiation

Planning, Negotiation

Planning, Negotiation

Planning, Negotiation
What do we (generally) do?

“Scenario”

Competition for slots on networks and in facilities

\[ k(t,r,j)_{i,n} \]

Mental map

Activity scheduling

Population

\[ \beta_{i,t,r,j,k} \]

Demand \( q \) are the \( i^{th} \) movements of person \( p \) from the current location at time \( t \) on route (connection) \( r \) to location \( j \). The resulting generalised costs \( k \) are used to adjust the schedules and to change the capacities \( C \) and prices \( P \) of facilities.
What should we do?

- Competition for slots on networks and in facilities
- Activity scheduling
- Parameter calibration

\[ k(t,r,j)_{i,n} \]

\[ q_i \equiv (t,r,j)_n \]

\[ \beta_{i,t,r,j,k} \]

Observed schedules and generalised costs
What would we like to do?

Locating, sizing and pricing of slots

Competition for slots on networks and in facilities

Mental map

Activity scheduling (private, commercial)

\[ k(t,r,j)_{i,n} \]

\[ C_f(t); P_f(t) \]

\[ q \equiv (t,r,j)_h \]
Classification criteria

• Steady state (equilibrium) ?

• Aggregate demands ?

• Complete and perfect knowledge ?

• Optimised schedules ?

• Degrees of freedom and detail of scheduling

• Modelling of capacity restrictions (movement, activities) ?
MATSIM-T aims (1): Steady-state version

- Steady state within 12 hours on a small multi-CPU machine
- 7.5 mio agents, parcels, navigation networks (Switzerland)
- Shared time-of-day dependent generalised costs of travel and activity participation
- Optimised scheduling
- Continuous time resolution; space: parcels; social networks
- Queuing for slots for movement and activities
MATSIM-T aims (2): Path-dependent version

- Path-dependent development; precise estimates within 12h on a small multi-CPU machine
- Large scale scenario
- Agent-specific, learned time-of-day dependent generalised cost of travel and activity participation
- Optimised scheduling at multiple decision points
- Continuous time resolution; space: parcels; social networks
- Queuing for slots for movement and activities
Current state (with a focus on ETH work)
Scenario: Facilities for 140’000 hectares

<facilities name="Swiss National Enterprise Census">
  <facility id="101" x="606300" y="281549">
    <activity type="shop">
      <capacity value="50"/>
      <opentime day="wkday" start_time="8:00:00" end_time="19:00:00"/>
      <opentime day="sat" start_time="8:00:00" end_time="16:00:00"/>
    </activity>
    <activity type="work">
      <capacity value="5"/>
      <opentime day="wkday" start_time="8:00:00" end_time="19:00:00"/>
      <opentime day="sat" start_time="8:00:00" end_time="16:00:00"/>
    </activity>
  </facility>
</facilities>
Scenario: Facilities – Data sources

Census of Workplace tables by hectare for:

<table>
<thead>
<tr>
<th>Description</th>
<th>Precision for</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Firms</td>
</tr>
<tr>
<td>Firm and employees per Sector 2/3</td>
<td>low</td>
</tr>
<tr>
<td>Firms and employees in classes per NOGA-2 digit class</td>
<td>middle</td>
</tr>
<tr>
<td>Firms per NOGA-4 digit class</td>
<td>precise</td>
</tr>
</tbody>
</table>
## Scenario: Facilities – Data sources

Example:

<table>
<thead>
<tr>
<th>NOGA Code</th>
<th>Description</th>
<th>Firms</th>
<th>FT-equivalents</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-45</td>
<td>(sector 2)</td>
<td>6</td>
<td>200</td>
</tr>
<tr>
<td>19</td>
<td>Leather and shoe production</td>
<td>2</td>
<td>0-9</td>
</tr>
<tr>
<td>20</td>
<td>Chemical industry</td>
<td>3</td>
<td>50-249</td>
</tr>
<tr>
<td>19.10</td>
<td>Leather production</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>19.20</td>
<td>Leather goods</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>19.30</td>
<td>Shoe production</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>
Scenario: Facilities – current allocation

- Read census of employment for each hectare
- Generate the required number of facilities
- Add activity work
- Set number of work place to minimum of class
- Distribute remainder proportionally to class size
- Add activity of use
- Add standard opening hours
- Randomly distribute on the hectare and attach to nearest link
Scenarios: Facilities - results
Scenario: Facilities - results
Scenario: Facilities - results
Scenario: Network

