

Research plan for using SPALDIA to develop ABM models for exposure assessment

Meng Lu

General plan

The SPALDIA cohort will be used for model development, activity pattern discovery, testing, and the scalability of the developed method. From the ABM (agent-based modelling) developed using SPALDIA, we will gradually add in more assumptions, and derive patterns and relationships to scale it to data-sparse situations.

Specifically, the following goals are to be achieved:

- A. Scale the ABM model developed using SPALDIA to the national level exposure assessment.
- B. Develop a process to quantify uncertainties from human activity modelling.
- C. Compare exposure assessed with ABM vs. time geography models (e.g. distance decay).

For A, we plan to Find relationships between occupation, work address, education, ethnicity vs. activity patterns (travel duration and traffic model), model them (e.g. time geography, linear regression model). Then make assumptions, and use the models identified to replace the information we know and test the mode. The model we developed is probabilistic, for B. For example,

- We assume to find strong relationship between distance to travel vs. travel mode, and the relationship can be modelled (e.g. the probability of travelling by car increases with the distance).
- Transportation studies have found strong relationship between occupation and ethnicity vs. travel mode, e.g. most Dutch students travel by bike or train. We use these relationships to assign activities to each population subgroup (e.g. socioeconomical group).

For B, as information becomes sparse, we expect higher uncertainty, and it is important to provide uncertainty measures to the ABM model and the model indicating relationships found in A. As mentioned in A, this will be achieved through developing probabilistic models and use the SPALDIA dataset to validate our model.

For C, we will compare models developed using SPALDIA, using ABM, time geography functions, and hybrid models.

At last, with the study matures, we plan to enable ABM modelling of human-environment-health interactions.

Detailed plan of ABM

Four activity models are developed, the core is to estimate for each person an activity schedule and according to it the continuous daily geographical locations. Montecarlo simulation is used to quantify uncertainty of each model.

- **Model 1: ABM- survey:**

Activity model based on detailed activity information provided in the SPALDIA.

Known variables:

Days per week of cycling, vehicles, walking. (road_act, road_act_time)

Primary commuting tools to work, if public transport, the time needed to the public transportation. (comm_tool, time2pub)

How many hours per day and days per week at work place. (work_hour, work_day)

10pm – 6 am activity (one hour somewhere?) (night_act)

Occupation (outdoor, indoor)

Work location (work_loc) Home location (home_loc)

Uncertainty sources:

Time of activities,

Tour types (e.g. home-work-home, home-work-activity-home)

Routes

Objectives:

1) This model is used to assess exposure for the **SPALDIA** cohort, as everyone in the cohort has activity information, imputation is not needed (this is the reason that our study is not based on purely activity/transportation surveys.

2) This model is used as a base model for the subsequent model designed with less known variables, to assess accuracy of subsequent models when detailed info are not available.

- **Model 2-1 and 2-2: ABM-generalised**

These two models are research models, to provide theoretical support for modelling information-sparse situations (SNC, EPIC-NL).

Model 2-1:

This model is used to understand how occupation and socioeconomic variables relate to the activity patterns and classify activity patterns for each subpopulation group, the subpopulation groups are

1) defined by statistical relationships between characteristics (e.g. occupation, age, income, number of children, sex) and activities (travel mode, walk time)

2) predefined (e.g. epidemiologically)

Statistical modelling will be applied to SPALDIA, activity surveys (if applicable), and results from existing literatures (possibly as a prior), to find the relationships. For each socioeconomical group, an activity scheduler is going to be designed. People in SCN and EPIC-NL are classified into these socioeconomical groups and the activities are assigned to them.

Model 2-2:

This model is compared to model 1, to understand the accuracy that can be reached with reduced information (e.g. unknown travel mode).

