Integration of the MATSim model into a four step transport model, using scientific workflow systems, DAFNI and OpenMOLE

J. Raimbault\textsuperscript{1,*} and M. Batty\textsuperscript{1}

\*j.raimbault@ucl.ac.uk

\textsuperscript{1}Center for Advanced Spatial Analysis, University College London

DAFNI Roadshow - UCL
November 22nd 2021
Urban transportation models

**MATSim model: heterogenous data and integration of many sub-models**

Source: [Balmer et al., 2009]
Land-use transport models as a progressive complexification through coupling of detailed sub-models

<table>
<thead>
<tr>
<th>Transport model</th>
<th>T1 No public transport no modal split</th>
<th>T2 Public transport no logit 24 h</th>
<th>T3 Public transport logit peak hour</th>
<th>T4 Multi-modal activity-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land-use model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1 None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2 Activity and judgement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3 No market-based land allocation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L4 Logit allocation with price signals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L5 Market-based land-use model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L6 Activity-based land-use model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Models

<table>
<thead>
<tr>
<th>Models</th>
<th>Very slow Networks</th>
<th>Land use</th>
<th>Slow Workplaces</th>
<th>Housing</th>
<th>Fast Employment</th>
<th>Population</th>
<th>Immediate Goods transport</th>
<th>Travel</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOYCE</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CUFM</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>DELTA/START</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>HUDS</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IMREL</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IRPUD</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>ITLUP</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>KIM</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>LILT</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MEPLAN</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>METROSIM</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MUSSA</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>POLIS</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RURBAN</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>STASA</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TRANUS</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>URBANSIM</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Source: [Wegener and Fürst, 2004]
Sensitivity analysis of MATSim

Large scale urban/transport ABMs must be validated for relevant and robust policy applications

A few example of MATSim validation or sensitivity analysis in the literature: uncertainty [Bienzeisler et al., 2021], sensitivity analysis [Zhuge et al., 2019], discrete choice parameters [Hörl, 2021]

Research objective:

Provide a modular and open implementation of MATSim generic to any UK urban area and test global sensitivity analysis methods on it
MATSim model integration

*Modular four-step multimodal transportation model using open source projects and data*

**Integrated models:**

- MATSim model (MATSim Community) for the transportation system
  https://www.matsim.org/  [Axhausen et al., 2016]
- SPENSER model (University of Leeds) for the synthetic population
  https://github.com/nismod/microsimulation  [Spooner et al., 2021]
- QUANT model (CASA, University College London) for spatial interactions to generate home-work plans
  http://quant.casa.ucl.ac.uk/  [Batty and Milton, 2021]
- spatialdata library (OpenMOLE community) for data processing
  https://github.com/openmole/spatialdata  [Raimbault et al., 2020]
Data and implementation

**Data:** Generic for any Functional Urban Area (GHSL [Florczyk et al., 2019]) or any arbitrary area in the UK: NOMIS census, Ordnance Survey roads, Traveline National Dataset for public transport

**Workflow systems for the integration of submodels:**
- DAFNI facility (UKCRIC) [https://dafni.ac.uk](https://dafni.ac.uk)
- OpenMOLE software [https://openmole.org/](https://openmole.org/) [Reuillon et al., 2013]

**Implementation**
- Synthetic SPENSER population distributed at the micro level using OSM buildings
- QUANT model to generate home-work commuting flows, job locations determined by sampling flows
- Network and plans (simple uniform commuting plans) prepared into MATSim xml files and fed into a multimodal MATSim model
- Models integrated as Docker containers
Road network preprocessing: implemented into the `spatialdata scala` library [Raimbault et al., 2020]

Public transport data: from TransXchange (TNDS) to GTFS using UK2GTFS R package [Morgan, 2021]; GTFS to MATSim xml schedule using `pt2matsim` library
OpenMOLE model exploration open source software
[Reuillon et al., 2013]

Enables seamlessly (i) model embedding; (ii) access to HPC resources; (iii) exploration and optimization algorithms

https://openmole.org/
Explored parameters

Parameter sampled for the sensitivity analysis:

- Functional Urban Area (spatial context [Raimbault et al., 2019])
- Random seed (influence of stochasticity [Bienzeisler et al., 2021])
- Synthetic population sampling
- Modal choice parameters [Hörl, 2021]: mode constants in scoring function (car, public transport, walking)
Role of stochasticity

FUA: Taunton

Trip departure time

Density

Random seed
- 1001
- 1008
- 1557
- 2104
- 2428
- 2436
- 2533
- 309
- 3671
- 918
Global Sensitivity Analysis

*Method based on the estimation of conditional relative variances* [Saltelli et al., 2010]

**First order index**

\[
S_i = \frac{\text{Var} \left( E_{X \sim i} (Y|X_i) \right)}{\text{Var} [Y]}
\]

is the expected relative variance reduction if \(X_i\) would be fixed.

**Total effect index**

\[
ST_i = \frac{E_{X \sim i} \left[ \text{Var} (Y|X_{\sim i}) \right]}{\text{Var} [Y]}
\]

is the expected relative variance if all factors but \(X_i\) are fixed (includes interaction effects).
## GSA results

<table>
<thead>
<tr>
<th>output</th>
<th>$\beta_W$</th>
<th>$\beta_{PT}$</th>
<th>$\beta_C$</th>
<th>S</th>
<th>FUA</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>carShare</td>
<td>0.023</td>
<td>0.0058</td>
<td>0.0079</td>
<td>3.94</td>
<td>0.165</td>
<td>0.379</td>
</tr>
<tr>
<td>ptShare</td>
<td>0.0081</td>
<td>0.0074</td>
<td>0.0030</td>
<td>2.164</td>
<td>0.04</td>
<td>0.0169</td>
</tr>
<tr>
<td>walkShare</td>
<td>0.0059</td>
<td>0.0017</td>
<td>0.0074</td>
<td>0.834</td>
<td>0.16</td>
<td>0.082</td>
</tr>
<tr>
<td>avgTripDistance</td>
<td>0.11</td>
<td>0.19</td>
<td>0.087</td>
<td>0.04</td>
<td>1.51</td>
<td>0.049</td>
</tr>
<tr>
<td>avgScore</td>
<td>0.43</td>
<td>0.0003</td>
<td>0.0039</td>
<td>0.057</td>
<td>0.0085</td>
<td>0.0073</td>
</tr>
</tbody>
</table>

*Total order Saltelli indices obtained with $\simeq 50$ model runs*
MATSim sensitivity analysis
→ Preliminary results, but suggest a strong influence of stochasticity, context and parameters

Large scale open, reproducible and validated models?
→ Still a long way to go: a lot of tuning even with containers; issue of infrastructure (memory vs CPUs)

Role of visualisation
→ Some models are intrinsically interactive/visual (cf QUANT): compatible with workflow systems / integration? (change in model function)

Future developments
→ dynamical strong coupling of models (SPENSER/QUANT); applications to policies
Conclusion

→ Open, reproducible and validated urban models as elementary bricks towards larger integrated models
→ Workflow systems provide model construction and exploration/validation
→ A preliminary global sensitivity analysis of multimodal MATSim for a generic implementation on UK FUAs

Open repositories

https://github.com/JusteRaimbault/UrbanDynamics/Models/Matsim for containers and workflows

https://github.com/openmole/spatialdata for data processing

Acknowledgements

DAFNI platform/Champions program; Urban Dynamics Lab Grant EP-SRC EP/M023583/1


