Southeast University
School of Transportation

Urban Transportation Planning

Chinese-English course (2019)

Lecturer: Dr. Qiong Bao

Email: baoqiong@seu.edu.cn

Phone: 17351788445

9:50AM, Sunday, 5th May

School of Transportation | Southeast University of China | Campus Jiulonghu | Building Jizhong Y311

Lecture schedule

Lecture	Week	Date/Time	Topic
1	9	28 April 9：50-12：15	Transportation planning & demand and supply & trip-based model
2	10	5 May 9：50-12：15	ABM: data process
3	11	10 May 9：50-12：15	ABM: scheduling
4	12	17 May 9：50-12：15	ABM: uncertainty analysis
5	13	24 May 9：50-12：15	ABM: sensitivity analysis
6	14	31 May 9：50-12：15	Project Evaluation I
7	15	7 June 9：50-12：15	Festival
8	16	14 June 9：50-12：15	Project Evaluation II

Transportation Modelling

- Since transportation problems are multi-dimensional,

  (e.g., Traffic congestion, Emissions, Impact on economy, Traffic accidents)

- and the need for transportation infrastructure is high, due to:

  Globalization, Urbanization,

  Governments cannot afford transportation constraints to have a great negative impact on future competitiveness, foreign investments,…

- Moreover, changing the existing infrastructure is:

  Expensive and have significant long-term effects;

  No guarantee for success;

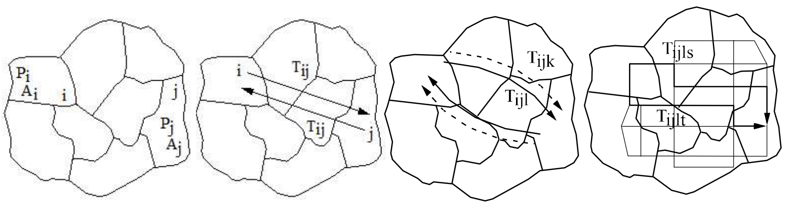
  Not trivial (existing spatial zones, restricted by local regulations, legislation, etc.)

Transportation Modelling

- Therefore transportation models are often used. By using transportation models to represent human behavior, we can:
  - understand individual's travel behavior
  - support management decision making
  - make predictions in uncertain circumstances
  - forecast traffic trend for new or expanded transportation system
  - test scenarios (e.g., alternative land uses, policies)
  - influence future decision-making
- AIM: to portray reality as accurate as possible.
- Commonly used approaches in different countries :
  - Trip-based approach
  - Tour-based approach
  - Activity-based approach

Review of the four step model

- Step1 Trip Generation
- Step2 Trip Distribution
- Step3 Mode Choice (Modal Split)
- Step4 Traffic Assignment (Route Choice)



Trip Generation Trip Distribution Mode Choice Traffic Assignment

Questions: could answer

- How many trips are produced by each household?
- How many trips are attracted to each zone?
- What is the distribution of trip lengths by purpose?
- Where are the destinations of different trip purposes?
- What mode was used for each individual trip?
- How is the traffic flow affected by cost and level of service?
- How are trips distributed to the network?

Questions: couldn't answer

- How complex is the travel? Are there any stops? Before or after the main destination?
- Where are the intermediate stop destinations?
- What mode (combinations of modes) was used to visit intermediate stops?
- How do major improvements affect induced travel?
- What factors affect time-of-day travel decisions?

Limitations of the trip based models

- With the lack of integration of land use forecasting models in the conventional models, the overall travel demand is fixed and essentially independent of the future transportation network;
- Focus on individual trips, ignoring the spatial and temporal interrelationship between all trips and activities comprising the individual activity pattern;
- Ignorance of travel as a demand derived from activity participation decisions;
- The construction of models based strictly on the concept of utility maximization, neglecting substantial evidence relative to alternative decision strategies, e.g. habit formation, choice complexity.

Requirement

- Need a policy-sensitive model
  - Range of transportation policies under study:
    - Congestion pricing policy (cost of travel)
    - Parking policy (cost of travel)
    - Land use and growth management policy (development of infrastructure)
    - Highway and transit improvements policy (level of service of transportation systems)
- Need a model integrated appraisal for air quality, environmental impact assessment, and induced demand

Motivations for ABM

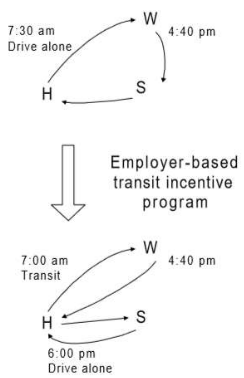
- The motivation for activity based travel demand modeling is that travel decisions are activity based.
- ABM can capture complex travel behavior patterns.
- In the 1970s, ABM emerged based on the theories of Hägerstrand (1970) and Chapin (1974).
  - Hägerstrand focused on personal and social constraints
  - Chapin focused on opportunities and choices
- Policies aimed to affect travel decisions (e.g. tele-commuting program) influence the problem indirectly through the behavior of individuals. Individuals, moreover, adjust their behavior in complex ways.

Motivations for ABM

- The figure illustrates the daily activity and travel pattern of one person.

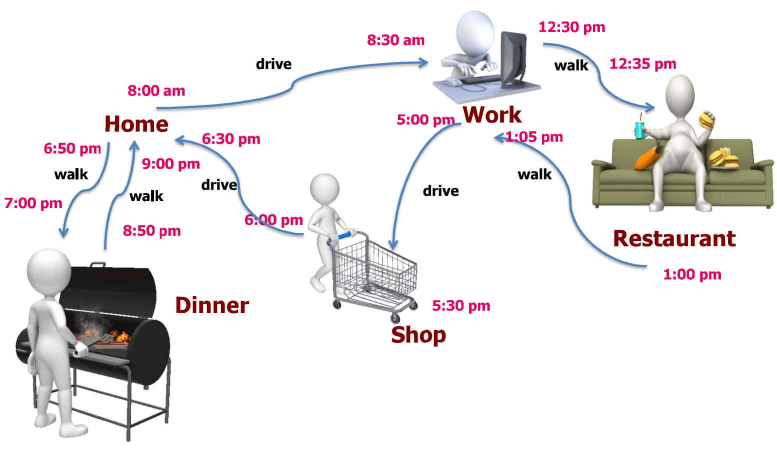
  This traveler switched to transit in response to an employer-sponsored program. The switch required him to begin the commute earlier, at 7:00 a.m., in order to arrive punctually. Because the shopping destination was not on the transit path, he decided to go home after work and shop later.
- This response was rooted in demand of activity, and involved a complex adjustment to the entire day's pattern.
- Models could capture this kind of response only if the models represent how people schedule their daily activities.

