Decision-aid methodologies in transportation

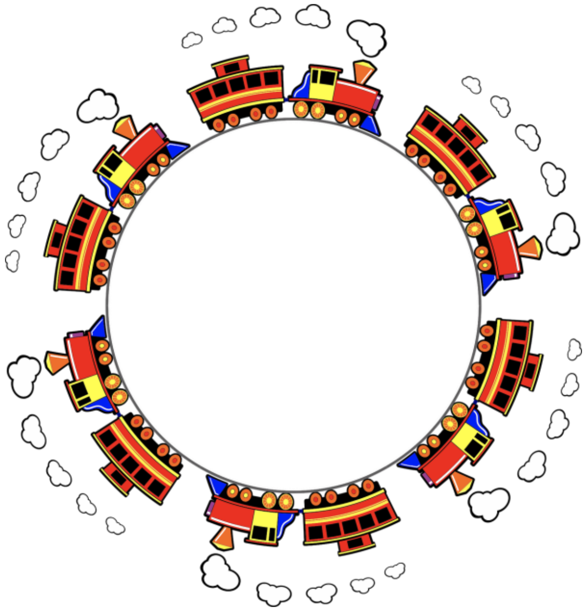
CIVIL-557

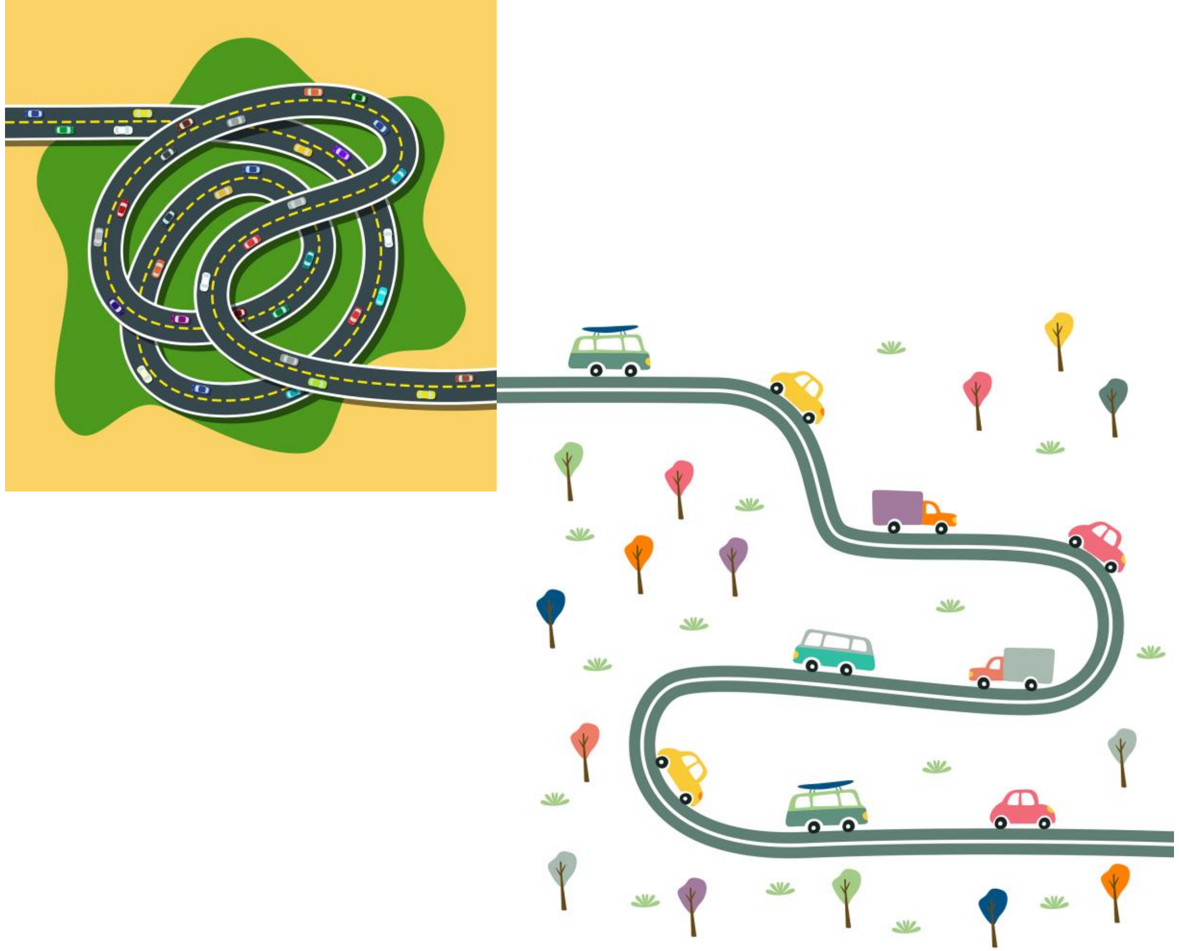
Modelling transportation systems
1. Introduction

