A FIRST APPROACH TOWARDS AN AUTOMATIZED PREPARATION OF INPUT DATA FOR THE AGENT-BASED DEMAND MODEL TAPAS

D. Krajzewicz, A. von Schmidt, S. Nieland, M. Heinrichs, A. Schengen

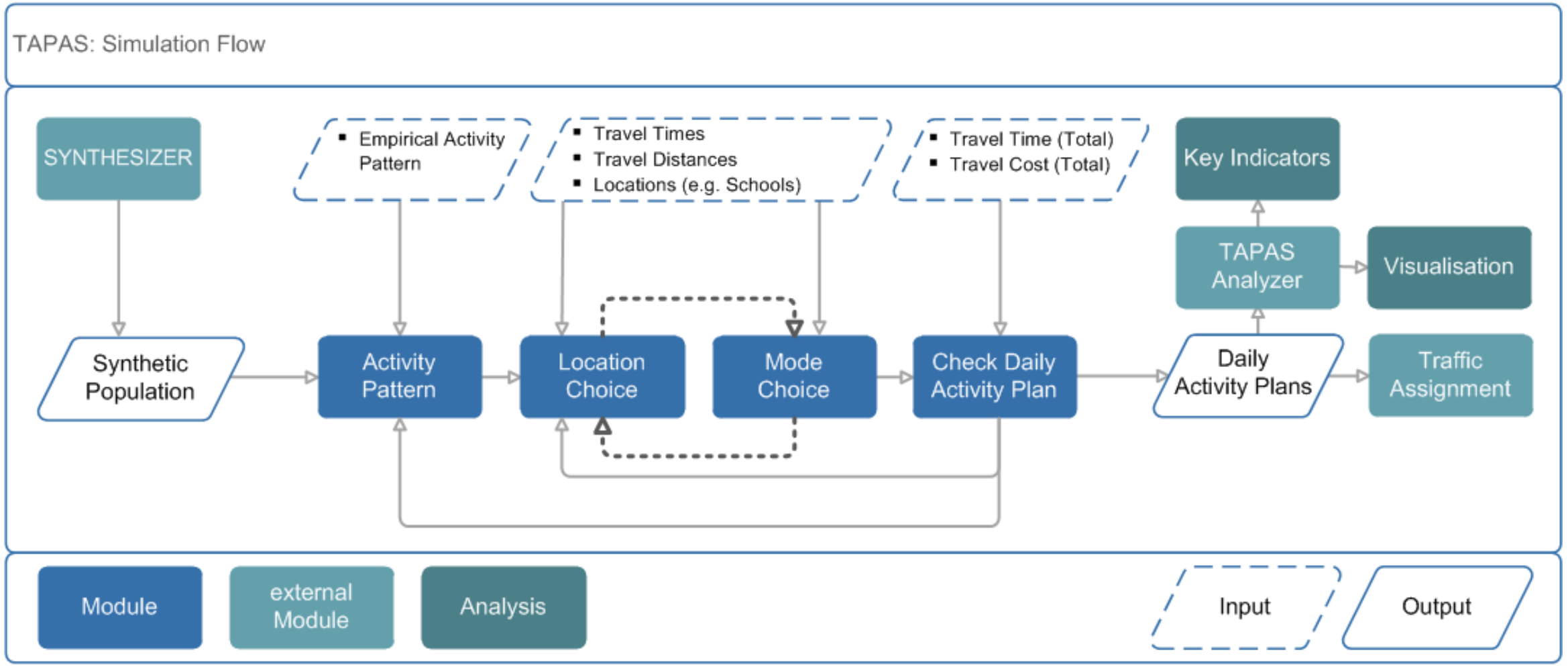
ABMTrans 2025, 22nd-24th of April 2025, Patras, Greece

DLR

TAPAS
Introduction

- TAPAS is an agent-based demand model
  - Every single person is modelled for itself
  - Has attributes, such as age, gender, employment status (including pupils and students), maybe a driver license, a public transport season ticket and/or a bike
  - Is classified into a person group
  - Chooses its daily activities matching the groups‘ behavior from a set of 50k empirically collected daily activity plans
- Peculiarity: Grouping to households
  - Certain mobility options (e.g. the owned car) may be used by different persons along the day
  - The ownership and the access to the option is represented on household level
  - This also counts for the mobility budget
- Available as open source (https://github.com/DLR-VF/TAPAS)

