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, Friday, 17th May
School of Transportation | Southeast University of China | Campus Jiulonghu | Building Jizhong Y311

Review of data scheduling in ABM

- ABM scheduling process:
  - Scheduling process
  - Methodology
    - Rule based method
    - Decision tree technology
      - Gini Index
      - Information Gain
      - χ^2 contingency table statistic
  - Case study

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

Outline

- Uncertainty analysis:
  - Input uncertainty
    - Coefficient of variation
    - Effects of population fraction size on uncertainty
  - Model uncertainty
    - Stochastic error
      - Confidence Interval
    - Impact of specific zonal characteristics
  - Case study

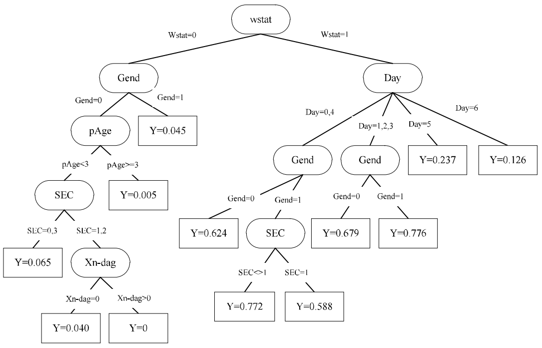
Uncertainty analysis

- Uncertainty is inherently associated with their models. Uncertainty analysis of travel demand forecasting model (e.g., ABM) is concerned with the amount and nature of uncertainty in the outcomes of the model.
- Consider an equation $y = f(x)$ that defines the mathematical relationship f between the derived y and the observed x. Assume this measurement is prone to various sources of error. The relationship between x and its error is given by $x = \tilde{x} + \varepsilon_x$. Where, $\tilde{x}$ is the true value. Uncertainty analysis is concerned with the mapping of the error in x and/or the stochastic nature of f on the uncertainty in outcomes y.

Uncertainty analysis

- Uncertainty analysis plays an important role in the validation of simulation models and interpretation of their results.
- The purpose of uncertainty analysis is to provide uncertainty intervals around the mean estimate of one or more outcomes. In uncertainty analysis, the model analyst attempts to quantify the uncertainty around the outcomes that is propagated through the model from different sources of uncertainty.
- How much we can make sure the output of our model?

DT: concerning work activity choice



Condition variable	Definition
Wstat	Work status
Day	Day of the week
Gend	Gender of individual
pAge	Age of the person
SEC	Income
Xn-dag	Number of employees

There are 6 input condition variables involved in this decision tree and collectively determine whether a work-related activity will be implemented or not.

Target Y: the probability of making the final decision, i.e. implement work activity or not

Sources of uncertainty in ABM

- The sources of uncertainty in ABM include:
  - stochastic error (i.e., random error);
  - parameter uncertainty (i.e., parameters used in simulation models are estimated and not truly known);
  - structural uncertainty (i.e., uncertainty associated with the choice of the models);
  - and possibly other sources (e.g., the choice of the starting population and sources of data that built the model).
- Mainly attributed to two basic sources:

  <u>input uncertainty and model uncertainty</u>

Sources of uncertainty in ABM

- Input uncertainty

  <u>due to the fact that data is not necessarily error-free</u>
  - sampling bias: the travel behavior of the non-response group may significantly differ from the behavior of the respondents.
  - survey design: survey questions (or reported travel patterns) may not sufficiently capture household-level activities or task allocation.
  - reporting errors: resulted from simple mistakes. E.g., respondents may not report certain trips to reduce respondent burden or because they believe the trip is not important for the purpose of the study.
  - coding assumptions: all by definitions are uncertainty, even those attempt to express the most likely future possibility. E.g., the defined average travel speed for train/bus as speed=50,60,70 km/h

Analysis of input uncertainty

- Model forecasts are directly dependent on the values of the input data. Commonly, these data are considered error-free. However, this is an unrealistic assumption.
- Data are sensitive to different types of errors.
- Input uncertainty: different values of the input data, sampled from distributions of measurement error, will lead to variation and, therefore, uncertainty in predicted activity-travel sequences.

Standard deviation

- In statistics, the standard deviation (SD) is a measure used to quantify the amount of variation or dispersion of a set of data values.
  - The formula for the sample standard deviation is:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

    - *A low standard deviation indicates that the data points tend to be close to the mean of the set;*
    - *while a high standard deviation indicates that the data points are spread out over a wider range of values.*

Coefficient of variation

- The coefficient of variation (CV) is defined as the ratio of the standard deviation σ to the mean μ: $CV = \frac{\sigma}{\mu}$
  - The coefficient of variation shows the extent of variability in relation to the mean of the population.
  - In probability theory and statistics, the coefficient of variation, also known as relative standard deviation (RSD), is a standardized measure of dispersion of a probability distribution or frequency distribution.
  - The input (model) uncertainty could be measured by the coefficient of variation.

