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

J. Raimbault[1,*] and M. Batty[1]

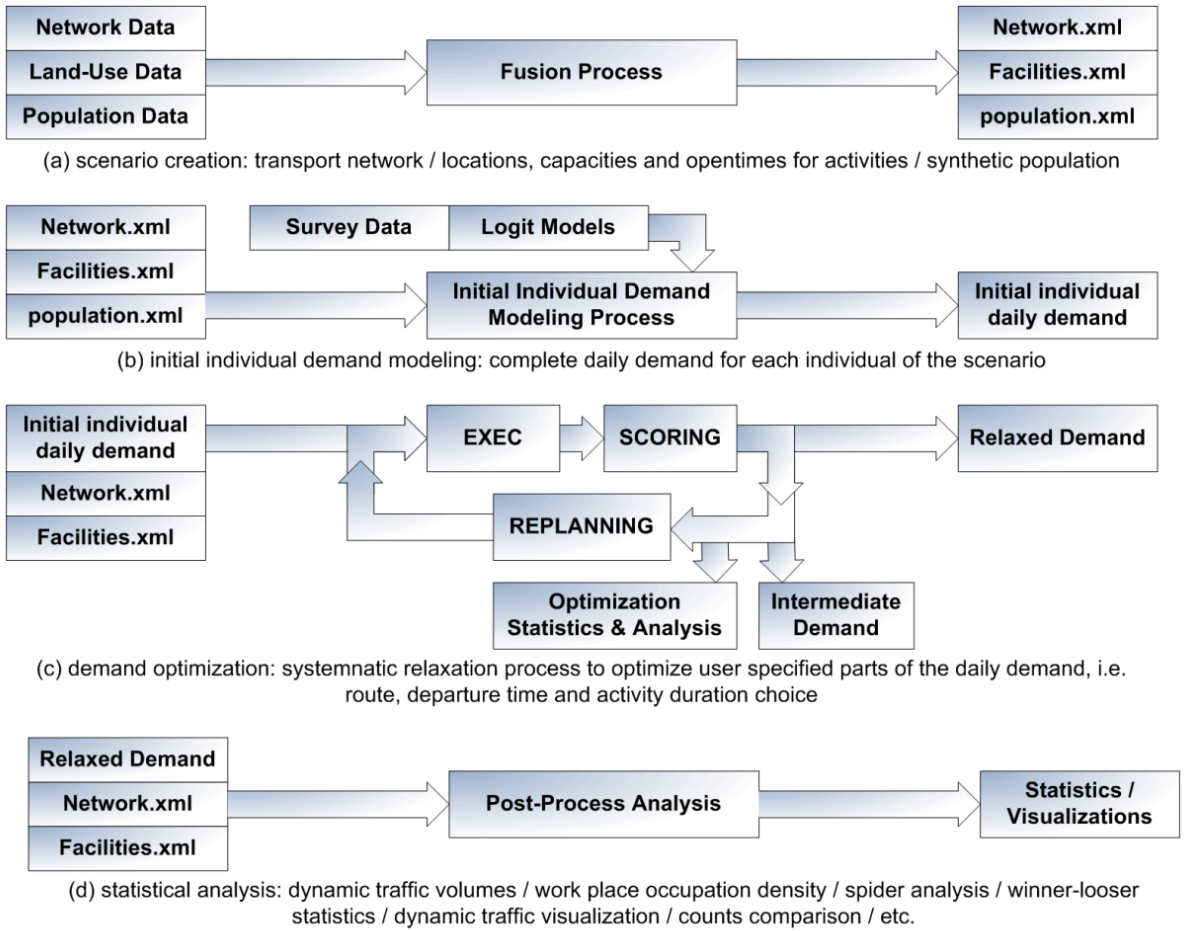
*j.raimbault@ucl.ac.uk

[1]Center for Advanced Spatial Analysis, University College London

DAFNI Roadshow - UCL
November 22nd 2021

Urban transportation models

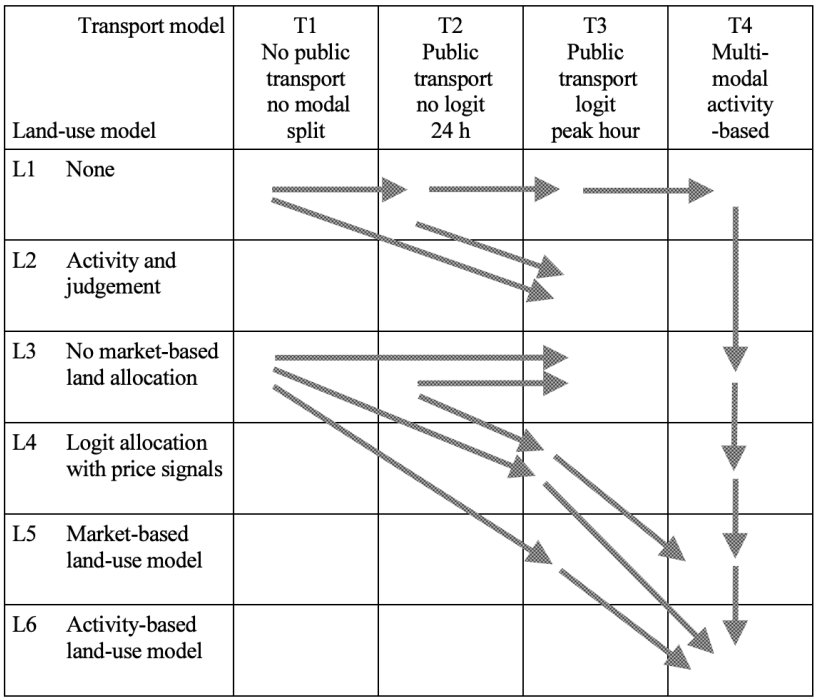
MATSim model: heterogenous data and integration of many sub-models



Source: [Balmer et al., 2009]

Land-use transport models

Land-use transport models as a progressive complexification through coupling of detailed sub-models



Models	Speed of change							
	Very slow		Slow		Fast		Immediate	
	Networks	Land use	Work-places	Housing	Employ-ment	Popula-tion	Goods transport	Travel