Scenario1: Before intervention

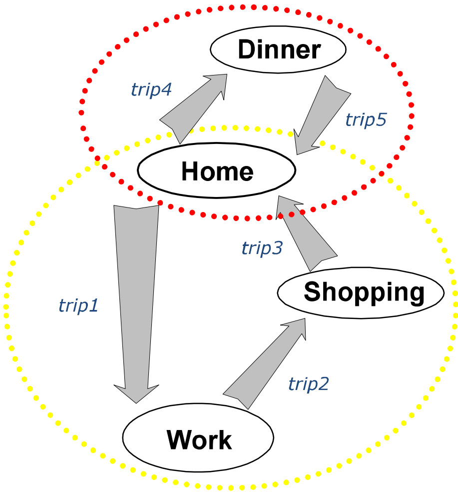


Scenario2: After intervention

Daily travel patterns by an individual



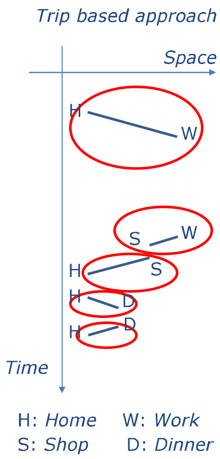
Activities completed in a working day by an individual

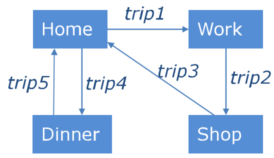


Involve:

- 4 activities
- 5 trips
- 2 tours

Trip based approach

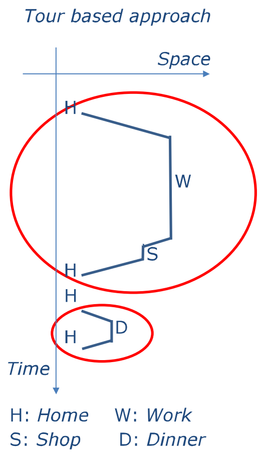


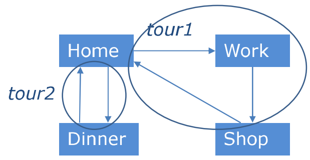


Trip based approach:

- The entities of the model is separate, independent ***trips***;
- Trip is defined as a travel from one location to another and without intermediate stops;
- no connections between different trips;
- no time component;
- no sequential infomation.

Tour based approach

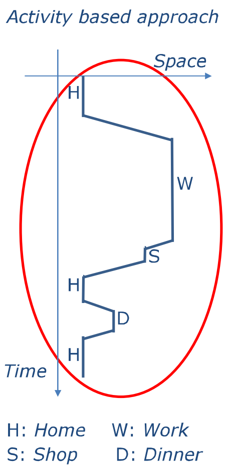


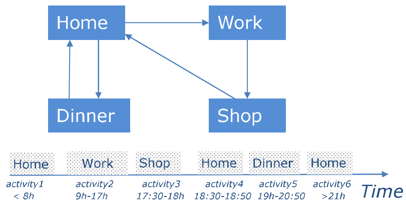


Tour based approach:

- *Tour* is defined as a cycle of trips start and end in the same location;
- In this example, there are two home-based tours;
- In tour-based approach, the temporal and spatial constrains could be represent within a tour;
- but there is no linkage between multiple tours.

Activity based approach





Activity based approach:

- In activity-based approach, the research unit is the sequence of behavior, instead of individual trips;
- Therefore, the full travel pattern of individual could be reflected in temporal and spatial dimensions;
- In this example, there are 6 activities happened.

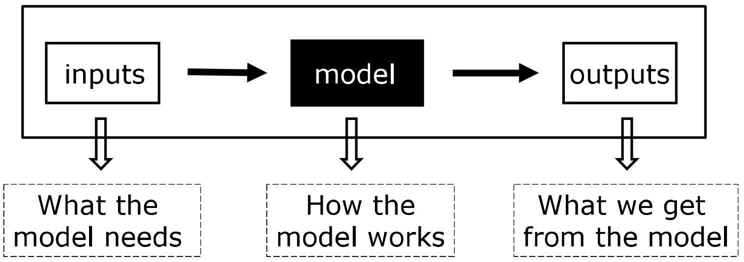
Recall the features of TBM

- Trips are treated independently
  - Departure times
  - Different household activities
  - Locations of subsequent stops in a tour
- All daily trips are "added up" across all households
- Difficult to account for induced travel
- Difficult to support a policy-sensitive analysis

Recall the features of ABM

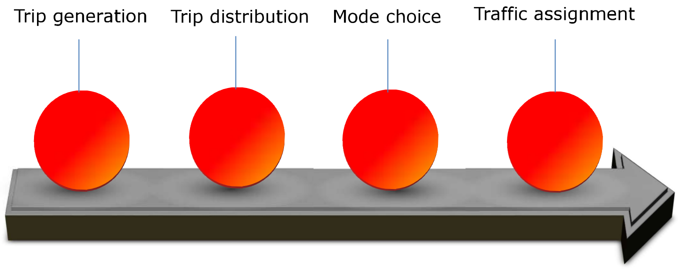
- Travel demand is derived from the activities that individuals need/wish to perform.
- Sequences or patterns of behaviour, and not individual trips are the unit of analysis.
- Household and other social structures influence travel and activity behavior.
- Spatial, temporal, transportation and even interpersonal interdependencies constrain activity/travel behavior.
- Activity-based approaches reflect the scheduling of activities in both the spatial and temporal dimensions.

Activity-based model



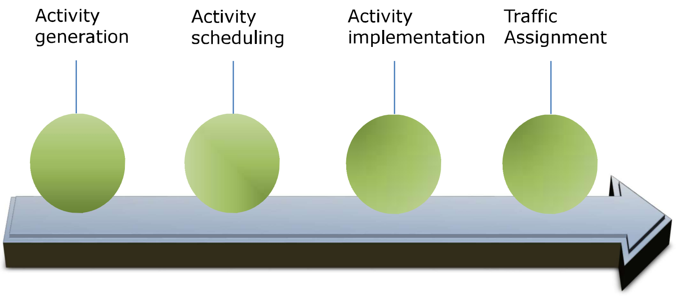
What we get from the TBM?

- Trip-based models



What we get from the ABM?