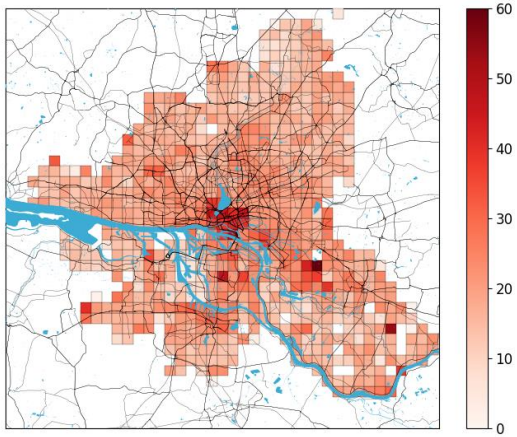
TAPAS
Introduction



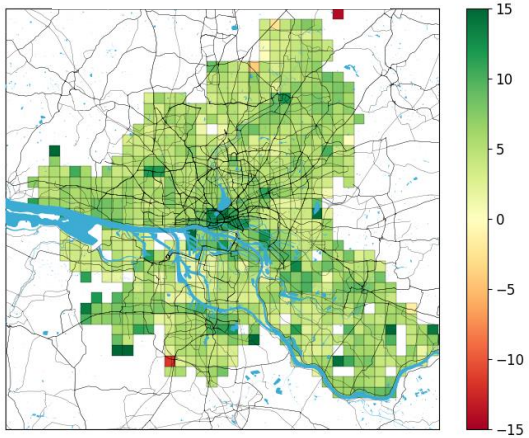
TAPAS
Introduction

- TAPAS is own of our Institute's most-used models
  - Mainly for showing changes in mobility behavior when introducing new mobility options (feeder shuttles, automated private vehicles, new public transport lines, on-demand traffic, etc.)
  - Allows for showing the effects on a fine spatial scale and distinguishing different person groups
  - Example: effects of introducing free public transport in the city of Hamburg

