Using ANN to assess the mode choice resulting from MATSim

Ruurd Buijs Thomas Koch Luk Knapen

CWI (Centrum voor Wiskunde en Informatica)
VU Amsterdam and Hasselt University

May 31st 2022

Overview: Project Goal

For the Amsterdam region

1. mode choice models (MNL, ANN) have been estimated from revealed preference based on GPS traces
2. estimate a *trip based* mode choice model from the MATSim output
3. compare the properties of the implicit mode choice mechanism in MATSim to a choice model extracted from observations

Overview: Project Stages

Stages

1. Prepare MATSim plans from FEATHERS (activity based travel demand predictor) : The Netherlands, nationwide
2. Run MATSim simulation
3. Estimate trip based mode choice model from MATSim output
4. Compare choice models (from MATSim, from GPS observations

Status

1. Building new `plans.xml` generator delayed by health event.

Location Disaggregation: Locations and Positions

Locations

1. Nearly all *activity based models* predict travel between TAZ (travel analysis zones)
2. Location choice models operate at TAZ level spatial resolution

Positions

1. MATSim requires position (street address) level resolution

Mind *Schedule Morphing* by trip parameter updating

1. by location to position disaggregation
2. by MATSim

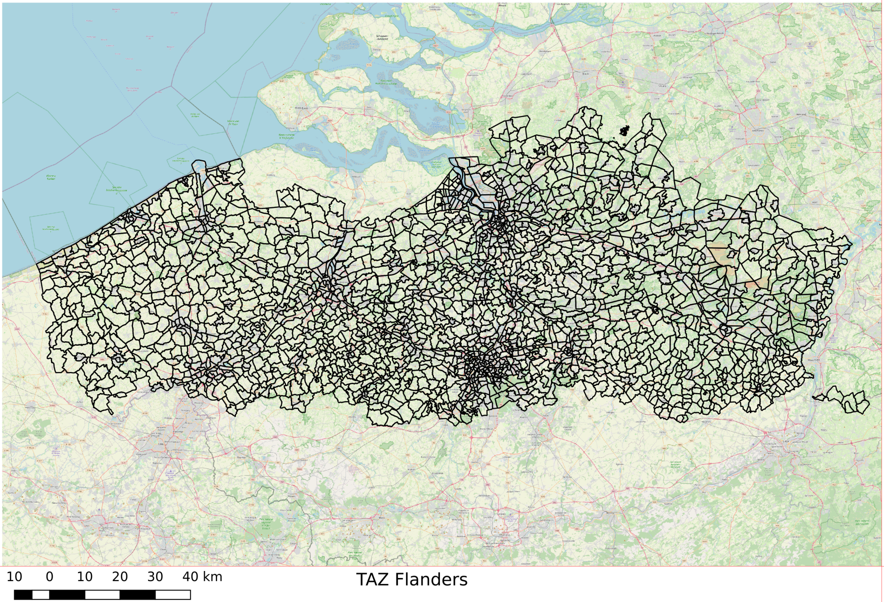
Location Disaggregation: Examples

Flanders

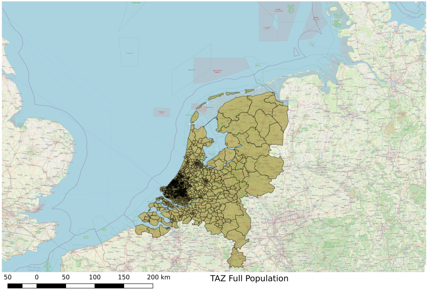
1. TAZ area: more or less homogenous ($5km^2$)
2. no address types in CRAB database

The Netherlands

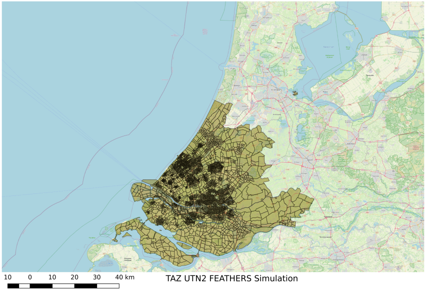
1. TAZ area: wide range $[0.01 \dots 1500]km^2$
2. address labels in BAG database
    1. each address database bears non-empty set of functional labels (*industry*, *meeting*, *residential*, *shopping*, *health*, *sports*, ...)
    2. relation $activTypes \times addrLabels$ to find candidate addresses for activity type



TAZ Flanders

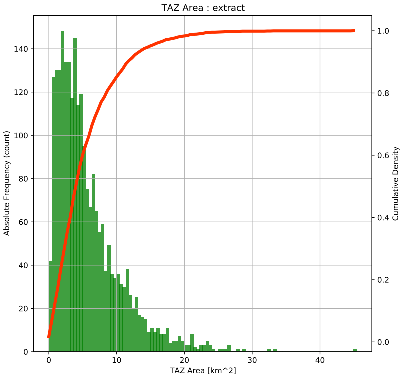


The Netherlands
Nationwide

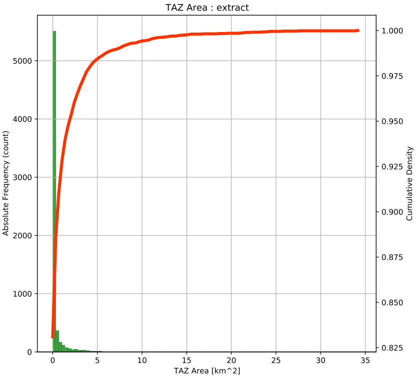


Simulated Rotterdam region
(UTN2 project)

TAZ area distributions: mind rightmost vertical axis ranges.