- **Model 3: ABM-SNC**

*in this model: home and work locations (need to be checked) are known, compared to the earlier Utrecht study the occupation and socioeconomical status are known. The activities are assigned to each socioeconomical groups according to model 2. This model is used to assess exposure for the **SNC** cohort.*

- **Model 4: ABM-EPIC**

*in this model: only home location is known, this model is used to assess exposure for the **EPIC-NL** cohort.*

Design of model 1:

(Known) required variables:

Days per week of cycling, vehicles, walking. (road_act, road_act_time)

Primary commuting tools to work, if public transport, the time needed to the public transportation. (comm_tool, time2pub)

How many hours per day and days per week at work place. (work_hour, work_day)

10pm – 6 am activity (one hour somewhere?) (night_act)

Occupation (outdoor, indoor)

Work location (work_loc)

Home location (home_loc)

Literature:

This model identifies human activities from the survey of SPALDIA. The activity patterns identified are assigned to sub-population groups (e.g. socioeconomic groups). The design is based on research from transportation science. Related survey-based activity scheduling patterns include:

- 1) Econometric models, which impute personal day activities (Bowman 2000);
- 2) Sequence Alignment model (SAMs), which studies the sequential dependencies between daily activities, and attempt to classify activity-chains into clusters based on their sequencing characteristics and composition. Profile Hidden Markov Models (pHMM) are developed to capture irregular activity patterns;

Probabilistic activity scheduler

Analyse the activity scheduling patterns and the probability that the schedule occurs.

The probability that a schedule occurs is determined by the probability that a **pattern** is chosen and the probability that the **tours** are conditional on patterns.

$$P(\text{schedule}) = P(\text{pattern})P(\text{tours}|\text{pattern}) \text{ (Bowman 2000)}$$

Suppose $P(\text{schedule})$ is 60%, then 60% this schedule is chosen. There will be multiple schedules combining activities, tour types, frequencies and purposes of tours, as described below.

A **pattern** is a function of the attributes of all its available tour alternatives, is characterised by the

- 1) primary activity,
- 2) the **type of tour** for the day's primary activity,
- 3) the number and purpose of secondary tours.

In this study, the pattern can be classified based on the socioeconomic groups.

A **tour** has a base point that a person will start the trip from and return to, it can consists subtours, e.g. break at work. The tour model is hierarchical:

Primary tour activity time – primary activity modes and destinations - secondary activity time – secondary activity modes and destinations

With this hierarchy, we introduced space-time constraints, e.g. going back to home from work will not happen in the morning. And a secondary tour (e.g. doing sports), happen after or before the primary tour (e.g. to work and back).

The secondary tour is conditioned on the primary tour

$$P(\text{tours}|\text{pattern}) = p(\text{primary tours}|\text{pattern})p(\text{secondary tours}|\text{pattern})$$

$$P(\text{tour}) = p(\text{timing})p(\text{mode, destination}|\text{timing})$$

A tour has a tour type, it is defined by the number, purpose and sequence of activity stops on the tour. For example: (1) home->work->home, (2) home->work->home->additional activity.

All the tours are assigned a **priority**. The tour contains highest priority activity will be given the highest priority (highest probability).

It is therefore essential to define the primary activity, in this study, it is determined by activity priority, which is determined from the SPALDIA.

The **activities are classified based on primary tour type, (purposes), and frequencies**. The purposes are classified to

- 1) Tight schedule constraints include work, work related, school and banking,
- 2) unconstrained. The frequencies are the number of secondary tours.

Tour destinations and mode choice model. From SPADIA or literature.

Example:

```
Primary_tourtype=["home2work2home","home2activity2home",  
"home2work2home2activity", "home2school2home2activity2activity" ]
```

```
constrained_activity=["home", "work", "work related", "shopping"]  
unconstrained_activity=["sports"]  
number_purpose_secondary_activity["1_unconstrained", "2_constrained"]  
Primary_activity=["Home", "work"]  
activity_priority={"home":0.6, "work":0.3, "work related":0.05, "shopping":0.05, "sports":  
0.05}
```

```
Pattern = {[busyworker:  
primary_activity = "work",  
type_of_tour = "home2work2home"  
number_purpose_secondary_activity = 1_unconstrained],  
  
[student:  
primary_activity = "school",  
type_of_tour = "home2school2home2activity2activity",  
number_purpose_secondary_activity = 2_unconstrained]}
```

```
tour_priority = f(tourtype, activity_priority)
```