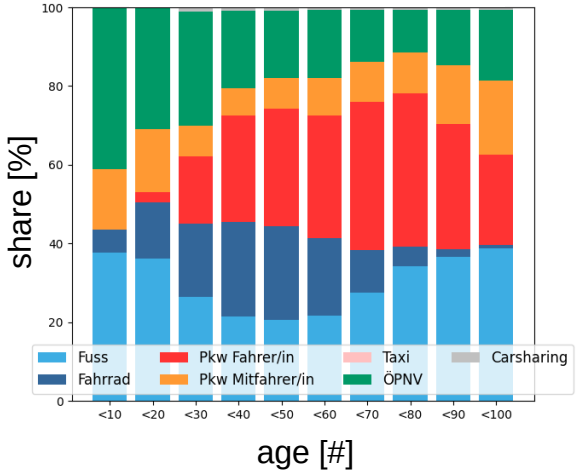
share of pt use 2030



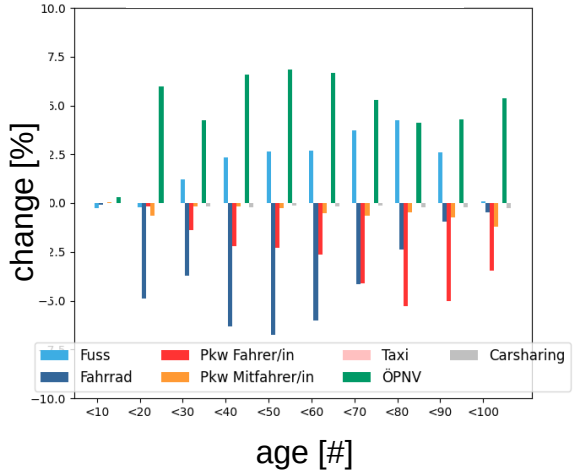
change when free



mode choice by age



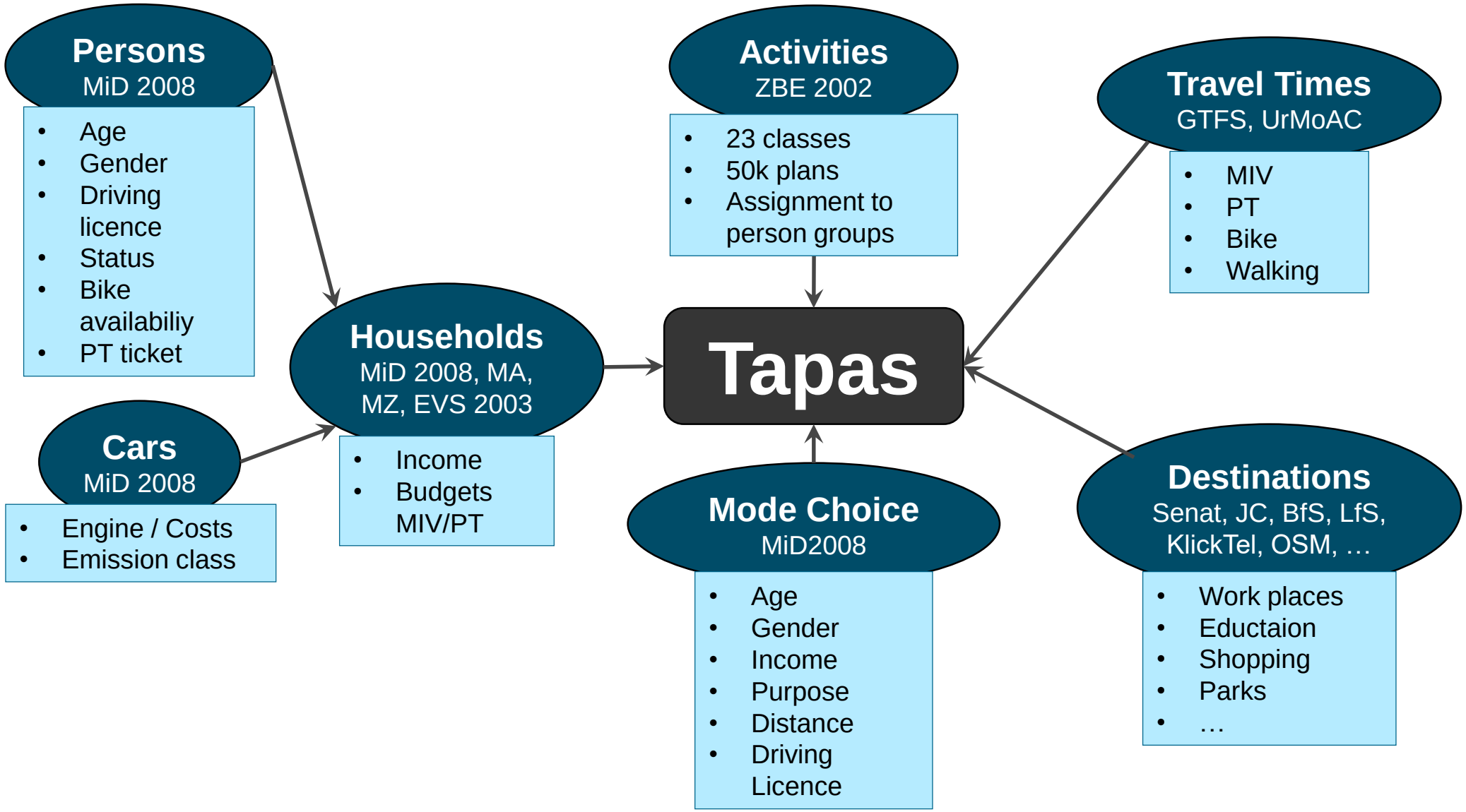
change when free



TAPAS
Data Needs

- TAPAS is a reliable and valuable tool
- … but …
  - A lot of data is needed
  - Data is usually available at different spatial aggregation levels (districts, cities, …)
  - The process of setting up a new simulation area takes time and is expensive
    - Manual collection of data sources
    - Cleaning and consolidation
- … so we try to set something up that it with two clicks
  - not yet there – we need to write a config file and start four Python scripts subsequently…
- … work in progress …