Simplification:

- Not connected links and subnetworks
- Remove nodes between unchanging link types
- Dead ends cut off

Results:

- Navteq (882'120 links) (25% loss of links/nodes)
- Teleatlas (1’288’757 links) (currently not in use)
- National transport model network
Scenario: Population

Alternatives:

• Artificial sample generation from Census marginals

• Census

• Private census with additional imputations (datapuls, Lucerne) plus household “formation”
Scenario: Mobility tools

Approach:

- MNL of mobility tool packages (Beige)
- Socio-demographics
- Location type
- Travel times to main centre by road and public transport

Data:

- National travel survey (MZ 2000)
Scenario: Demand

Data

- MZ 2000
- MZ 2005

Approach:

- Selection via conditional probability distribution from Chains
  * Person types frequencies
- ca. 50 activity chains for MZ 2000 (>95% of all cases)
- extended with activity durations given by MZ 2000
  ➔ 450 different activity chains (weighted)
Scenario: Demand - destinations

Home:
  - Random location in hectare

Work/school
  - Random allocation from census commuter matrices
  - Disaggregation to facilities inside municipality

Other locations
  - “Neighbourhood search” from given home/work/school locations (a variation of Gravity Model)
Scenario: Mode choice

Approach:

Fixed mode choice at the tour level (subtours are identified) as a function of
- Driving licence
- Mobility tool ownership
- Distance
- Age * season ticket

Data:
- MZ 2005
Scenario: Mode choice – Observed mode public transport
Scenario: Tour mode choice - Results

Market shares [%]

- Car
- Walk
- Public Transport
- Bicycle
- Other

Observed
Predicted
Interim summary: Demand generation

- **Number and type of activities**
- **Sequence of activities**

- Start and duration of activity
- Composition of the group undertaking the activity
- *Location of the activity*

- Connection between sequential locations

- Location of access and egress from the mean of transport
- *Vehicle/means of transport*
- Route/service
- Group travelling together
Progress on shortest – path calculation

Possibilities:

- **Bounding boxes**: Find out whether certain nodes can at all be on a shortest path and if not, do not consider them.

- **Multi-level approach**: Add shortcuts to the network where possible to bypass several edges at a time when routing.

- **Bi-directed search**: Start routing at the end and at the start node at the same time.

- **Goal-directed search**: Change the way the nodes are ranked such that nodes which are less likely to be on the shortest path are also less likely to be visited. The most popular algorithm that uses this technique is A*.
Basic algorithm: A*

Step 1: Improve data structure for list of candidate nodes (7% reduction)
Step 2: Improve handling of the “visited flag” (~ length of route)
Step 3: Detect dead ends (1/4 of nodes in the NavTeq network) (50% reduction)
Step 4: Use euclidian-distance to destination to rank candidates (50-80% reduction)
Step 5: Use intermediate landmarks (80-90% reduction)
Speed-up: Landmark selection

- Divide network into sectors containing an equal number of nodes
- For each sector, choose a node that is farthest away from the center
- Check that the landmarks are not too close to each other. If so, narrow one sector and choose a new landmark within the sector
Speed up of shortest-path calculations

Free flow conditions: 97-99% reductions
Loaded conditions: 95% reductions

on navigation networks
Interim summary: Plan

<person id="22018">
  <plan score="157.72" selected="yes">
    <act type="h" x100="703600" y100="236900" link="5757" end_time="07:35:04" />
    <leg num="0" mode="car" dep_time="07:35:04" trav_time="00:16:31">
      <route>1900 1899 1897</route>
    </leg>
    <act type="w" x100="702500" y100="236400" link="5749" dur="08:12:05" />
    <leg num="1" mode="car" dep_time="16:03:40" trav_time="01:10:22">
      <route>1899 1848 1925 1924 1923 1922 1068</route>
    </leg>
    <act type="l" x100="681450" y100="246550" link="2140" dur="01:20:00" />
    <leg num="2" mode="car" dep_time="" trav_time="00:34:35">
      <route>1067 1136 1137 1921 1922 1923 1925 1848 1899</route>
    </leg>
    <act type="h" x100="703600" y100="236900" link="5757" />
  </plan>
</person>
Traffic flow model