For a person A

#Assume A is a student and 80% he will follow the student pattern.

```
P(pattern) = 80%  
Pattern_ = Pattern["student"]  
Tour_type, tour_primary, tour_secondary = retrieve_from (Pattern_)
```

```
P(primary_tour|pattern) = f(activity_priority, tourtype, timming, mode)  
P(secondary_tour|pattern) = f(activity_priority, tourtype, timming, mode)
```

```
P(tours|pattern) = p(primary tours|pattern)p(secondary tours|pattern)  
#P(tour) = p(timming)p(mode, destination|timing)  
P(schedule) = P(pattern)P(tours|pattern)
```

Design of Model 2-1:

Linear and non-linear regression analysis is used to identify the relationship between a person's socioeconomical status and his/her daily activity patterns identified in **Model 1**.

Design of Model 2-2:

Applying the model designed in Model 1 to gradually reduced information: for example, the first model has information of

- 1) On average, time spent traveling to and from the workplace each day.
- 2) 10pm – 6 am activity (one hour somewhere?)
- 3) Occupation (outdoor, indoor)

4) Work location

The subsequent models will not use 1), and then 1) and 2) but with MonteCarlo simulations, the priors can be identified from SPALDIA or existing literature.

Design of Model 3 and 4:

Spatial activity: a distance decay model, or an activity model that is heavily based on assumptions based on model 2, will be used to identify the probability that a person is at a certain location.

Temporal activity: with model 2, we expect to cluster activities as figure 1. A statistical model can be used to assign probabilities that a person is at a certain location at a certain time of the day.

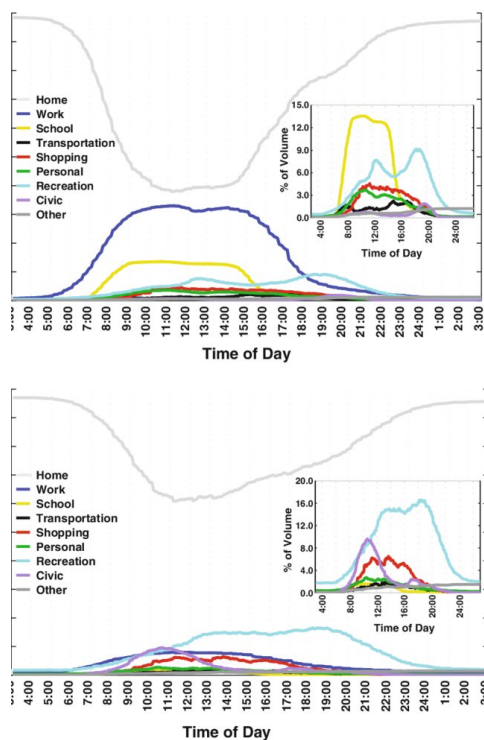


Figure 1: top: weekday activity, bottom, weekend activity. Source: (Jiang et al. Clustering daily patterns of human activities in the city)

Required variables from SPALDIA

The variables below are useful for simulating or identifying

For model 1:

traffic mode and commuting duration:

- days/week of cycling for transport
- min/day of cycling for transport truncated to 180 min/day
- min/week of cycling for transport based on truncated daily minutes (comm_cycl_dpw_new2 * comm_cycl_minpd_trunc)
- sum of all days/week in transport-related activities (comm_cycl_dpw_new2 + comm_walk_dpw_new2)

- days/week of motorized transport
- min/day of motorized transport
- min/week of motorized transport (days/week of motorized transport * min/day of motorized transport)
- days/week of walking for transport
- min/day of walking for transport truncated to 180 min/day
- min/week of walking for transport based on truncated daily minutes (comm_walk_dpw_new2 * comm_walk_dpw_new2)
- Length of optimal route from subject's address to nearest bus stop [m]
- Length of optimal route from subject's address to nearest train/tram station [m]
- On average, how much time do you spend traveling to and from your primary workplace each day?
- How do you primarily commute to your work place?
- How do you primarily commute to your work place? - other, which one
- On average, how much time do you spend traveling to and from your workplace each day (total for all workplaces)

For identifying primary activity:

- On average, how many hours do you usually spend at your primary workplace?
- On average, how many hours do you usually spend at your workplace?
- Which professional position do you have
- How many days per week do you go to your workplace?
- How many hours per week do you usually work?
- Do you still work regularly?