TAPAS
Data Needs

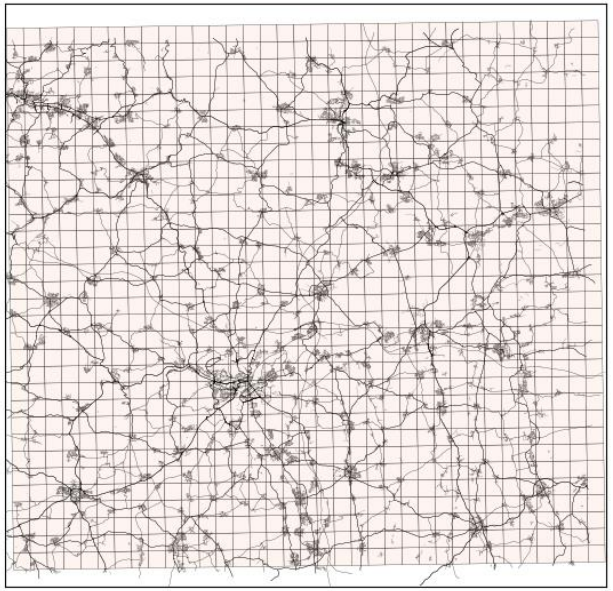


TAPAS
Data Needs

- Four “areas” of needed data
  - Traffic assignment zones
  - The disaggregated and allocated population (persons/households, mobility options like car, bike, public transport ticket)
  - Allocated destinations (activity locations) per purpose (work, education, shopping, leisure, …) with capacities
  - Travel times and distances per mode + number of interchanges for public transport

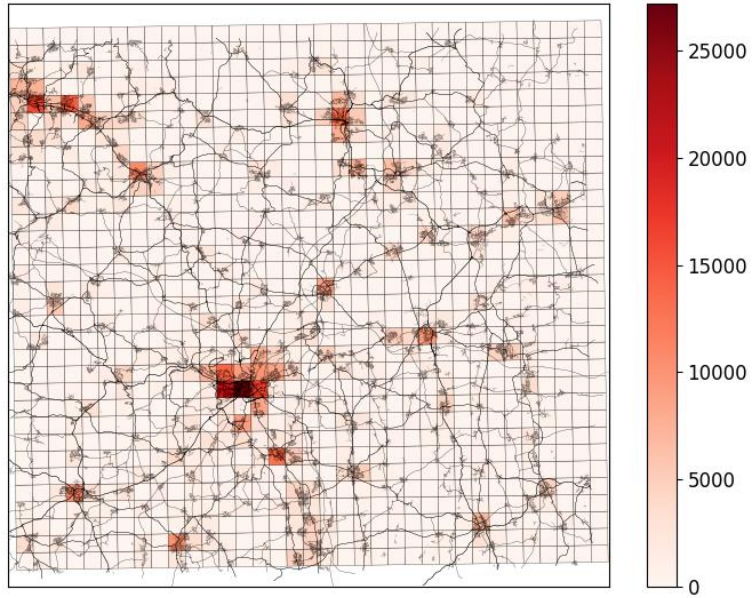
Generated Data
Traffic Analysis Zones (TAZ)

- In a first step, we decided to divide the area into plain rectangular cells…
- 2 km × 2 km seems a good balance between spatial resolution and computation time
  - Frankly, when computing the destinations, TAPAS collects destinations from surrounding cells; the smaller the cells, the more have to be visited
  - On the other hand, we want to have a high spatial resolution
- Usually, TAZ are built by building areas with similar mobility behavior
- Future improvements:
  - Take into account barriers etc.
  - Investigate, whether cells can be joined



Generated Data
Population (I)

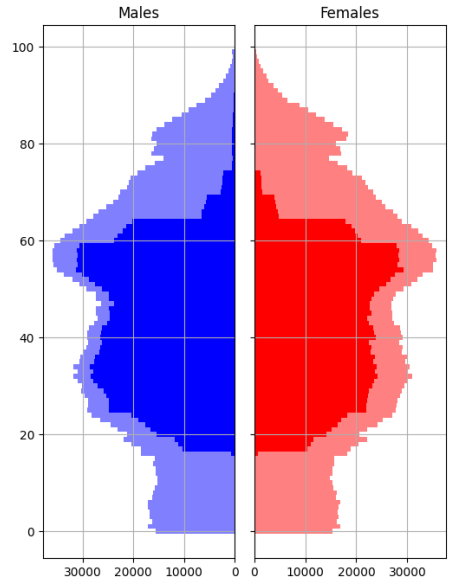
- In Germany, a data set with all addresses with estimated numbers of households and inhabitants is available
- The tool uses this data, converts the addresses to the format used by TAPAS and builds
  - The households
  - The persons
  - Builds cars for households
  - Computes the households‘ incomes



Generated Data
Population (II)

Building the population

- First, the number of persons per household is built
- Then, the persons are generated
  - The first person is always an adult
    - the age is selected randomly from the age distribution of persons above 20 years
    - the gender is selected randomly, taking onto account the respective share
  - The second person is an adult, or a child
    - If it‘s an adult, the age is selected from the first adult built +/i 5 years
  - All subsequent persons are children
    - All child ages are selected from the age distribution