Coefficient of variation

- The formula for the sample standard deviation is:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

- The coefficient of variation (CV) is defined as the ratio of the standard deviation σ to the mean μ: $CV = \frac{\sigma}{\mu}$

Example:

– DataSet1 = [90, 100, 110]

– DataSet2 = [1, 5, 10]

Coefficient of variation

Which dataset is the one with the least variance?

Standard deviation comparison:

- *DataSet1=10 > DataSet2=4.509*

The coefficient of variation comparison:

- *DataSet1=0.1 < DataSet2=0.845*

DataSet1: less variability;

DataSet2: more variability;

– DataSet2 = [1, 5, 10]

- *Standard deviation is 4.509,*
- *Average is 5.33,*
- *the coefficient of variation of 4.509 / 5.33 = 0.845*

Analysis of input uncertainty

- The analysis concerns the Albatross model system, applied to a 10% fraction of the synthetic population of Rotterdam.
- Effects of input uncertainty:

– The nr. of employees in daily goods retailing

– Travel times in a specific corridor of the transportation network

- *Variability is expected to significantly affect the results*
- *To prevent the effect of model uncertainty in different runs, the random number seed was held constant for different runs.*

- Performance indicators:

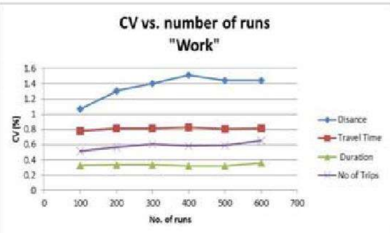
– distance travelled, total number of trips, travel time / (day • person) by activity type, transport mode

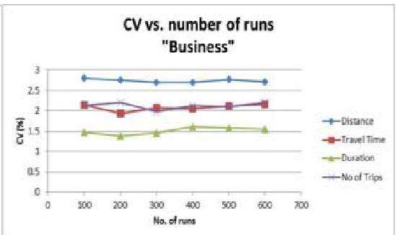
Analysis of input uncertainty

- Steps of experiment:

1. Using Monte Carlo draws, sample a realization of the probability distributions of travel times on the corridor on the transportation network and the employment in each postal code area for all manipulated input data;
2. Run for each sampled individual the agent-based model one time to obtain a value for the considered performance indicators ;
3. Repeat the above steps 600 times, using the same seed points;
4. Calculate the coefficient of variation and its confidence intervals for each performance indicator across the sample.

Input uncertainty by activity types



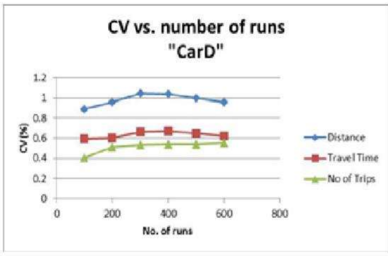


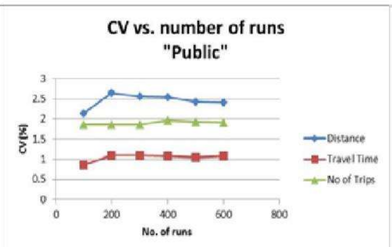
- Input uncertainty on the performance indicators by activities:

– The most uncertain indicator of all activities is distance, followed in most cases by travel time.

– For "Work" activities, distance is the only indicator with a relatively high fluctuation, reaching its maximum coefficient of variation at 400 runs. The remaining indicators are more or less stable.

Input uncertainty by transport modes





- Input uncertainty on the performance indicators by TransMode:

– The input uncertainty of distance, which is presented by blue lines, also shows most fluctuation. The lowest variability pertains to the number of trips.

– Consistent with the segmentation by activity type, distance has the highest uncertainty of all performance indicators.

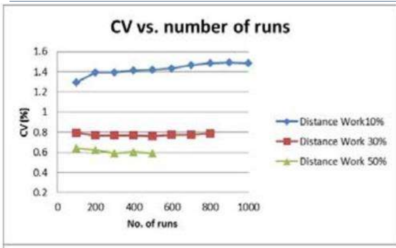
	Dist (10% diff from final mean	Travel time (10% diff from final mean	Duration (10% diff from final mean	No of trips (10% diff from final mean	Dist (5% diff from final mean	Travel time (5% diff from final mean	Duration (5% diff from final mean	No of trips (5% diff from final mean	Dist (2% diff from final mean	Travel time (2% diff from final mean	Duration (2% diff from final mean	No of trips (2% diff from final mean
Total	1	1	1	1	1	1	1	1	1	1	1	1
CarD	1	1	-	1	1	1	-	1	1	1	-	1
Slow	1	1	-	1	1	1	-	1	1	1	-	1
Public	1	1	-	1	1	1	-	1	3	1	-	1
CarP	1	1	-	1	1	1	-	1	20	1	-	1
Work	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Business	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
BrgGet	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Shop1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