- Activity-based models



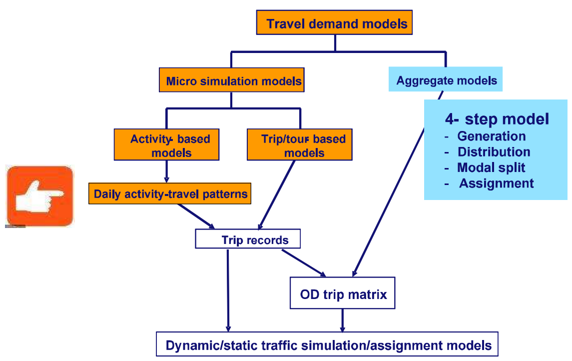
What we get from the ABM?

- ABM is operational at the individual level. A full daily travel pattern of individual describes which activities are conducted, where (location), when (start time), for how long (duration), with whom, the transport mode (mode choice) involved and ideally also the implied route decisions is determined.
- This kind of model therefore not only presents the predicted trips, but also the complete activities taken place during these trips. Therefore, an enriched travel data set can be generated that allows for the assessment of more detailed transportation plans, projects and policies.

Output Variable	Definition	Values
H_hhid	Uniquely defined household ID for population data	1,2, … 4023602
H_hhlocid	Traffic analysis zone (TAZ) ID of the most detailed geographical level	
H_comp	Household composition	0: single, no children, 1: single, with children, 2: single, with parents, 3: double, no children, 4: double, with children
H_income	Household income	0: [0, 1250), 1: [1250, 2250), 2: [2250, 3250), 3: >=3250
H_age	The oldest age in household	0: <35, 1: [35, 55), 2: [55, 65), 3: [65, 75), 4: >=75
H_child	Number of children	0: no children, 1: <6, 2: [6, 12], 3: >12
H_ncar	Number of cars in household	0: no cars, 1: 1car, 2: 2 or more cars
P_pid	Person ID	
P_age	Person age of household member	0: <35, 1: [35, 55), 2: [55, 65), 3: [65, 75), 4: >=75
P_twork	Work status of household member	0: no work, 2: work
P_gend	Gender of household member	0: male, 1: female
P_isDriv	Having driving license	0: no license, 1: with license
A_activityCounter	Activity ID	
A_day	Day of the week	0: Monday … 6: Sunday
A_activityType	Activity type	0: Being at home, 1: Work, 3: Bring/get, 4: Shopping(one), 5: Shopping(mul), 6: Services, 7: Social visits, 8: Leisure, 9: Touring, 10: other
A_beginningTime	Beginning time of each activity	
A_duration	Activity duration	unit: minute
A_location	Location ID	
J_duration	Journey duration	
J_transportMode	Transport mode	1: Car driver, 3: Non-motorised, 4: Public transport, 6: Car passenger
J_distance	Travel distance for each journey	

What we get from the ABM?

- **Trip-based *vs* activity-based**



What the ABM needs?

- The typical input data includes:
  - Population data

age, occupations, income level, vehicle ownership, family size,

  - Land use data

current land use and zoning system data, maps and topography, activity for each area, land use and zoning plans

  - Transportation system information

cost, service levels of public transportation, speeds and travel times, speed limits, parking capacity and cost

  - Observed travel patterns

number of trips and purpose, origins and destinations (OD), transport mode, start time, and time duration, past data for trends

Sources of input data of ABM

- Sources of input data:
  - Census data

(e.g. nr. of employees in the area, nr. of residents)

  - Employment office

(e.g. nr. of employees in the restaurant/post office)

  - Utility company records

(e.g. departure intervals of bus/train, in vehicle time)

  - Studies (observed data)

(e.g. access/egress time, waiting time, max. transfer distance)

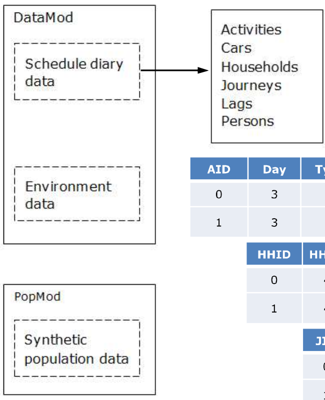
  - Surveys

(e.g. activity types, transport modes, locations of activity)

FEATHERS/ALBATROSS (ABM)

- FEATHERS is one of the operational activity based simulation systems, and was developed by the Transportation Research Institute (IMOB) of Hasselt University of Belgium.
- It's a micro-simulation framework to facilitate the implementation of activity-based models for travel demand forecasting.
- Currently, the embedded scheduler is adapted from the ALBATROSS model (Arentze and Timmermans, 2004).
- A sequence of 26 decision trees is used to simulate how individuals perform their daily activities.

Input data of ABM (ALBATROSS or FEATHERS)



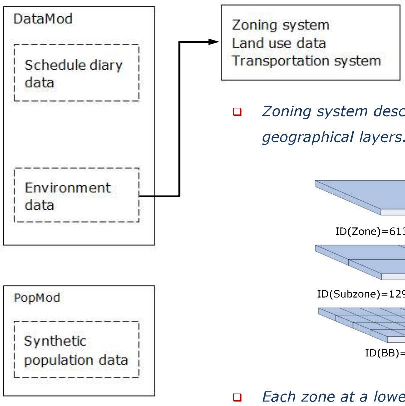
- Activity-based schedule diary data
  - *a trip-based survey*
  - *grouped into six data sets*

AID	Day	Type	BT	Dur	LOC_ID
0	3	0	300	295	4797
1	3	6	815	225	4282

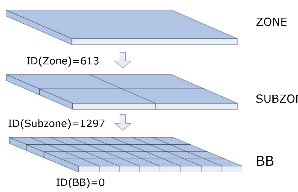
HHID	HHLOCID	Comp	Income	NCar	Npers	Child
0	4797	4	2	1	1	0
1	4794	5	4	2	1	1

JID	PID	Day	BT	Dur	Mode	LOC_O	LOC_D
0	0	3	755	20	1	4797	4282
1	0	3	1200	20	1	4282	4797
2	0	3	1300	10	1	4797	3758

Input data of ABM (ALBATROSS or FEATHERS)



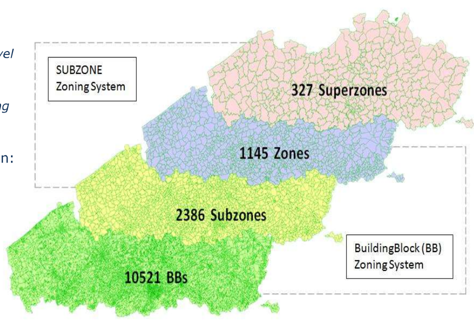
- Environment data: *factors which might potentially influence the travel demand and travel behavior in the study area*
- *Zoning system describes the hierarchical structure of the geographical layers.*