For tour types:

- Which professional position do you have
- sum of all days/week in walking activities
- sum of all min/week in walking activities

For secondary tours:

- Which means of commuting do you use regularly (e.g. to work, for shopping, going out)?
- at least 1 hour work inside or outside at night (10pm-6am)
- How is your usual work rhythm best described? This is for your actual main activity or, in case you are not working, your usual work rhythm
Nighttime work
- How is your usual work rhythm best described? This is for your actual main activity or, in case you are not working, your usual work rhythm
Shift work
- How is your usual work rhythm best described? This is for your actual main activity or, in case you are not working, your usual work rhythm
Daytime work

For separating between seasons (not needed in the initial model)

- percent time spent in commute on non-working day in summer
- percent time spent in commute on non-working day in summer after imputation
- percent time spent in commute on non-workday in winter
- percent time spent in commute on non-workday in winter after imputation
- percent time spent in commute on workday in summer
- percent time spent in commute on workday in summer after imputation

- percent time spent in commute on workday in winter
- percent time spent in commute on workday in winter after imputation

For model 2: (socioeconomical variables)

- would you describe your ethnicity as non-European?
- What country were you born in?
- As these days civil status is not necessarily equivalent to the form of living, may we ask you to state
- civil status harmonised over the 3 surveys with 4 answer categories
- Highest educational level, combination Sap1 & Sap2
- employment category 5cat
- employment category 4cat
- employment category 4cat of partner
- What describes best your current work situation?
- What describes best the current work situation or your partner? If you don't live in a relationship, p
- Which professional position do you have? (if you are not working anymore, please give the last pos
- Which professional position does your mother have? (if she is not working anymore, please give th
- Which professional position does your father have? (if he is not working anymore, please give the l
- Which professional position does your partner have? (if she/he is not working anymore, please give

Additionally, for corresponding activity schedules to locations:

- Walkability index at 1x1km resolution at the SAPALDIA extent
- Walkability index at 1x1km resolution at the Swiss extent

Exposure assessment:

According to the activities, a person's geographical location is identified (e.g. with known home, work, activity locations), predicted (e.g. activity location is close to home), or simulated (e.g. unknown work locations). The last step is to calculate exposure. Exposures are calculated at commuting time (exp_route), working time (exp_work), off-work time (home and leisure) (exp_off)

air pollution: airpol

Air pollution is calculated for each hour of the season or year.

1. Work day exposure

Work day exposure:

- Loop over: each person in the cohort
- Loop over: each hour

Route exposure: exp_route

if comm_tool== 'walk' #primary commuting to work tools is walking
 exp_route = airpol (on route) * time spent on route
 (if the time is not available, use length/ average speed)


```

elif comm_tool == 'car'
    exp_route = airpol * time spent on route * car-filter

elif comm_tool == 'public'
    exp_route = airpol * time2pub + airpol * time on bus * bus-filter.

```

Work exposure: exp_work

```

If occupation == 'indoor'
    Exp_work = airpol * work_hour * indoor_filter
Else
    Exp_work = airpol in working area * work_hour

```

Off-work exposure: exp_off

```

If night_act = T
    Exp_off = airpol * 1 + airpol * (home_hour-2)* indoor_filter + buffered airpol
around home *1
Else
    Exp_off = airpol*(home_hour-1)* indoor_filter + buffered airpol around home *1
Total exposure = exp_work + exp_route + exp_off

```

2. Weekend exposure: purely simulation, same for all the models (is it necessary ?)