Flanders



Rotterdam region (UTN2 project)

Location Disaggregation: Problems to solve

Intrazonal trips predicted by FEATHERS

1. FL: intrazonal trip duration $= 5[min]$
2. NL: intrazonal trip duration (distance) is $\langle TAZ, mode \rangle$ specific
3. intrazonal trip duration (distance) *not* <u>activity</u> or <u>individual</u> specific

Tour Delimiting

1. ambiguity: how to disaggregate *revisited location* into *position*(s) ?
2. e.g. *pickUp/dropOff* vs. *shopping*

Location Disaggregation: Choice models

Required

1. Two level choice modeling

   ActivBM : *location* choice at TAZ resolution

   disAggregation : *position* choice at street address resolution

2. Challenges
   1. keep the models compatible
   2. independent sampling *within travel plan* is *behaviourally inaccurate*

Actual practice

1. Nearly all research papers mention
   1. *independent* position *sampling*
   2. using *uniform* distributions
2. In general: causes unrealistic schedules (travel plans)
3. Suitable only for fine grained homogeneous TAZ area distribution

Location Disaggregation: Proposed method

Position type ⇔ *Activity type* in a schedule (travel plan)

stable : same for each activity of given type (*work*, *school*, . . .)

volatile : discretionary (*shopping*, *services*, . . .)

Concept

1. Assign *position* to each stable *location* in schedule
2. Keep the predicted travel modes
3. For each tour: sample N tour alternatives (by sampling *position* for each volatile *location*)
4. Select the alternative with smallest total travel duration (distance) using fast coarse euclidean distance based estimate

Location Disaggregation: Determining N for tour T

$nAct(T)$	number of activities in tour T
A_i	i-th activity in tour
$Z(A_i)$	TAZ for i-th activity in tour
$nAddr(Z(A_i), A_i)$	number of addresses in TAZ $Z(A_i)$ for activity type A_i

$$nComb = \prod_{i \in [1, nAct(T)]} nAddr((Z(A_i), A_i)) \tag{1}$$

$$N = min(200, nComb/10) \tag{2}$$

1. The number 300 is somewhat arbitrarily chosen (function of computational power), but has desirable probability of selecting a 'good' tour for realistic values of *nComb*.
2. $nComb/10$ is used to avoid systematic selection of shortest option when insufficient alternatives are available.

Location Disaggregation: Proposed method

```
for all tp ∈ TravelPlans do
    assign position to each stable location in tp
    for all tour ∈ tp do
        selectedOption ← null
        for i ∈ [0, N − 1] do
            alt ← tour
            for all loc ∈ alt do
                if loc.isVolatile() then
                    loc ← uniformly sampled from suitable positions
                end if
            end for
            selectedOption ← keepShortest(selectedOption, alt)
        end for
        tour ← selectedOption          ▷ all volatile locations substituted in tour
    end for                        ▷ all volatile locations substituted in travel plan
end for                                        ▷ all travel plans processed
```

ANN and MATSim: Alternative generation and training

Alternative generation

1. Using Conveyal R5[1], trip alternatives for each modality will be generated for each trip in the output of MATSim simulation that uses the *plans.xml* file.

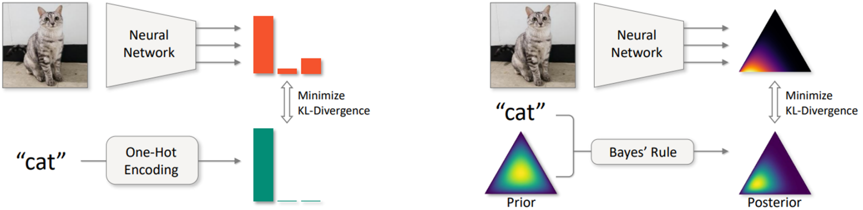
Training

1. A choice model will be estimated by training an Artificial Neural Network trained with belief matching loss on the generated trip alternatives, using the chosen trip mode in MATSim as *ground truth*

[1] https://github.com/conveyal/r5

ANN and MATSim: Loss function

Joo, T., Chung, U., and Seo, M.-G. (2020). Being bayesian about categorical probability. In *International Conference on Machine Learning*, pages 4950–4961. PMLR.



(a) Softmax cross-entropy loss

(b) Belief matching framework

$$\psi(\alpha_y^{\mathbf{w}}(\mathbf{x},\mathbf{s})) - \psi(\alpha_0^{\mathbf{w}}(\mathbf{x},\mathbf{s})) - \lambda(log(\frac{\Gamma(\alpha_0^{\mathbf{w}}(\mathbf{x},\mathbf{s}))\Pi_k\Gamma(\beta_k)}{\Pi_k\Gamma(\alpha_k^{\mathbf{w}}(\mathbf{x},\mathbf{s}))\Gamma(\beta_0)}) + \sum(\alpha_k^{\mathbf{w}}(\mathbf{x},\mathbf{s}) - \beta_k)(\psi(\alpha_k^{\mathbf{w}}(\mathbf{x},\mathbf{s})) - \psi(\alpha_0^{\mathbf{w}}(\mathbf{x},\mathbf{s}))))$$

ANN and MATSim: Comparing validation performance

Validation performance will be computed on:

1. Withheld validation partition of generated alternatives from MATSim output plans.
2. Generated alternatives based on trips matched using tracking data from the Amsterdam region collected from an app.

Training and validation metrics can be compared:

1. Comparing training performance and validation performance on generated alternatives from MATSim output will yield an indication of the ability of the choice model to generalize
2. Comparing validation performances on MATSim vs real world data will yield an indication of the ability of the MATSim model to capture real world mode choice dynamics