BOYCE	+				+	+		+
CUFM		+	+	+	+	+		
DELTA/START	+	+	+	+	+	+	+	+
HUDS				+	+	+		
IMREL	+	+	+	+	+	+		+
IRPUD	+	+	+	+	+	+		+
ITLUP	+	+			+	+		+
KIM	+				+	+	+	+
LILT	+	+	+	+	+	+		+
MEPLAN	+	+	+	+	+	+	+	+
METROSIM	+	+	+	+	+	+		+
MUSSA	+	+			+	+		+
POLIS		+			+	+		+
RURBAN		+			+	+		+
STASA	+	+	+	+	+	+	+	+
TRANUS	+	+	+	+	+	+	+	+
URBANSIM		+	+	+	+	+		+

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, OrdnanceSurvey roads, Traveline National Dataset for public transport

Workflow systems for the integration of submodels:

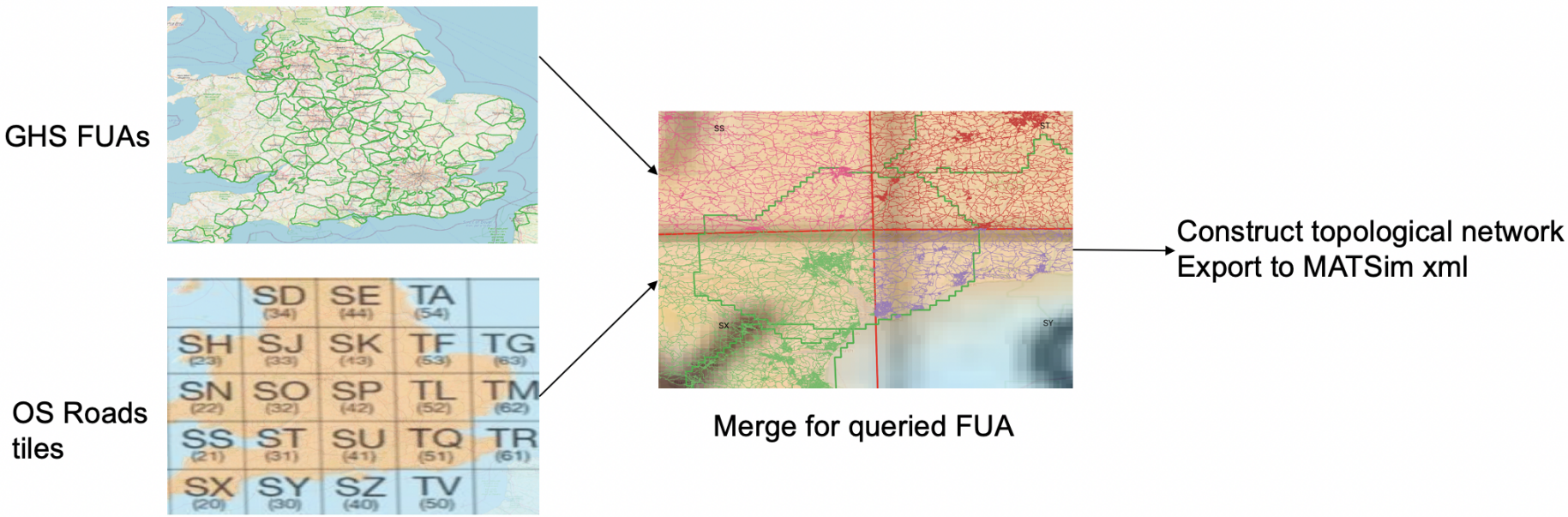
- DAFNI facility (UKCRIC) `https://dafni.ac.uk`
- OpenMOLE software `https://openmole.org/` [Reuillon et al., 2013]