“Scenario”

Competition for slots on networks and in facilities

$\beta_{i,t,r,j,k}$

Population

$\beta_{i,t,r,j,k}$

Mental map

Activity scheduling

$q_i \equiv (t,r,j)_h$
Approach

- Physical (VISSIM)
- CA (TRANSIMS)
- Q (Cetin)
- Q event (MATSIM)
- parallel Q event (MATSIM)
- Meso (METROPOLIS)
- Macro (VISUM)
Parallel q-event driven simulation with gaps

- Approach
- Fundamental diagram (on a ring motorway)
- Domain decomposition
- Test
Q-event: Approach without gaps
Q-event: Approach with gaps

Link 1

Link 2

\( t \)

\( v_{\text{gap}} \)
Q-event: Implementation details

• Squeezing to avoid grid-lock

• Inflow capacity = 110% of outflow capacity (1800 veh/h* lanes)

• Vehicles are served in order of arrival at the junctions

• C++ with binary data interface to MATSim-T
Q-event: Fundamental diagram
Q-event: Test setup

- Road network of the federal states of Germany Berlin and Brandenburg
- 11.6k nodes, 27.7k links
- 7.05M person days
- Average number of trips per agent: 2.02
- Average length of a trip: 17.5 links
- Total: 249M road segments to be traveled
- 77 min on a single dual-core CPU (1.6GH; 256 GB RAM)
Q-event: Integrated domain decomposition
Q-event: Parallelisation

The graph shows the relationship between the number of processors and the speedup factor for microsimulation and linear speedup. The data points for microsimulation are represented by a solid line, while the dashed line represents the linear speedup. The speedup factor increases linearly with the number of processors.
Scheduling and its utility function
Utility function: Individual schedules

\[ U_{\text{plan}} = \sum_{i=1}^{n} U_{\text{act},i} + \sum_{i=2}^{n} U_{\text{trav},i-1,i} \]

\[ U_{\text{act},i} = U_{\text{dur},i} + U_{\text{wait},i} + U_{\text{late} \cdot \text{ar},i} + U_{\text{early} \cdot \text{dp},i} + U_{\text{short} \cdot \text{dur},i} \]
Utility function: Travel

\[ U_{\text{trav},i-1,i} = \begin{cases} \beta_{\text{trav}} \cdot t_{\text{trav},i-1,i} & \text{if } t_{\text{trav},i-1,i} \geq 0 \\ 0 & \text{else} \end{cases} \]

\[ \beta_{\text{trav}} = -12 \text{ Euro/h} \]
Utility function: Waiting

\[
U_{\text{wait},i} = \begin{cases} 
\beta_{\text{wait}} \cdot t_{\text{wait},i} & \text{if } t_{\text{wait},i} \geq 0 \\
0 & \text{else}
\end{cases}
\]

\[\beta_{\text{wait}} = -6 \text{ Euro/h} \]
Utility function: Late arrival

\[
U_{\text{late.ar},i} = \begin{cases} 
\beta_{\text{late.ar}} \cdot (t_{\text{start},i} - t_{\text{latest.ar},i}) & \text{if } t_{\text{start},i} \geq t_{\text{latest.ar},i} \\
0 & \text{else}
\end{cases}
\]

\[\beta_{\text{late.ar}} = -18 \text{ Euro/h}\]
Utility function: Departing early

\[ U_{\text{early.dp},i} = \begin{cases} 
\beta_{\text{early.dp}} \cdot (t_{\text{earliest.dp},i} - t_{\text{end},i}) & \text{if } t_{\text{end},i} \leq t_{\text{earliest.dp},i} \\
0 & \text{else} 
\end{cases} \]

\[ \beta_{\text{early.dp}} = -6 \text{ Euro/h} \]
Utility function: Duration too short

\[ U_{short\_dur, i} = \begin{cases} \beta_{short\_dur} \cdot \left( t_{shortest\_dur, i} - (t_{end, i} - t_{start, i}) \right) & \text{if } t_{end, i} - t_{start, i} \leq t_{shortest\_dur, i} \\ 0 & \text{else} \end{cases} \]

\[ \beta_{short\_dur} = -6 \text{ Euro/h} \]
Utility function: Activity performance