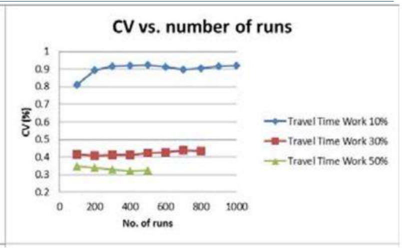
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Total	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CarD	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Slow	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Public	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CarP	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Work	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Business	19	209	207	331	234	490	317	540	506	504	530	560
BrgGet	226	228	174	153	352	240	308	428	574	574	543	446
Shop1	46	227	89	417	448	430	169	453	523	492	407	550
Shopn	225	224	417	233	376	369	499	318	520	561	552	457
Service	127	343	166	259	279	423	388	308	448	561	386	456
Social	36	124	82	325	364	206	339	421	575	514	508	459
Leisure	260	152	131	263	483	164	319	305	536	250	549	337
Tour	53	237	82	171	420	364	263	366	470	433	283	451

Required number of runs to obtain an average/a standard deviation for different indicators with a certain accuracy

- *The required nr. of model runs for a stable standard deviation > that for the average*
- *For the average indicator: accuracy level is increased to a 2%, a major jump occurs*

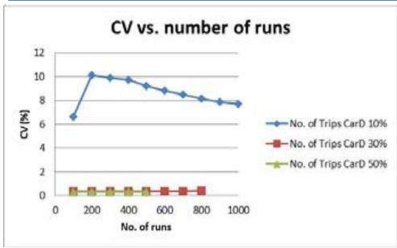
Effects of population fraction size on uncertainty

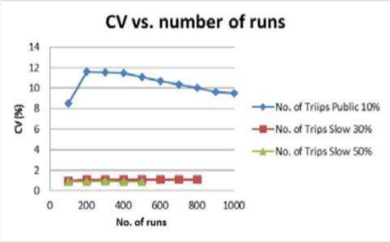




- Uncertainty on the performance indicators by work activity:
  - Uncertainty decreases with increasing fraction size (from 10%-50%)
  - Interestingly, uncertainty in some indicators may increase with an increasing number of model runs. (for the 10% sample fraction)

Effects of population fraction size on uncertainty





- Uncertainty on the performance indicators by TransMods:
  - Uncertainty decreases with increasing fraction size (from 10%-50%)
  - The uncertainty drops to low uncertainty values of around 0.5% for all the performance indicators for sample fractions of 30% and 50%.

Sources of uncertainty in ABM

- Model uncertainty: specification and estimation errors
  - specification errors: results from a failure of identifying the true model.
    - E.g., a simplification of the model: deleting one or more variables will increase model uncertainty from a statistical perspective.
    - In our forecasting, researchers are usually restricted by existing data sources, implying they often can only include variables with available data. Thus, we have to decide some thresholds or the boundaries of our model: some processes are modeled, others are not.
  - estimation error: takes place in estimating the values of various variables and parameters of the model. These models may be inherently probabilistic and the researcher cannot be sure that the true relationship has been depicted.

Stochastic error in ABM

- The stochastic error effects inherently included in the ABM due to using of micro-simulation approach.
  - Activity-based models focus on activity-travel generation and activity scheduling decisions, in which heterogeneity and randomness are fundamental characteristics.
  - Since they simulate individual activity patterns by drawing randomly from marginal and conditional probability distributions that are defined for the various choice facets that make up an activity pattern.
  - As a result, running a transport micro-simulation model several times with the same input will generate different outputs due to the random number seed used in each run.

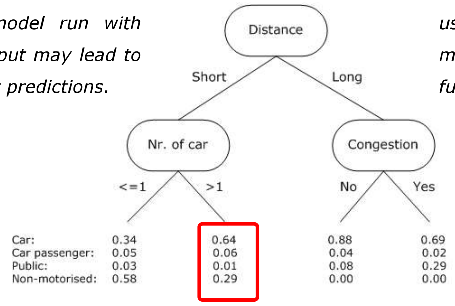
Stochastic errors in Feathers

- Micro-simulation in FEATHERS framework:
  - performed by predicting the activities and travel choices of each individual via selecting the activity type, duration, start time, location and transport mode from a sequence of 26 decision trees.
  - The DT technology assumes the choice is driven by a set of rules.
  - These rules predict a probability distribution across choice alternatives rather than an all-or-nothing decision.
  - Therefore, the output of decision tree represents the probability of each choice alternative.
  - The final choice is determined by randomly selecting an alternative from the probability array.

Stochastic errors reflected in our model

Remark:

every model run with same input may lead to different predictions.



Solution:

use the average value of multiple predictions for further analysis.

The output of DT represents the probability of each choice alternative.