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

Data preparation

→ 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 workflow engine

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)

DAFNI Integration

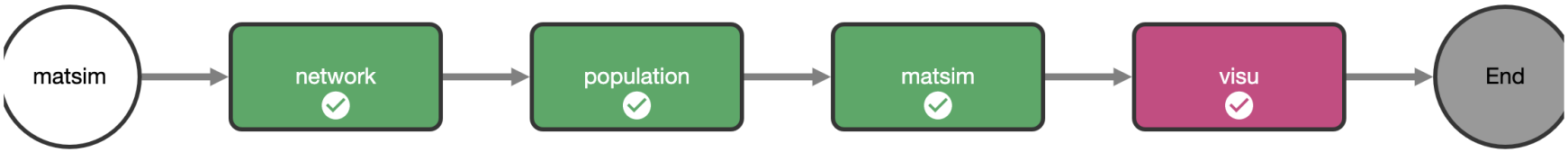
Home / Workflows / Workflow Status

Workflow Status

Key: Iterator | Template | Model | Publisher | Visualisation

Reset position

Click and drag on the white area around the Workflow to pan the canvas.

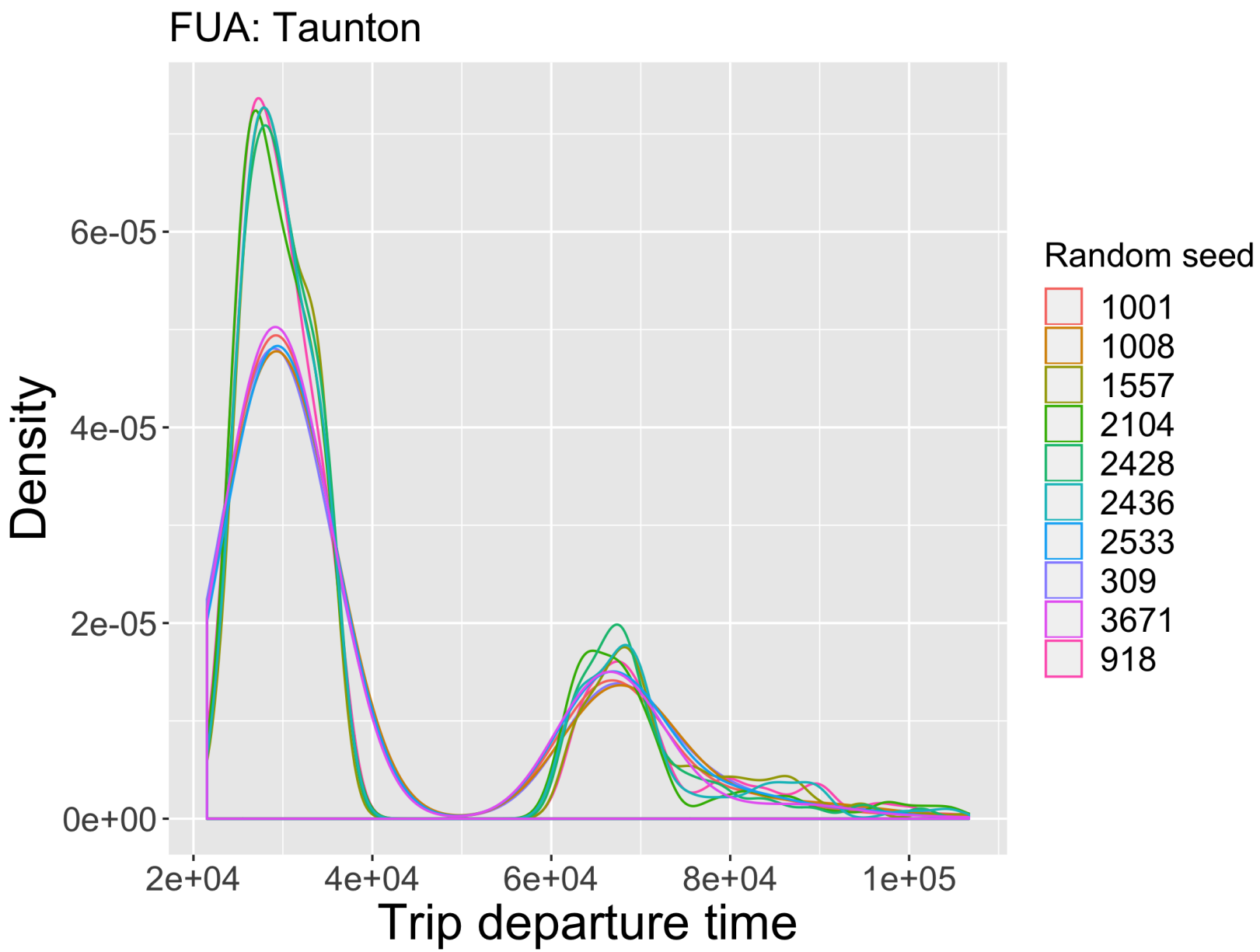


OpenMOLE integration: script

Saltelli.oms

```
val popSample = Val[Double]; val seed = Val[Double]; val fuaInd = Val[Double]; val utilitywalk = Val[Double]; val utilitycar = Val[Double]; val utilitypt = Val[Double]
val fuaName = Val[String];val fuaFile = Val[File]
val iterations = Val[Int]; val threads = Val[Int]
val dataDir = Val[File]; val nwDir = Val[File]
val planFile = Val[File]; val networkFile = Val[File]; val transitScheduleFile = Val[File]; val transitVehiclesFile = Val[File]
val outputDir = Val[File]
val carShare = Val[Double]; val ptShare = Val[Double]; val walkShare = Val[Double]; val avgTripDistance = Val[Double]; val avgScore = Val[Double]

val setFUA = ScalaTask(
  """
  val potFUAs = fuaFile.content.split("\n").toSeq
  val fuaName = potFUAs((fuaInd*potFUAs.size.toDouble).toInt)
  """
) set (
   (inputs,outputs) += (popSample,seed,utilitywalk,utilitycar,utilitypt),
   inputs += (fuaInd,fuaFile),
   outputs += (fuaName),
   fuaFile := workDirectory / "areas.txt"
)