- *Each zone at a lower (disaggregated) level belongs to one and only one zone at a higher level (aggregated).*

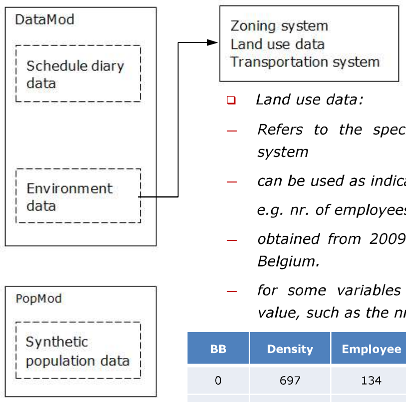
Zoning system of FEATHERS

- Subzone zoning system:
  - Superzone (*municipalities*),
  - Zone (*administrative units at one level lower than municipality*),
  - and Subzone (*virtual areas according to homogeneous characteristics*)
- Most detailed travel demand information: Subzone
- BB zoning system:
  - Zone
  - Subzone
  - and BB (*the most disaggregated geographical level currently applicable in Belgium*)
- Most detailed travel demand information: BB



Geographical Levels of Flanders used in Feathers

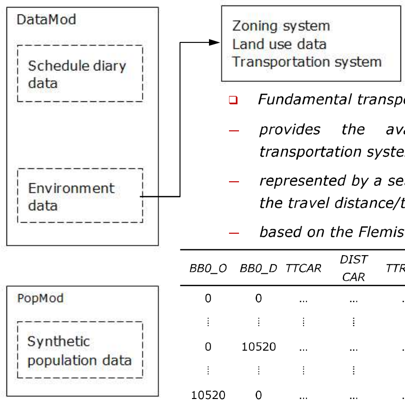
Input data of ABM (ALBATROSS or FEATHERS)



- *Land use data:*
  - *Refers to the specific data at different levels of zoning system*
  - *can be used as indicators for conducting particular activities,*
    e.g. nr. of employees, nr. of school children
  - *obtained from 2009 year based socio-demographic data in Belgium.*
  - *for some variables with no exact value, we use default value, such as the nr. of free parking places.*

BB	Density	Employee	Children	Households	Bank/PostOffice
0	697	134	0	319	22
1	2313	106	130	393	6
2	3160	386	0	597	39

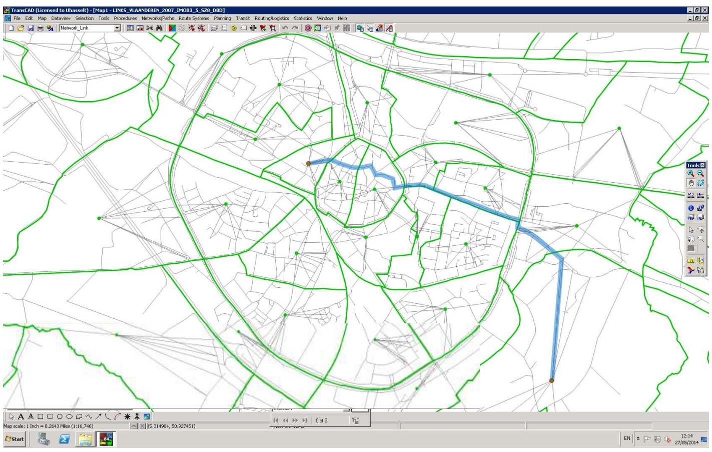
Input data of ABM (ALBATROSS or FEATHERS)



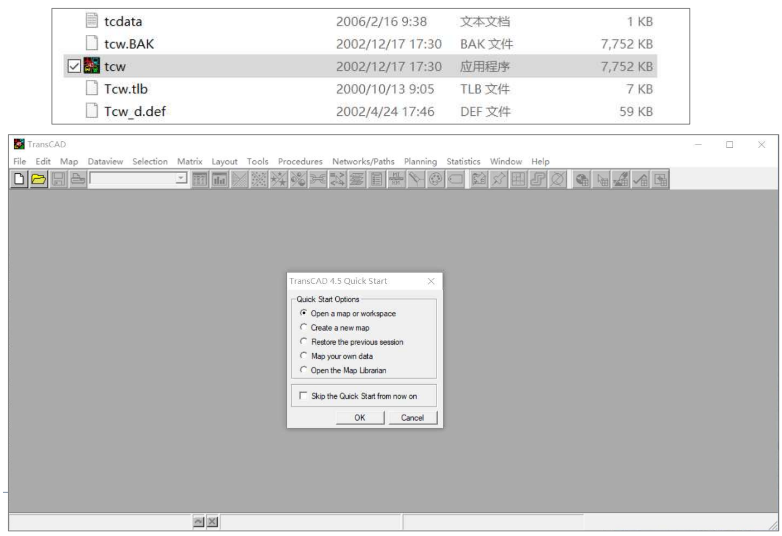
- *Fundamental transportation system data:*
  - *provides the availability and performance of the transportation system between every two zones*
  - *represented by a set of Level of Service (LOS) matrices (e.g. the travel distance/time w.r.t. different transport modes)*
  - *based on the Flemish road and public transport networks*

BB0_O	*BB0_D*	*TTCAR*	*DIST CAR*	*TTRAIN*	*Agress TRAIN*	*DIST TRAIN*	*TBTM*	*Agress BTM*	*COST BTM*	…
0	0	…	…	…	…	…	…	…	…	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	10520	…	…	…	…	…	…	…	…	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10520	0	…	…	…	…	…	…	…	…	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10520	10520	…	…	…	…	…	…	…	…	…

Distance and time (O-D pair)