(64%, 6%, 1%, 29%)

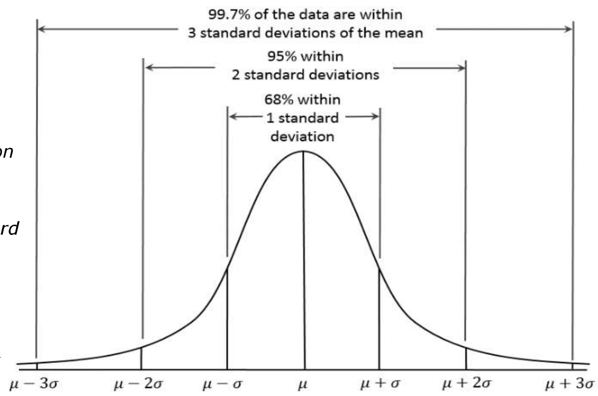
The final choice is determined by randomly selecting an alternative from the probability array.

Stochastic errors reflected in our model

- Research Question: How to determine the minimum nr. of model runs to reach a stable prediction?
- To answer this question, the concept of Confidence Interval is adopted.

Given:

a normal distribution with the mean u, and the standard deviation σ.



Confidence Interval

- A point estimate
  - A sample statistic could be used to estimate a population parameter, e.g., the sample mean $\bar{X}$ is a point estimator of the population mean μ
  - Cannot be expected to provide the exact value of the population parameter.
- An interval estimate
  - Computed by adding and subtracting a margin of error to the point estimate with a certain level of confidence:

$$\text{Point estimate} \pm \text{Margin of error}$$

  - Provide information about how close the point estimate is to the value of the population parameter.

Confidence Interval

- An interval estimate
  - Computed by adding and subtracting a margin of error to the point estimate with a certain level of confidence:

$$\text{Point estimate} \pm \text{Margin of error}$$

  - Provide information about how close the point estimate is to the value of the population parameter.
- Specific situation: Confidence interval of a population mean μ with unknown population standard deviation σ

$$\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where s is the sample standard deviation, *(1-a)* is the confidence coefficient and $t_{a/2}$ is the t value providing an area of $a/2$ in the upper tail of the t distribution with n-1 degree of freedom.

Example of Confidence Interval

- Estimate the true mean 'CR' of the whole roadway with 95% confidence level, when the population standard deviation is unknown.

ID	Segment	CR	ID	Segment	CR
1	26	0.09	21	2	0.399
2	35	0.161	22	17	0.425
3	11	0.172	23	1	0.426
4	6	0.174	24	7	0.445
5	8	0.195	25	28	0.454
6	9	0.283	26	33	0.471
7	27	0.312	27	5	0.479
8	21	0.342	28	3	0.496
9	31	0.345	29	37	0.496
10	29	0.358	30	12	0.502
11	20	0.373	31	23	0.503
12	34	0.374	32	25	0.504
13	18	0.382	33	19	0.541
14	22	0.383	34	32	0.545
15	39	0.385	35	36	0.556
16	38	0.387	36	16	0.583
17	4	0.387	37	15	0.586
18	24	0.393	38	13	0.591
19	30	0.394	39	14	0.618
20	40	0.395	40	10	0.757

Example of Confidence Interval

- Step1: calculate mean value $\bar{X}$ based on the sample data

$$\bar{X} = \sum_{i=1}^{n} x_i/n = 0.417$$

- Step2: calculate standard deviation s

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2} = 0.137$$

- Step3: look for $t-value$ in ***t-distribution table*** with n-1 degree of freedom and 95% confidence level

  95% confidence level: $a/2=0.05/2=0.025$

  n=40: degree of freedom=39

$$t_{0.025} = 2.023$$

Table 2 (Continued)

Degrees of freedom	Area in upper tail .20	.10	.05	.025	.01	.005
30	.854	1.310	1.697	2.042	2.457	2.750
31	.853	1.309	1.696	2.040	2.453	2.744
32	.853	1.309	1.694	2.037	2.449	2.738
33	.853	1.308	1.692	2.035	2.445	2.733
34	.852	1.307	1.691	2.032	2.441	2.728
35	.852	1.306	1.690	2.030	2.438	2.724
36	.852	1.306	1.688	2.028	2.434	2.719
37	.851	1.305	1.687	2.026	2.431	2.715
38	.851	1.304	1.686	2.024	2.429	2.712
39	.851	1.304	1.685	$\boxed{2.023}$	2.426	2.708
40	.851	1.303	1.684	2.021	2.423	2.704
41	.850	1.303	1.683	2.020	2.421	2.701
42	.850	1.302	1.682	2.018	2.418	2.698
43	.850	1.302	1.681	2.017	2.416	2.695
44	.850	1.301	1.680	2.015	2.414	2.692

Example of Confidence Interval

- Step4: substitute the values into our confidence interval formula

$$\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} = 0.417 \pm 2.023 \times \frac{0.137}{\sqrt{40}} = 0.417 \pm 0.044$$