val popcommand = "ls /data/LAD;export SBT_OPTS=\"-Xmx16G\";sbt \"project spatialdata; runMain org.openmole.spatialdata.application.matsim.RunMatsim --synthpop --popMode=detailed --jobMode=sample --planMode=default --sample=${popSample} --FUAName=${fuaName}"+
"--FUAFile=/data/GHS_FUA_UCDB2015_GLOBE_R2019A_54009_1K_V1_0_WGS84.gpkg --LAFile=/data/LAD/LAD_WGS84.shp --OAFile=/data/OA/OA2011_WGS84.shp --MSOAFile=/data/MSOA/EnglandWalesScotland_MSOAWGS84.shp --SPENSERDirs=/data/SPENSER/England,/data/SPENSER/Scotland,/data/SPENSER/
" --OSMBuildingsDirs=/data/OSM/England,/data/OSM/Scotland,/data/OSM/Wales --QUANTDataDir=/data/QUANT/ --output=Plans\";ls /data;[ ! -f Plans_${fuaName}.xml ] && touch Plans_${fuaName}.xml"

val population = ContainerTask(
    image = workDirectory / "images" / "matsim-population-1.0-0a52e7b80f.tar.gz",
    command = popcommand,
    workDirectory = "/root/spatialdata/library",
    containerSystem = Singularity(),
    errorOnReturnValue = false
) set (
   (inputs,outputs) += (fuaName,popSample,seed,utilitywalk,utilitycar,utilitypt),
   inputFiles += (dataDir,"/data/"),
   outputFiles += ("Plans_${fuaName}.xml",planFile),
   dataDir := workDirectory / "rawdata"
)

val readfiles = ScalaTask("""
   val networkFile = nwDir / (fuaName+".xml.gz")
   val transitScheduleFile = nwDir / (fuaName+"_transit_schedule.xml.gz")
   val transitVehiclesFile = nwDir / (fuaName+"_transit_vehicles.xml.gz")
""
) set (
  (inputs,outputs) += (fuaName,popSample,seed,utilitywalk,utilitycar,utilitypt,planFile),
  inputs += (nwDir),
  outputs += (networkFile, transitScheduleFile, transitVehiclesFile),
  nwDir := workDirectory / "Network" / "runtime"
)
```