TransCAD Interface

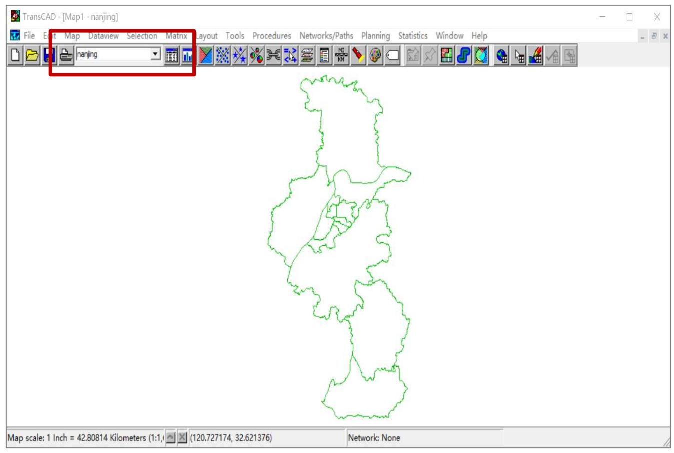


Map and dataview

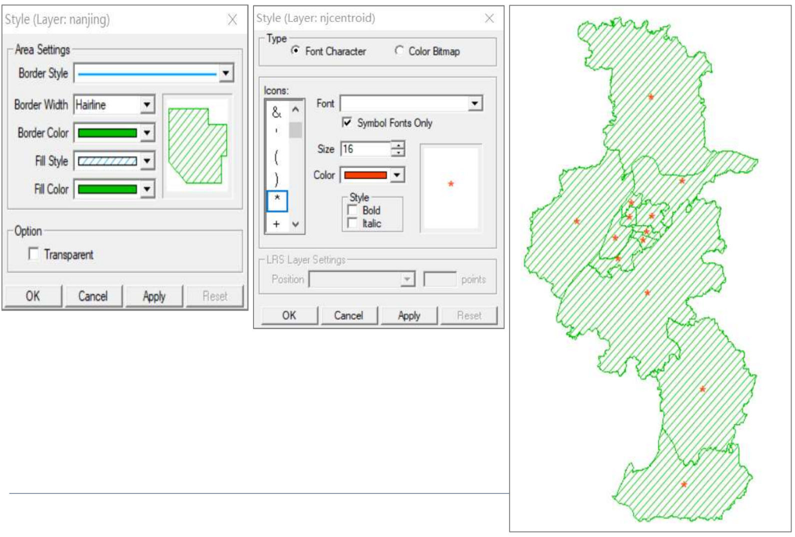

Dataview1 - nanjing

ID	Area	Name	Kind
4	1061.33	溧水县	0137
1	803.98	高淳县	0137
3	1474.71	六合区	0137
13	1624.60	江宁区	0137
2	907.71	浦口区	0137
8	372.76	栖霞区	0137
5	91.59	雨花台区	0137
6	90.95	建邺区	0137
7	24.59	秦淮区	0137
9	27.77	下关区	0137
10	25.81	鼓楼区	0137
11	77.13	玄武区	0137
12	23.96	白下区	0137

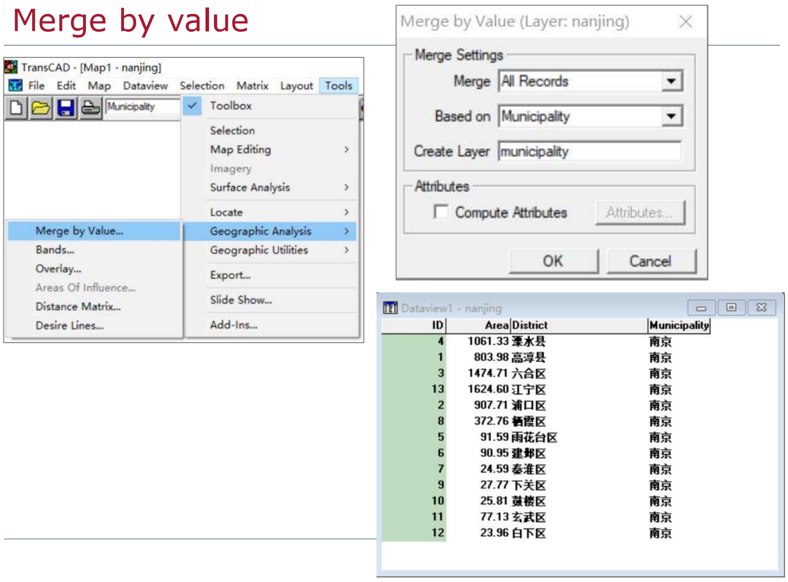
Layers



Layer setting

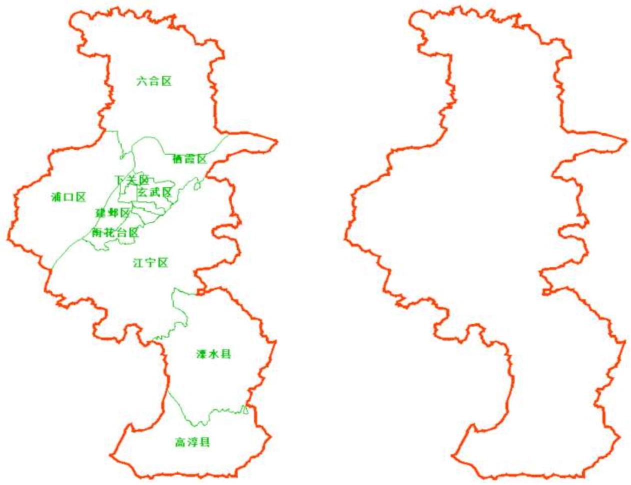


Merge by value



ID	Area	District	Municipality
4	1061.33	溧水县	南京
1	803.98	高淳县	南京
3	1474.71	六合区	南京
13	1624.60	江宁区	南京
2	907.71	浦口区	南京
8	372.76	栖霞区	南京
5	91.59	雨花台区	南京
6	90.95	建邺区	南京
7	24.59	秦淮区	南京
9	27.77	下关区	南京
10	25.81	鼓楼区	南京
11	77.13	玄武区	南京
12	23.96	白下区	南京

Merge by value



Input data of ABM (ALBATROSS or FEATHERS)

DataMod

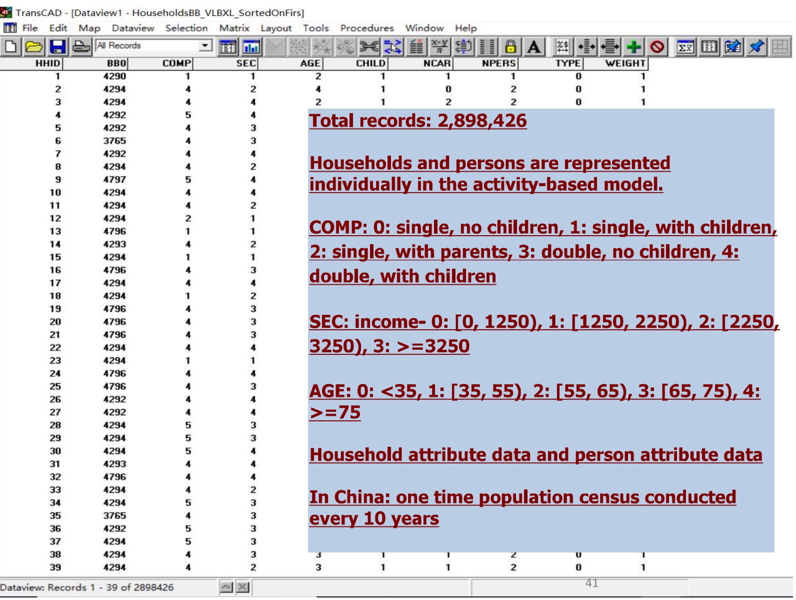
- Schedule diary data
- Environment data

PID	HHID	Age	Work	Gend	Driver
1	1	2	2	2	1
2	2	4	0	1	1

HHID	HHLOCID	Comp	Income	NCar	Npers
1	4290	1	1	1	1
2	4294	4	2	0	2

PopMod

- Synthetic population data → Households data, Persons data



Total records: 2,898,426

Households and persons are represented individually in the activity-based model.