- Step5: obtain the interval estimate

$$[0.417 - 0.044, 0.417 + 0.044] = [0.373, 0.461]$$

- Step6: obtain the confidence interval

$$CI = 0.461 - 0.373 = 0.088 = 2 \times 0.044$$

Stochastic errors reflected in our model

- Research Question: How to determine the minimum nr. of model runs to reach a stable prediction?
- Based on the concept of Confidence Interval:

$$CI_{(1-\alpha)\%} = 2 \times t_{(\alpha/2,n-1)}\frac{S}{\sqrt{n}}$$

$CI_{(1-\alpha)\%}$: $(1-\alpha)\%$ confidence interval for the true average value;

α: the probability of the true average value not lying within the CI;

$t_{(\alpha/2,n-1)}$:Student's t -statistic with $n-1$ degrees of freedom;

n : required number of model runs;

$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2}$: estimated standard deviation.

Experiment

$$CI_{(1-\alpha)\%} = 2 \times t_{(\alpha/2,n-1)}\frac{S}{\sqrt{n}}, S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

- Based on the formula, an iterative computation is applied in our experiment.
- Specific steps:

– Step 1: Feathers is performed 100 successive runs using 10% fraction of full population. Thus, the standard deviation in our formula can be estimated w.r.t. the indices x under study.

– Step 2: By selecting a 95% level of confidence and predefine the stable condition as 10% fraction of the average value of the index (X) after 100 runs $CI_{(1-\alpha)\%} \leq 0.1 \times \bar{x}_{100}$, the required number of model runs could be calculated using this formula.

Important Indices

- Three travel indices:

– average daily number of activities per person

– average daily number of trips per person

– average daily distance travelled per person

- Segmentation:

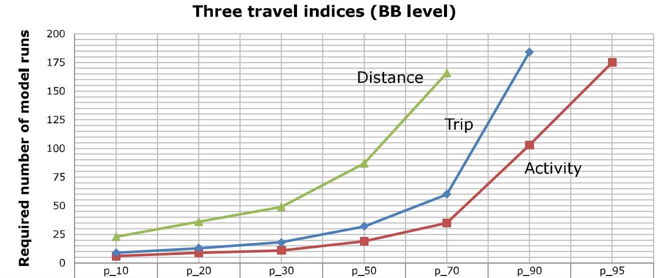
– gender (male, female)

– age (18-34, 35-54, 55-64, 65-74, 75+ years)

– transport mode (car as driver, car as passenger, non-motorised mode, public transport)

– activity type (home-related activity, work-related activity, shopping activity, touring activity)

Three travel indices



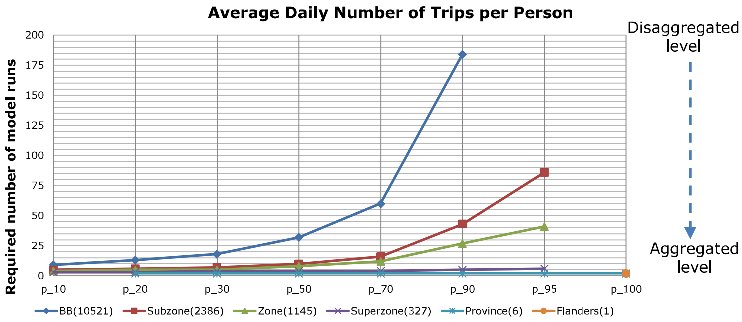
	p_10	p_20	p_30	p_50	p_70	p_90	p_95
Activity	6	9	11	19	35	103	175
Trip	9	13	18	32	60	184	>200
Distance	23	36	49	87	166	>200	>200

- For three travel indices:

the daily distance travelled > the daily number of trips > the daily number of activities

due to the accumulation of stochastic error over steps

Three travel indices (geographical level)



- For six geographical levels:

At highly aggregated levels:

require limited nr. of model runs

Average daily number of trips --- Percentile 70%					
BB	Subzone	Zone	Superzone	Province	Flanders
60	16	12	4	2	1

Segmentation

- Segmentation:

Gender, age, transport mode, activity type

- Consistent results:

1. *A more aggregated level: easier to achieve the predefined stable condition.*

aggregated: the size of the geographical scale and the detailed level of segmentation

Age segmentation

BB (10521)	Nr. of persons	required minimum nr of runs				Province (6)	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
18-34	119657	81	161	>200	>200	18-34	119657	2	2	3	3
35-54	181022	59	113	>200	>200	35-54	181022	2	2	2	2
55-64	67781	143	>200	>200	>200	55-64	67781	3	3	3	3
65-74	63261	186	>200	>200	>200	65-74	63261	3	3	3	3
75+	47409	>200	>200	>200	>200	75+	47409	3	3	3	3