Generated Data
Population (III) – Drawbacks and Issues

- Very uniform households
  - No same-sex households
  - Quite a fix age differences
  - No dependency between the age of the adults and the ages of the children
- It seems, like students should be treated explicitly
  - Specific types of living
  - Specific age range
  - Additional to the base population
- Future improvements:
  - Incorporate further statistics
  - Build a valid distribution of households / living groups (flat shares)

Generated Data
Activity Places (I)

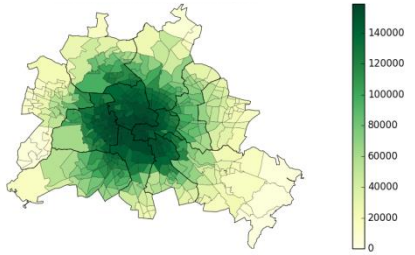
- Usually the data set that is the most difficult one to gather
  - No common statistics
  - A large variety of types
- We rely on OpenStreetMap (OSM) here
- In OSM, buildings may have a tag that describes their purpose
- We built a complete mapping from all OSM-tags to TAPAS activities
  - Overall, OSM delivers about 95% of the needed data
  - Both, numbers and capacities have to be rescaled

Generated Data
Activity Places (II)

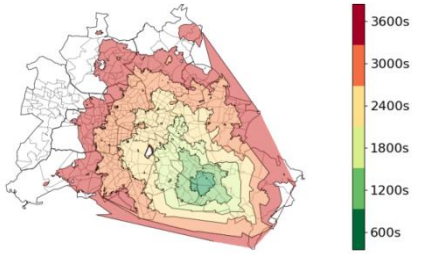
- Besides the purpose, buildings need capacities
  - For the visitors
  - For the working places
- We use Bosserhoff-values here
  - Bosserhoff has developed scaling factors from purpose to the number of visitors / workers in dependence of the size
  - The size is determined from the footprint and the number of the building's levels – if not directly given
  - Unfortunately, Bosserhoff-constants are not freely available…

Data Preparation
Travel Time Matrices

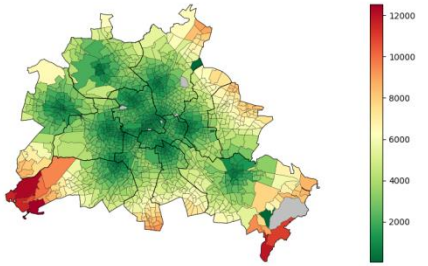
- … we already covered that in a talk last year
  - Krajzewicz, Daniel und Schengen, Alain (2024) Computation of mode-dependent travel time matrices for an agent-based demand model computed using a standalone accessibility tool. Procedia Computer Science (238), Seiten 779-784. Elsevier. doi: 10.1016/j.procs.2024.06.091. ISSN 1877-0509.
- We use an open source tool dedicated to computing accessibility measures named UrMoAC (https://github.com/DLR-VF/UrMoAC)
- … fits perfect for this purpose
- For TAPAS, travel time and distance matrices for walking, using a bike, motorized individual transport and public transport are computed
  - + number of interchanges for public transport



Average amount of accessible workplaces using motorized public transport



An isochrone for intermodal usage of public transport and bike starting at the DLR in Berlin-Adlershof



Average travel time to the next StEP-3-center by foot

Observations

- TAPAS is almost 20 years old
  - A lot of work was “archeology” – why is this data field here, what is it for, is it still needed?
  - Complexity of computation and of needed data had to be revisited

→ The development not only speeds up our work, we got new points of view at the underlying data

Summary

- With four calls to Python scripts, we have a running TAPAS scenario
  - 1. call: Build population and activity places
  - 2. call: Upload generated data to the database
  - 3. call: Build travel time / distance matrices
  - 4. call: Convert matrices to TAPAS
- Overall duration: less than one hour
- But: there are still some things to improve…
  - Mix data available at different spatial levels
  - Find data sources usable to replicate European areas (Germany is basically covered)
  - Improve determination of location capacities
  - Including a “two click” interface…

Impressum

Topic:	**A first approach towards an automatized preparation of input data for the agent-based demand model TAPAS**
Date:	24.04.2025
Authors:	D. Krajzewicz, A. von Schmidt, S. Nieland, M. Heinrichs, A. Schengen
Institute:	Institute of Transport Research
Images:	DLR, CC-BY