COMP: 0: single, no children, 1: single, with children, 2: single, with parents, 3: double, no children, 4: double, with children

SEC: income- 0: [0, 1250), 1: [1250, 2250), 2: [2250, 3250), 3: >=3250

AGE: 0: <35, 1: [35, 55), 2: [55, 65), 3: [65, 75), 4: >=75

Household attribute data and person attribute data

In China: one time population census conducted every 10 years

Population attribute data

- Population attribute data
  - Due to the privacy protection or limited by the high cost, the attributed population data is not directly available.
- Challenge: How to obtain the household and person attribute data for the entire population in the study region from available data?
- Methods:
  - Population simulator
  - Population synthesis

Both of the above methods could "create" the household and person data which possess all of the demographic attributes needed for model inputs.

Population attribute data

- Objective:
  - to obtain the household and person attribute data for the entire population from the available data

Population attribute data

Sample population
+
Marginal data

→

Population attribute data

Entire population

Population simulator

Population attribute data

Base year data

→

Population attribute data

Target year data

E.g.: 2000 year population attribute data

AGE: 0: <35, 1: [35, 55), 2: [55, 65), 3: [65, 75), 4: >=75

AGE2000
Child: 17-year
Category value: 0

AGE2019
Adult: 36-year
Category value: 1

Population synthesis

- Practical in application:
  - Relatively easy (feasible) to obtain a sample data;
  - Available to obtain the forecasted or projected aggregate distributions of household and person attributes for the population;
- Commonly used approaches:
  - Iterative proportional fitting algorithm (IPF)
  - Iterative proportional updating algorithm (IPU)

Iterative proportional fitting (IPF)

- Deming and Stephan (1941) presented the method to adjust sample frequency tables to match known marginal distributions using a least squares approach.
- Beckman (1996) proposed a method to create synthetic population based on IPF.
  - Joint distribution of household attributes was estimated using IPF.
  - Synthetic households were generated by randomly selecting households from the sample based on estimated joint distributions.
  - Synthetic persons were comprised from the selected households.
  - IPF is one of the most commonly used techniques for generating synthetic population.

Iterative proportional fitting (IPF)

- Input of IPF:
  - Sample data: frequency tables or household travel surveys
  - Marginal distributions from the Census files (base year), population forecasts (future year) are used for controls
- Output of IPF:
  - Joint distribution of household attributes
  - A set of cross tables

IPF example

- 2 attributes: household size, household income

Frequency table (sample)

Marginal distributions

		Household income category		Total	Marginal-size
		High	Low		
Household size category	1	3	1	4	30
	2	2	4	6	40
	3	2	1	3	30
Total		7	6		
Marginal-Income		60	40		

Marginal distributions

IPF example

Iteration1: Adjust for income

		Household income category: High			Household income category: Low			Adjust Total	Marginal-size	Adjust for size
Household size category	1	3	×8.57 =	25.7	1	×6.67 =	6.67	32.37	30	0.93
	2	2		17.14	4		26.67	43.81	40	0.91
	3	2		17.14	1		6.67	23.81	30	1.26
Total		7			6					
Adjust Total		60			40					
Marginal-Income		60			40					
Adjust for income		8.57=60/7			6.67=40/6					

IPF example

Iteration1: Adjust for size

		Household income category: High	Household income category: Low	Adjust Total	Marginal-size	Adjust for size
Household size category	1	23.82	6.18	30	30	0.93
	2	15.65	24.35	40	40	0.91
	3	21.6	8.4	30	30	1.26
Adjust Total		61.08	38.92			
Marginal-Income		60	40			

Iteration2: …

IPF example

Iteration3: c

Frequency table (joint distribution)

		Household income category: High	Household income category: Low	Adjust Total	Marginal-size
Household size category	1	23.57	6.43	30	30
	2	15.17	24.83	40	40
	3	21.29	8.71	30	30
Adjust Total		60.03	39.97		
Marginal-Income		60	40		
Adjust for income		0.9997	1.0005		

Steps of IPF

- Step 1: (Column-adjust) Each column of cells is proportionally adjusted to equal the marginal column totals. Specifically, each cell is divided by the actual sum of the column of cells, then multiplied by the marginal column total.
- Step 2: (Row-adjust) Each row of cells is proportionally adjusted to equal the marginal row totals. Specifically, each cell is divided by the actual sum of the row of cells, then multiplied by the marginal row total. This is the end of the first 'Iteration'.
- Steps 3: The above steps are repeated until the selected level of convergence is reached.

Iterative proportional fitting (IPF)

- Limitations of IPF:
  - The procedure only controls for household attributes and not person attributes.
  - As a result, synthetic populations fail to match given distributions of person characteristics.
  - The method assumes that all households in the sample contributing to a particular household type have the same structure (i.e., a similar individual structure).
  - However, the structure of households even within a same household type is generally different and hence different weights are needed based on household structure.

Iterative proportional updating (IPU)

- The IPU algorithm simultaneously controls for both household and person attributes of interest.
  - Reallocates the weights of the households within a same household type to account for the differences in their household structures.
- Input of IPU:
  - Sample data: frequency tables or household travel surveys
  - Household and person constrains from IPF algorithm
- Output of IPU:
  - Joint distribution of household and person attributes

IPU example

- 2 attributes: household income, person age

Frequency table

HHID	Initial weights	HH Income High	HH Income Low	Person Age<30	Person Age 30-55	Person Age >55
1	1	1	0	1	1	1
2	1	1	0	1	0	1
3	1	1	0	2	1	0
4	1	0	1	1	0	2
5	1	0	1			
6						
7				2	1	2
8	1	0	1	1	1	0
Constraints		35	65	91	65	104
Weighted Sum		3	5	9	7	7