Overall index

BB (10521)	Nr. of persons	required minimum nr of runs				Province (6)	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
Overall	479130	32	60	184	>200	overall	479130	2	2	2	2

Segmentation

- Consistent results:

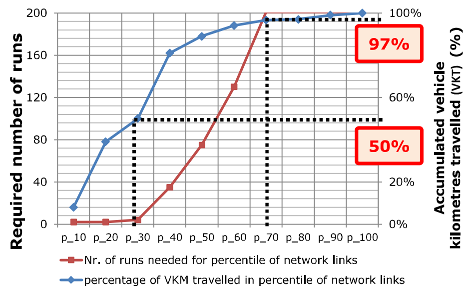
2. *Fewer model runs: segmentations that potentially involved more trips or activities.*

Activity segmentation

Zone (1145)	required minimum nr of runs			
	p_50	p_70	p_90	p_100
Home-related Activity	4	5	10	68
Work-related Activity	12	22	60	>200
Shopping Activity	31	60	151	>200
Touring Activity	133	>200	>200	>200

consistent with the frequency happened in our daily life

Vehicle kilometers travelled (Network Link Level)



The required nr. of model runs and the accumulated VKT at each network link level

- Each network link level:
  - 30% of network links: Less than 5 model runs

    cover 50% of the total vehicle kilometres travelled on the road network.
  - 70% of network links: require more than 200 model runs

    cover over 97% of the total vehicle kilometres travelled on the road network.

Vehicle kilometers travelled (Aggregated level)

The required minimum number of model runs at the aggregated level w.r.t vehicle kilometers travelled

Flanders	100 runs Average (×10⁵ kilometer)	minimum nr of runs required (CI<=0.1×average value)
overall	72.312	2
Linktype1	30.063	2
Linktype2	1.596	3
Linktype3	18.991	3
Linktype4	0.071	7
Linktype5	9.036	3
Linktype6	1.985	3
Linktype7	9.343	3
Linktype8	1.228	2

Note: Linktypes1-8 represent the route link type of Flemish road network.

- The whole Flanders:
  - The overall vehicle kilometres travelled reach the stability faster.
- Aggregated route link type level:
  - At most 7 model runs:

    satisfy the stable condition.

    (Linktype4: the one with the lowest amount of vehicle kilometres travelled)
- The application of FEATHERS at an aggregated level only requires limited model runs.

Impact of specific zonal characteristics

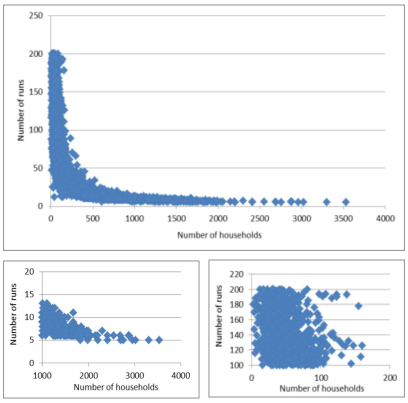
- Objective:

to investigate the potential influencing factors that may have significant impact on the variance of the minimum number of model runs for each zone.

- 4 variables describing specific zonal characteristics:
  - the number of households in each zone,
  - the number of employees in each zone,
  - the area size of each zone, and
  - the population density of each zone.

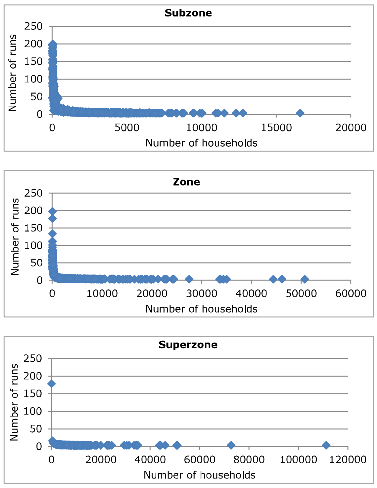
Nr. of households at BB level

- The number of model runs decreases when the number of households in each TAZ is increasing.
- Less than 15 model runs: the number of households (BB) > 1000
- More than 100 model runs: the number of households (BB) < 200



The relation between the nr. of households and the minimum nr. of model runs w.r.t. the average daily nr. of trips per person

Nr. of households at aggregated level



The relation between the nr. of households and the minimum nr. of model runs w.r.t. the average daily nr. of trips per person at the aggregated geographical level

Note:

The Province level and the whole Flanders level are not included since only two model runs are needed for these two levels.

- The number of model runs:
  - decreases when the nr. of households is increasing.
  - decreases along with the aggregation of geographical level.

Quantify the impact of zonal characteristics

- Apply a generalized linear model (GLM):

  to figure out the interaction of 4 zone-related variables on the variance of the required minimum number of model runs.
  - independent variable: the 4 zone-related variables,
  - dependent variable: the required number of model runs.
  - adopt Poisson loglinear model and Negative binomial model
- Data preparation: take the natural log of the independent variables.
  - replace 0 with 1.
  - remove zones requiring more than 200 model runs from the data set as no exact values are available.