\[ U_{dur,i} = \begin{cases} 
\beta_{dur} \cdot t^* \cdot \ln \left( \frac{t_{dur,i}}{t_{0,i}} \right) & \text{if } t_{0,i} \leq t_{dur,i} \\
0 & \text{if } 0 \leq t_{dur,i} < t_{0,i} \\
\alpha \cdot t_{dur,i} & \text{else (} \alpha > 0 \text{)}
\end{cases} \]

\[ \alpha = 10 \text{ Euro/h} \]
\[ \beta_{dur} = 6 \text{ Euro/h} \]
Utility function: Home-Work-Home example
Utility function: Revising the parameters

\[ \beta_{\text{dur}} = 6 \, \text{€/h}, \quad \beta_{\text{wait}} = 0 \, \text{€/h}, \]
\[ \beta_{\text{trav}} = -6 \, \text{€/h}, \quad \beta_{\text{late.ar}} = -18 \, \text{€/h}, \]
\[ \beta_{\text{early.dp}} = 0 \, \text{€/h}, \quad \beta_{\text{short.dur}} = 0 \, \text{€/h}, \]
\[ \beta_{\text{neg.dur}} = -18 \, \text{€/h}. \]

Vickrey model ratio is 1:2:3:

\[ \beta_{\text{wait}} : \beta_{\text{trav}} : \beta_{\text{late.ar}} = 0 : -6 : -18 \]

Considering the opportunity costs of not performing an activity while waiting or travelling, one has to subtract \( \beta_{\text{dur}} \):

\[ \beta_{\text{wait,eff}} : \beta_{\text{trav,eff}} : \beta_{\text{late.ar,eff}} = -6 : -12 : -18 \]
Scheduling: Current Planomat(s)

Version 1:

- GA optimiser of durations and starting times
- Retains time-of-day profile of generalised costs

Version 2:

- CMA-ES (Covariance Matrix Adaptation Evolutionary strategy) of durations and starting times
- Retains time-of-day profile of generalised costs
Scheduling: Once and future Planomat

- **Number and type of activities** out of an agenda
- **Sequence of activities**
  - **Start and duration of activity**
  - **Composition of the household group** undertaking the activity
  - **Location of the activity**
    - **Connection between sequential locations**
      - **Location of access and egress from the mean of transport**
      - **Vehicle/means of transport**
      - **Route/service**
      - **Group travelling together**
Improving the convergence
Initial approach

Method:

• Plan everybody at Iteration 1
• Replan for a fixed share

Convergence:

• On mean performance
Improved approach

Method:

• Plan for everyone at Iteration 1
• Replan for a fixed, but decreasing share over the number of iterations

Measurement:

• Aggregates
Impact of improved approach
System architecture

- XML standards
- Data handling tools (aggregation/disaggregation)
- Data base
- Iteration handling and control of convergence
- Models
Architecture: JAVA - packages
Architecture: Data flow
Architecture: Data flow

MATSIM-T (Multi-Agent Transport SIMulation Toolkit)

Agent Initialization (MATSim-INI)
- used by
  - Individual Initial Demand Modeling (MATSIM-IIDM)
    - activity location
    - choice set
    - allocation
    - activity chain
    - allocation
    - etc...

Data Base (MATSIM-DB)
- Agent DB
  - contains: description, knowledge and demand of each person

Data Preparation (MATSIM-DATA)
- Data Fusion & Consolidation (MATSIM-FUSION)
  - (dis-)aggregations
  - fusions
  - consistency check
  - etc...

Population Modeling (MATSim-POP)
- synthetic persons (IPF)

provides
- plans xml

MATSim-DB
- Parsers

world xml
- facilities xml
- network xml
- matrix xml
- etc...

Census DB
- refers to

Network DB
- refers to

Facilities DB
- refers to

World DB
- etc...

Matrix DB
- etc...