Household constrains -- from IPF using HH attributes

Person constrains -- from IPF using Person attributes

IPU example

Iteration1: Adjust for HH income high

HHID	Initial weight	HH Income High	HH Income Low	Person Age<30	Person Age 30-55	Person Age >55	Weight1
1	1	1	0	1	1	1	11.67
2	1	1	0	1	0	1	11.67
3	1	1	0	2	1	0	11.67
4	1	0	1	1	0	2	1
5	1	0	1	0	2	1	1
6	1	0	1	1	1	0	1
7	1	0	1	2	1	2	1
8	1	0	1	1	1	0	1
Constraints		35	65	91	65	104	
Weighted Sum0		3	5	9	7	7	
Weighted Sum1		35	5	51.67	28.33	28.33	

$W_1 = 11.67 = 35/3$

Weighted Sum1 $= \sum_{i=1}^{8} A_i \times W_i$

IPU example

Iteration1: Adjust for HH income low

HHID	Initial weight	HH Income High	HH Income Low	Person Age<30	Person Age 30-55	Person Age >55	Weight1	Weight2
1	1	1	0	1	1	1	11.67	11.67
2	1	1	0	1	0	1	11.67	11.67
3	1	1	0	2			11.67	11.67
4	1	0	1	1			1	13
5	1	0	1	0			1	13
6	1	0	1					13
7	1	0	1					13
8	1	0	1					13
Constraints		35	65	91	65	104		
Weighted Sum0		3	5	9	7	7		
Weighted Sum1		35	5	51.67	28.33	28.33		
Weighted Sum2		35	65	111.67	88.33	88.33		

$W_2 = 13 = 65/5$

Weighted Sum1 $= \sum_{i=1}^{8} A_i \times W_i$

IPU example

Iteration1: Adjust for HH & P attributes

HHID	Initial weight	HH Income High	HH Income Low	Person Age<30	Person Age 30-55	Person Age >55	Weight 1	Weight 2	Weight 3	Weight 4	Weight 5
1	1	1	0	1	1	1	11.67	11.67	9.51	8.05	12.37
2	1	1	0	1	0	1	11.67	11.67	9.51	9.51	14.61
3	1	1	0	2	1	0	11.67	11.67	9.51	8.05	8.05
4	1	0	1	1	0	2	1	13	10.59	10.59	16.28
5	1	0	1	0	2	1	1	13	13	11.00	16.91
6	1	0	1	1	1	0	1	13	10.59	8.97	8.97
7	1	0	1	2	1	2	1	13	10.59	8.97	13.78
8	1	0	1	1	1	0	1	13	10.59	8.97	8.97
Constraints		35	65	91	65	104					
Weighted Sum0		3	5	9	7	7					
Weighted Sum1		35	5	51.67	28.33	28.33					
Weighted Sum2		35	65	111.67	88.33	88.33					
Weighted Sum3		28.52	55.38	91.00	76.80	74.39					
Weighted Sum4		25.60	48.50	80.11	65.00	67.68					
Weighted Sum5		35.02	64.90	104.84	85.94	104					

IPU example

Iteration401: Adjust for HH and P attributes

HHID	Initial weight	HH Income High	HH Income Low	Person Age<30	Person Age 30-55	Person Age >55	Final IPU Weight
1	1	1	0	1	1	1	1.36
2	1	1	0	1	0	1	25.66
3	1	1	0	2	1	0	7.98
4	1	0	1	1	0	2	27.79
5	1	0	1	0	2	1	18.45
6	1	0	1	1	1	0	8.64
7	1	0	1	2	1	2	1.47
8	1	0	1	1	1	0	8.64
Constraints		35	65	91	65	104	
Weighted Sum0		3	5	9	7	7	
Final Weighted Sum		35	64.99	90.99	64.99	103.99	

Zero-cell problem

- Problem
  - The disaggregate sample for small TAZs might capture infrequent household types.
  - IPF algorithm fails to converge
- Solution
  - Add a small arbitrary number to the zero-cells
  - This procedure introduces an arbitrary bias
  - Borrow the prior information for the zero cells from last layer subject to an limit on the probabilities

Zero-cell problem

Zero-cell frequency table

		Household income category	
		High	Low
Household size category	1	3	0
	2	2	4
	3	2	1
Total			12

Higher level frequency table

		Household income category	
		High	Low
Household size category	1	7	2
	2	8	10
	3	3	3
Total			33

Threshold probability=1/12=0.083

		Household income category	
		High	Low
Household size category	1	0.25	0
	2	0.17	0.33
	3	0.17	0.08

Higher level probability table

		Household income category	
		High	Low
Household size category	1	0.21	0.06
	2	0.24	0.3
	3	0.09	0.09

Zero-cell problem

Zero-cell adjusted Probability table

		Household income category	
		High	Low
Household size category	1	0.25	0.06
	2	0.17	0.33
	3	0.17	0.08

Higher level probability table

		Household income category	
		High	Low
Household size category	1	0.21	0.06
	2	0.24	0.3
	3	0.09	0.09

Sum of probability=1.06>1

Zero-cell adjusted Probability table

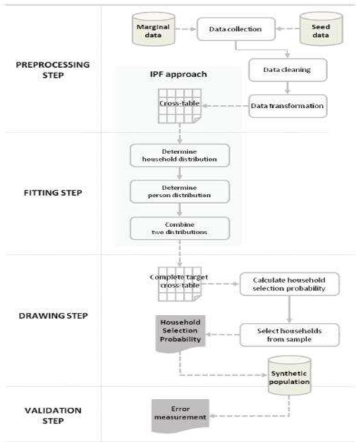
		Household income category	
		High	Low
Household size category	1	0.23	0.06
	2	0.16	0.31
	3	0.16	0.08

Iterative proportional updating (IPU)

- Joint household distribution from IPF gives the frequencies of different household types to be drawn
- IPU
  - IPF frequencies are rounded
  - The difference between the rounded frequency sum and the actual household total is adjusted
  - Households are drawn probabilistically based on IPU estimated weights for each household type.

Steps of generating synthetic population

- There are four main steps in the process of generating synthetic population:
  - Data preprocessing
  - Data fitting
  - Data drawing
  - Data validation



Steps of generating synthetic population

- Data preprocessing
  - To prepare input data for the synthetic population technique
    - Seed data: A sample data set containing disaggregated population with enough details of the demographic attributes needed.
    - Target marginal: Information about the sum of one dimension of an attribute.
  - A set of cross-tables is obtained
    - Consists of rows and columns that represent one dimension with cells and marginal totals of each attribute.

Steps of generating synthetic population

- Data fitting
  - To adjust seed data to target marginal distribution by iteratively re-scaling the value of a cell in the seed cross-table.
    - An IPF fitting process is applied to the household data to make them comply with household marginals.
    - Next, a person distribution is determined in the same way.
    - The two distributions are then combined using a person-per-household ratio at the end of the fitting step.
  - The fitting process can be classified into:
    - Single level fitting vs. Multi-level fitting.
    - Zone-by-zone fitting vs. Multi-zone fitting.
    - Temporal fitting vs. Spatial fitting.