Results of negative binomial model

The estimated parameters of the negative binomial model for BB and Subzone levels

	BB				Subzone			
Parameter	Coeffi.	Std. Err.	Wald Chi-Square	Sig.	Coeffi.	Std. Err.	Wald Chi-Square	Sig.
Intercept	8.072	0.053	23019.458	**	7.482	0.044	28463.447	**
Household	-0.386	0.037	107.763	**	-0.173	0.053	10.560	*
Employee	-0.284	0.036	61.681	**	-0.513	0.053	92.117	**
Area size	-0.169	0.048	12.346	**	-0.054	0.017	9,987	*
Population density	-0.161	0.049	10.998	**	-0.058	0.018	10.859	*

Significance codes: **: 0.001, *: 0.01

- All of the four variables have negative effects on the variation of the model runs.

  more model runs are needed if the studied variable is decreasing.
- BB level: all of the four variables significantly influence the nr. of model runs.

  'Household': a relatively higher coefficient
- Subzone level: 'Employee' is the only influential variable.

  A higher coefficient indicates 'Household' and 'Employee' are more influential.

Case study of ABM

- Case 1: Leuven light rail project

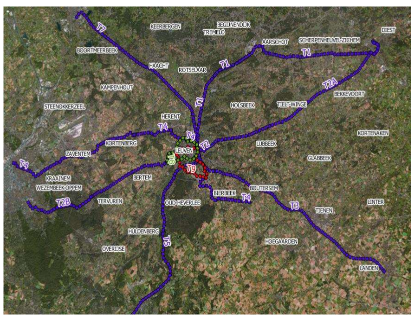
  (Transportation supply planning)
  - Background: in this case study, the city of Leuven is selected as a study area to perform prediction of the travel demand. The city owns quite large transport potential, and is yet reasonably compact in size. Nevertheless, the city has no urban or regional light rail system so far.

Case study: Leuven light rail project

- Objective:
  - to investigate the potential impact of a regional light rail system on travel demand.
- Perform in 2 scenarios:
  - Null scenario:

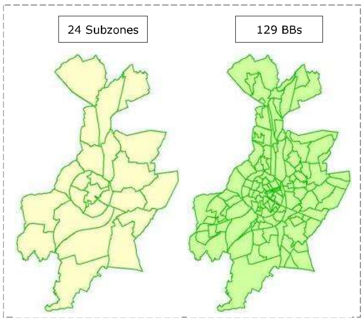
    limited to the situation where no light rail network is included. The public transport network contains only train and bus lines.
  - Light rail scenario:

    integrated with the proposed light rail network information.



The proposed regional light rail network surrounding Leuven consists of 10 different lines with a total length of about 250km.

City of Leuven



Geographical structure of Leuven

- Municipality Leuven:
  - a capital city in Flemish region of Belgium.
  - Superzone (ID: 124), consists of 24 Subzones and 129 BBs.

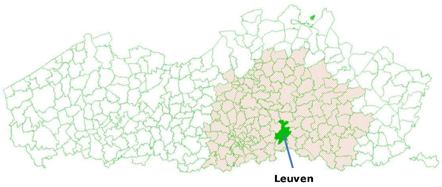
Restrain the size of study areas

- One of the practical limitations of applying ABM: computation time. *especially when large amount of population and detailed geographical level are taken into account.*
  - 16 hours: for a single model run based on 10% of the full population of Flanders at BB level.
  - at least two days: based on 50% of the full population of Flanders at BB level.
  - the computation time will be magnified dramatically: by considering the effects of stochastic errors and therefore multiple model runs are required.

Restrain the size of study areas

- It is often the case that merely a small territory (e.g., a municipality) rather than the whole country is the focus of a specific study.
- Therefore, a relatively small study area surrounding the target territory is needed for investigation rather than to take the whole region into account.
- Solution (tradeoff): To reduce the computation time, one tradeoff can be made in the application. Which is to restrain the size of the study area and conduct the computation only for the selected region.
  - specific study: only a small territory (e.g., a municipality) rather than the whole Flanders is the focus.

Study area of Leuven



Investigated study area for Leuven municipality

- Investigate the study area surrounding Leuven:
  - execute FEATHERS (ABM),
  - based on 50% of the full population of Flanders,
  - by comparing the travel demand difference between the study area and the whole Flanders, and predefine the accuracy level as 90%,
  - study area for Leuven: include 135 municipalities w.r.t the public transport.
  - It should be noted here: The restrained study area covers the proposed whole light rail network.