Agent DB
- used by
- refers to
- contains: description, knowledge and demand of each person

fuse
Architecture: Data flow
What the tools allow: Kt. Zürich (1)

Population

1 Conversion

World (raster & municipalities)

2 Mapping
   raster → municipality

3 Distribution
   adding activity chains

Plans
   - Person attributes
   - Home location (R, M)

Plans
   - Person attributes
   - Home location (R, M)

R: Hectare; M: Municipality
What the tools allow: Kt. Zürich (2)

**Plans**
- Person attributes
- Home location (R, M)
- Activity chain
- Prim-act location (M)

**4 Distribution**
work & education
commuters

**5 Disaggregation**
municipality → raster

World (raster & municipalities)

**6 Secondary Activity Location Choice (extern)**

**Plans**
- Person attributes
- Home location (R, M)
- Activity chain
- Prim-act location (R, M)
- Sec-act location (R)

**Plans**
- Person attributes
- Home location (R, M)
- Activity chain
- Prim-act location (R, M)

**4 Distribution**
work & education
commuters

**5 Disaggregation**
municipality → raster

Pendlermatrix 2000 (M)

Landuse (R)
MATSim-EA (C++) with „planomat“ (1% sample)
MATSim-EA (C++) with „planomat“: Iteration 0

The diagram shows the number of departures (in 5min time bins) over the course of the day, with distinct lines representing different types of activities:
- Leisure chains (hll,hllh,hlslh)
- Education chains (heeh,hehl)
- Work chains (hwh,hwhwh,hwswwh,hwwh)
- Shop chains (hs,hsh)
- Mixed chains (i.e. helh,hlsh,hwth,...)

Key events:
- Departure from work
- Departure from leisure
- Departure from first leisure act
- Departure from last work act

Opening times:
- Work & education
- Shop
- Leisure

The graph provides insights into the temporal patterns of activity departures throughout the day.
MATSim-EA (C++) with „planomat“: Iteration 400

The diagram shows the number of departures (in 5min time bins) over the hour of the day for different types of chains and activities:

- Leisure chains (hllh, hllh, hlslh)
- Education chains (heeh, heh)
- Mixed chains (i.e. helh, hlish, hwh, ...)
- Work chains (nwh, hwh, hwh, hwwwh, hwwwh)
- Shop chains (hsh, hssh)

Key events include:
- Departure from home to work
- Departure from home to leisure
- Departure from last work act
- Departure from last leisure act
Current tasks: Functionality

• Improving the realism of the scenario (e.g. Opening hours, parameter distributions)

• Validation of the current Switzerland scenario
  • Switzerland scenario in 12h to steady state

• Functional expansion the Planomat (Mode choice, Destination choice)

• Parameter estimation for Planomat

• Interface to UrbanSim
Future tasks: Functionality

- Integration of social network data structures
- Explicit social network choices
- Addition of supply agents (car sharing, demand responsive transport, retail location, parking pricing, road pricing) (Traffic control)
Current tasks: Usability (Shareability)

- Analysis tools
- Anonymous test data sets
- Improved documentation
- Web and sourceforge maintenance
- User support (Visits to ETH/Berlin)
## Acknowledgements

<table>
<thead>
<tr>
<th>Category</th>
<th>Contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facilities:</td>
<td>Konrad Meister</td>
</tr>
<tr>
<td>Mobility tools:</td>
<td>Francesco Ciari, Sigrun Beige</td>
</tr>
<tr>
<td>Tour mode choice:</td>
<td>Francesco Ciari</td>
</tr>
<tr>
<td>Router:</td>
<td>Nicolas Lefebvre</td>
</tr>
<tr>
<td>Traffic flow model:</td>
<td>David Charypar</td>
</tr>
<tr>
<td>Planomat:</td>
<td>Konrad Meister, David Charypar</td>
</tr>
<tr>
<td>Supply agents:</td>
<td>Michael Löchl, Francesco Ciari</td>
</tr>
<tr>
<td>Social networks:</td>
<td>Jeremy Hackney</td>
</tr>
<tr>
<td>System architecture:</td>
<td>Michael Balmer</td>
</tr>
<tr>
<td>System integration:</td>
<td>Michael Balmer, Marcel Rieser</td>
</tr>
<tr>
<td>System management:</td>
<td>Michael Balmer, Marcel Rieser</td>
</tr>
</tbody>
</table>
Acknowledgements

Past contributors:

- Michael Bernard, ETH
- Sigrun Beige, ETH
- Ulrike Beuk, TU-Berlin
- Dr. Nurhan Cetin
- Phillip Fröhlich, Modus, Zürich
- Dr. Christian Gloor
- Dr. Bryan Raney, HP America
- Marc Schmitt, ETH
- Andreas Vogel, TU Berlin
More information

www.matsim.org

www.vsp.tu-berlin.de

www.ivt.ethz.ch/vpl/publications/reports