Steps of generating synthetic population

- Data drawing
  - To draw synthetic population by adding individuals/households to the population up to the desired size of the population.
    - Household selection probability
    - Monte-Carlo sampling
- Data validation
  - To check the accuracy of the estimation, the synthetic population needs to be validated against real population using a goodness-of-fit measure.
    - Traditional statistics, e.g., R^2 and chi-square.
    - Information-based statistics, e.g., entropy difference.
    - Distance-based statistics, e.g.,a standardized root mean square error

Example of source data

Example of source data: household data

Household ID	Residence	Income	Size
1	3	1	1
2	15	1	1
3	20	4	2
4			
5			
6			

Example of source data: person data

Person ID	Household ID	Age	Job	Gender	Driver License
1	1	2	2	2	1
2	2	3	0	2	0
3	3	4	0	2	0
4	3	4	0	1	1
5					
6					
7					
8					
9					

Example of target data

Subzone	Households	Persons	Income 1	Income 2	Size 1	Size 2	Age1
0	4138	7207	734	711	1451	2686	782
1	767	1442	136	132	269	498	141
2	1496	2709	265	257	525	972	314
3	8073	9169	1896	1905	3591	4483	1367
Age2	**No Work**	**Work**	**Male**	**Female**	**No Driving License**	**Driving License**	
1117	3031	4177	4052	3155	1688	5519	
202	679	763	751	691	416	1026	
448	1043	1666	1578	1131	566	2144	
1952	4267	4902	5074	4094	2533	6635	

Example of cross-table

Example of (multi-zone) cross-table

zone 1	Gender			
Income	Male	Female	Sum	Target margin
1	311	327	638	766
2	877	498	1375	1925
3				
4				
Sum				
Target margi				

zone 0	Gender			
Income	Male	Female	Sum	Target margin
1	984	501	1485	2228
2	709	531	1240	1860
3	791	634	1425	1995
4	654	961	1615	1938
Sum	3138	2627	5765	
Target margin	4393	3415	8071	

Cons and pros of ABM

- Advantages:
  - Travel is derived from the need to conduct activities
  - Predict where, when, how long, with whom, which route/mode
  - Better capture activity-travel behavior of individual
  - Sensitive to a larger spectrum of policies
- Main limitations:
  - ABM requires more detailed input data.
  - For a sustainable development and also to further improve the prediction accuracy of the model, the input data should be updated as an ongoing process.
  - In practice, the model execution is time-consuming.

Lecture summary

- Activity based model:
  - Difference among the trip-based approach, tour-based approach, and the activity-based approach;
  - The output of ABM;
  - The input of ABM;
    - Population attribute data
      - IPF
      - IPU

Questions

1. What are the main differences between trip based approach, tour based approach and activity based approach?
2. What are the outputs of an activity-based model?
3. IPF calculation

Unit 2 Assignment

- Famous activity based methods:
  - Constraint-based models

    (e.g. CARLA, MASTIC, PCATS)
  - Utility maximizing models

    (e.g. STARTCHILD, DAS, Tel Aviv)
  - Rule-based models (computational process models)

    (e.g. ALBATROSS, FEATHERS, TASHA, ADAPTS)
  - Micro-simulation models (mixed)

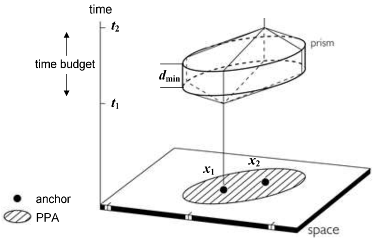
    (e.g. CEMDAP, HAPP, MATSim)

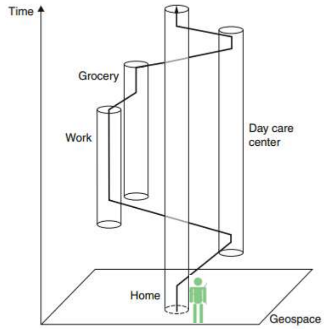
References: Bowman and Ben-Akiva, 2001; Arentze and Timmermans 2004; Miller, 2009; Balmer & Axhausen et al., 2009; Bhat et al., 2012;

Unit 2 Assignment

- Famous activity based methods:
  - Constraint-based models

    e.g. CARLA, MASTIC, PCATS (1970s ~ 2000s)





Unit 2 Assignment

- Famous activity based methods:
  - Utility maximizing model

    e.g. DAS (Bowman and Ben-Akiva, 2001)

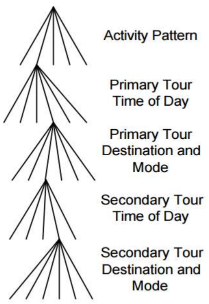


Fig. Activity schedule hierarchy, using nested logit model

Unit 2 Assignment

- Famous activity based methods:
  - Rule-based models

    e.g. ALBATROSS (Arentze and Timmermans, 2004)

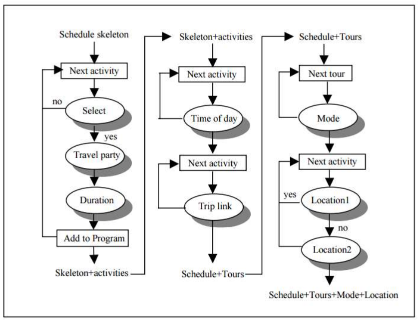


Fig. Activity schedule sequencing, using decision tree classification

Unit 2 Assignment

- Famous activity based methods:
  - Micro-simulation models (mixed)

    e.g. CEMDAP (Bhat), HAPP (Recker), MATSim (Axhausen)
    - CEMDAP: 20+ decomposed econometric models (hierarchical)
    - HAPP: mixed integral programming (simultaneously)
    - MATSim: evolutionary optimization process (simultaneously but)

Unit 2 Assignment

- Requirements:
  - Submit your report before June 7th. As a review of research papers, your report structure should cover the following parts:
    - *introduction (find the very beginning history of ABM),*
    - *summarize the literature's research question and clearly denote the reference paper or website,*
    - *describe the selected ABM (developed by ?, motivated by ?, applied for ? how the model be used?)*
    - *The incorporated methodology (how the methodology apply, why take the methodology into account/advantages of the method)*
    - *an case study is prefer.*
    - *the major findings and the results.*
    - *the reference list*

Thanks for your attention!

baoqiong@seu.edu.cn