Validation of Leuven study area

Comparison between Leuven study area and the whole Flanders w.r.t. public transport mode

Transport mode	Study area	No. of persons within the SA (50%)	No. of households within the SA (50%)	Running Time (hour)	Accuracy level	Time saving
Public transport	Leuven study area	1021660	633216	28.8	89.6%	55%
	Flanders	2395514	1449213	64.5	-	-

- Validation results of the study area compared with the whole Flanders w.r.t. the public transport mode:
  - The data shows in the study area of Leuven, the population is almost equal to half of that of whole Flanders.
  - Results show a high accuracy level (89.6%). And 55% of the computation time could be saved compared to that for whole Flanders.

Minimize the effects of stochastic errors:

Nr. of Super zones	Super zone ID	Nr. of model runs	Nr. of Super zones	Super zone ID	Nr. of model runs	Nr. of Super zones	Super zone ID	Nr. of model runs	Nr. of Super zones	Super zone ID	Nr. of model runs	Nr. of Super zones	Super zone ID	Nr. of model runs
1	8	**3**	28	129	**3**	55	21	**4**	82	103	**4**	109	311	**4**
2	16	**3**	29	266	**3**	56	25	**4**	83	104	**4**	110	316	**4**
3	28	**3**	30	271	**3**	57	27	**4**	84	105	**4**	111	310	**4**
4														**4**
5														**5**
6														**5**
7														**5**
8														**5**
9														**5**
10														**5**
11														**5**
12														**5**
13	69	**3**	40	320	**3**	67	58	**4**	94	125	**4**	121	108	**5**
14	71	**3**	41	321	**3**	68	61	**4**	95	126	**4**	122	110	**5**
15	72	**3**	42	323	**3**	69	67	**4**	96	127	**4**	123	114	**5**
16	74	**3**	43	324	**3**	70	70	**4**	97	128	**4**	124	117	**5**
17	77	**3**	44	325	**3**	71	80	**4**	98	130	**4**	125	118	**5**
18	78	**3**	45	326	**3**	72	81	**4**	99	133	**4**	126	121	**5**
19	86	**3**	46	327	**3**	73	84	**4**	100	134	**4**	127	131	**5**
20	89	**3**	47	1	**4**	74	85	**4**	101	211	**4**	128	132	**5**
21	91	**3**	48	3	**4**	75	87	**4**	102	259	**4**	129	135	**5**
22	94	**3**	49	4	**4**	76	88	**4**	103	270	**4**	130	269	**5**
23	95	**3**	50	5	**4**	77	92	**4**	104	272	**4**	131	275	**5**
24	96	**3**	51	10	**4**	78	93	**4**	105	274	**4**	132	100	**6**
25	106	**3**	52	14	**4**	79	99	**4**	106	278	**4**	133	112	**6**
26	113	**3**	53	18	**4**	80	101	**4**	107	281	**4**	134	90	**7**
27	124	**3**	54	19	**4**	81	102	**4**	108	296	**4**	135	98	**7**

- *In order to minimize the effect of stochastic error, 7 model runs is required to ensure all the 135 municipalities to be stable w.r.t. the average daily number of trips per person.*
- *In fact, most of them only need less than five model runs.*

Comparison of two scenarios

The predicted daily travel demand (i.e., the nr. of trips) of 4 transport modes for 2 scenarios

		Car as driver	Car as passenger	Non-motorised mode	Public transport
Null scenario	Run1	1,384,997	333,402	818,210	154,913
	Run2	1,386,325	332,784	818,387	155,000
	Run3	1,383,779	332,003	820,958	154,692
	Run4	1,383,371	333,463	815,492	156,014
	Run5	1,384,407	334,249	817,158	154,650
	Run6	1,384,264	333,080	818,885	154,290
	Run7	1,380,316	333,205	816,984	155,437
	Average	**1,383,923**	**333,169**	**818,011**	**154,999**
Light rail scenario	Run1	1,384,585	334,431	779,447	167,609
	Run2	1,386,632	335,106	781,763	166,223
	Run3	1,386,480	331,012	780,364	167,564
	Run4	1,386,265	333,690	782,130	167,427
	Run5	1,388,175	332,199	781,431	166,685
	Run6	1,385,187	333,697	781,025	167,742
	Run7	1,388,022	332,826	779,965	166,980
	Average	**1,386,478**	**333,280**	**780,875**	**167,176**
% change		**[-0.13%, 0.57%] 0.18%**	**[-0.97%, 0.93%] 0.03%**	**[-5.06%, -4,09%] -4.54%**	**[6.54%, 8.72%] 7.86%**

Lecture summary

- Uncertainty analysis:
  - Input uncertainty
    - Coefficient of variation
    - Effects of population fraction size on uncertainty
  - Model uncertainty
    - Stochastic error
      - Confidence Interval
    - Impact of specific zonal characteristics
  - Case study

Questions

1. Please indicate the purpose of uncertainty analysis.
2. Calculation: Coefficient of variation
3. Calculation: Confidence Interval

Thanks for your attention!

baoqiong@seu.edu.cn