Role of stochasticity



Global Sensitivity Analysis

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

First order index

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

is the expected relative variance reduction if X_i would be fixed

Total effect index

$$ST_i = \frac{E_{\mathbf{X}_{\sim i}}\left[Var\left(Y|\mathbf{X}_{\sim i}\right)\right]}{Var\left[Y\right]}$$

is the expected relative variance if all factors but X_i are fixed (includes interaction effects)

GSA results

output	β_W	β_{PT}	β_C	S	FUA	p
carShare	0.023	0.0058	0.0079	3.94	0.165	0.379
ptShare	0.0081	0.0074	0.0030	2.164	0.04	0.0169
walkShare	0.0059	0.0017	0.0074	0.834	0.16	0.082
avgTripDistance	0.11	0.19	0.087	0.04	1.51	0.049
avgScore	0.43	0.0003	0.0039	0.057	0.0085	0.0073

Total order Saltelli indices obtained with $\simeq 50$ *model runs*

Discussion

MATSim sensitivity analysis

$\rightarrow$ Preliminary results, but suggest a strong influence of stochasticity, context and parameters

Large scale open, reproducible and validated models?

$\rightarrow$ Still a long way to go: a lot of tuning even with containers; issue of infrastructure (memory vs CPUs)

Role of visualisation

$\rightarrow$ Some models are intrinsically interactive/visual (cf QUANT): compatible with workflow systems / integration? (change in model function)

Future developments

$\rightarrow$ 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 EPSRC EP/M023583/1

References I

Axhausen, K., Horni, A., and Nagel, K. (2016).
The multi-agent transport simulation MATSim.
Ubiquity Press.

Balmer, M., Rieser, M., Meister, K., Charypar, D., Lefebvre, N., and Nagel, K. (2009).
Matsim-t: Architecture and simulation times.
In *Multi-agent systems for traffic and transportation engineering*, pages 57–78. IGI Global.

Batty, M. and Milton, R. (2021).
A new framework for very large-scale urban modelling.
Urban Studies, page 004209802098225 2.

References II

Bienzeisler, L., T., L., Wage, O., Huck, L., and Friedrich, B. (2021).
Uncertainty and variability analysis of agent-based transport models.
EWGT2021 Proceedings.

Florczyk, A. J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., et al. (2019).
Ghsl data package 2019.
Luxembourg, EUR, 29788(10.2760):290498.

Hörl, S. (2021).
Integrating discrete choice models with matsim scoring.
Procedia Computer Science, 184:704–711.

References III

Morgan, M. (2021).
UK2GTFS: Converts UK transport timetable datasets to GTFS format.
https://github.com/itsleeds/uk2gtfs, https://itsleeds.git.io/uk2gtfs/.

Raimbault, J., Cottineau, C., Le Texier, M., Le Néchet, F., and Reuillon, R. (2019).
Space matters: Extending sensitivity analysis to initial spatial conditions in geosimulation models.
Journal of Artificial Societies and Social Simulation, 22(4).

Raimbault, J., Perret, J., and Reuillon, R. (2020).
A scala library for spatial sensitivity analysis.
GISRUK.

References IV

Reuillon, R., Leclaire, M., and Rey-Coyrehourcq, S. (2013).
Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models.
Future Generation Computer Systems, 29(8):1981–1990.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010).
Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index.
Computer physics communications, 181(2):259–270.

Spooner, F., Abrams, J. F., Morrissey, K., Shaddick, G., Batty, M., Milton, R., Dennett, A., Lomax, N., Malleson, N., Nelissen, N., et al. (2021).
A dynamic microsimulation model for epidemics.
Social Science & Medicine, page 114461.

References V

Wegener, M. and Fürst, F. (2004).
Land-use transport interaction: State of the art.
Available at SSRN 1434678.

Zhuge, C., Shao, C., Wang, S., and Hu, Y. (2019).
Sensitivity analysis of integrated activity-based model: Using matsim as an example.
Transportation Letters, 11(2):93–103.