Evangelos Paschalidis

The need for modelling transportation systems

Why do we need models of transportation systems?





The need for modelling transportation systems

Transportation (systems):

- enable the movement of people and goods
- facilitate economic activity
- facilitate social interactions
- provide access to essential services
- …
- …

The need for modelling transportation systems

Transportation is a result of the spatial distribution of land uses…

- Motivation: Social and economic growth
- Accessibility: induce more travel
- Aim: save travel time as a resource

The need for modelling transportation systems

… but…

The need for modelling transportation systems

… but…

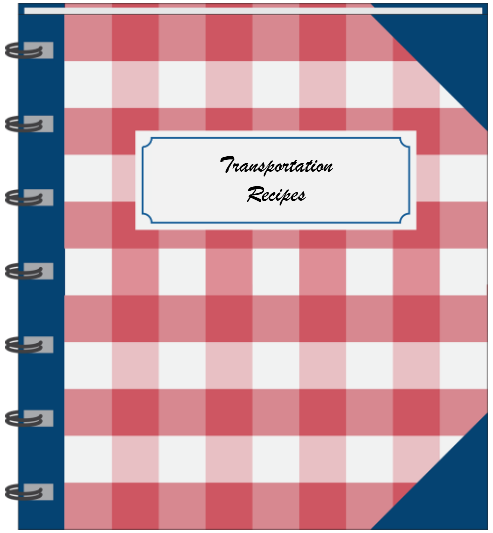
… we can better understand and forecast issues and problems of transportation systems using transport models…

Course objectives

- To understand the types, role, and purpose of transport models
- To be able to select an appropriate model for a given task
- To understand the advantages and disadvantages of the modelling techniques
- Analyse real-world problems

Course structure

- Applied rather than theoretical (maybe not so much in the first lecture though…)
- You are probably familiar with many of the concepts – in this class we will link them all together
- Draws heavily on computer exercises – uses transport related datasets
  - Modelling and data processing in Python
  - Use of the QGIS software for the visualisation of the network and traffic assignment
- Script templates will be provided before each lecture
- A collection of (good) practices in transport modelling



Course structure

Teaching:

- Lectures (Theory)
- Lab sessions (Exercises and coding)

Sometimes, lectures and lab sessions may be mixed

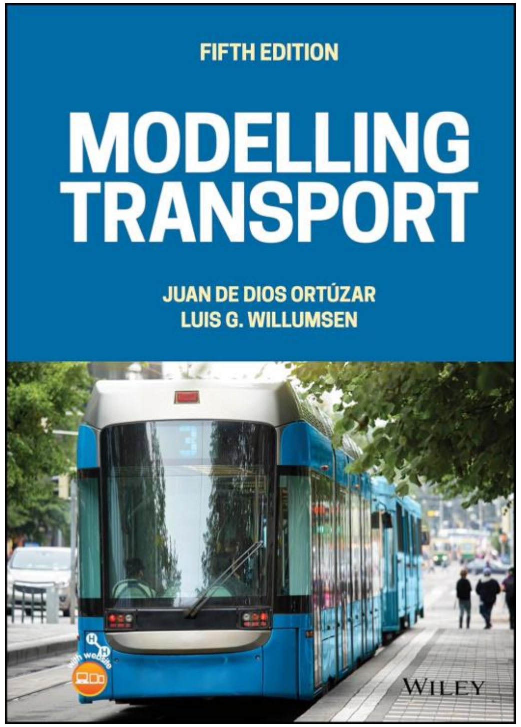
Evaluation:

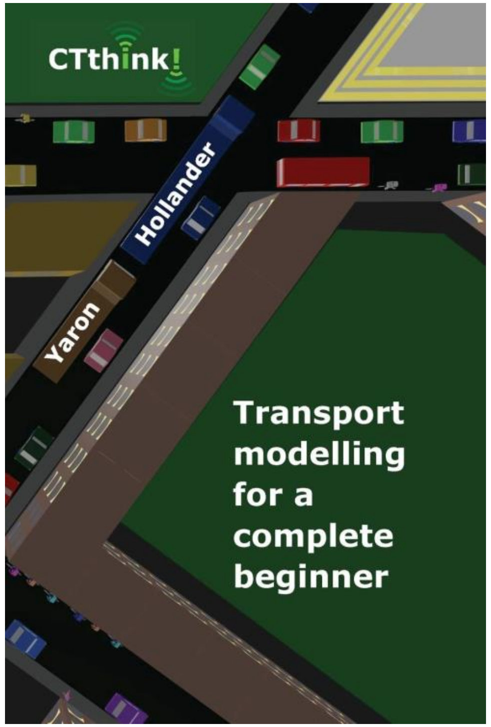
- Project (50% of the grade - Submission: July 4th)
- Written exam (50% of the grade - May 27th)

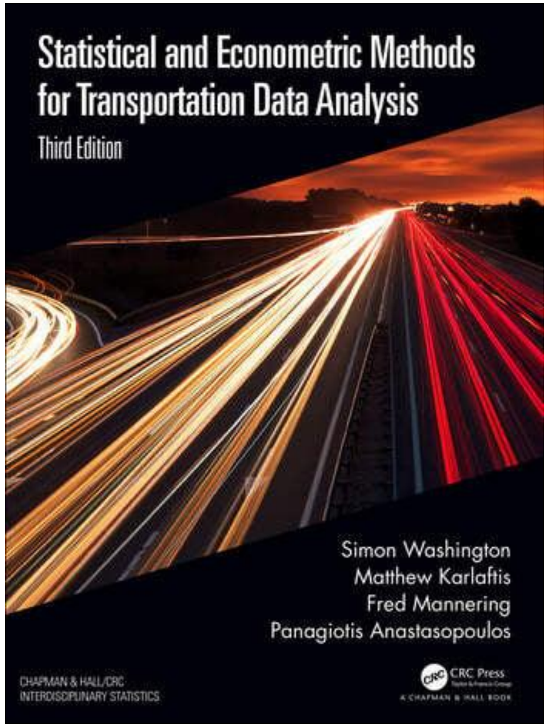
Reading material:

- Course slides and lab scripts: Exam and assignment material (hence some slides may be wordy)
- Key reference textbooks (next slide): Optional - only for own further reading

Key reference textbooks of the course



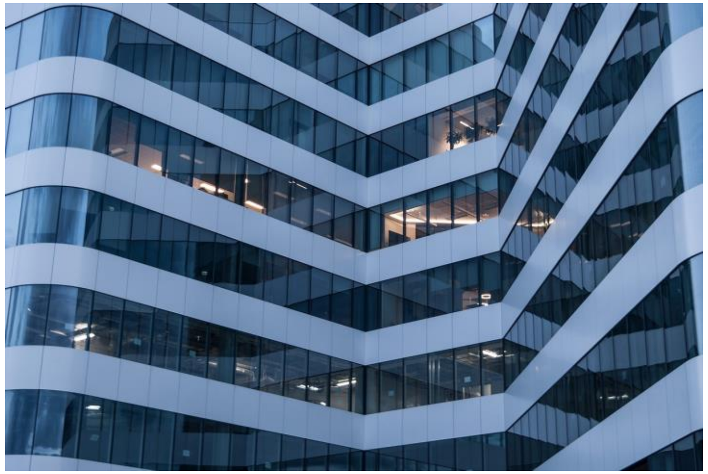
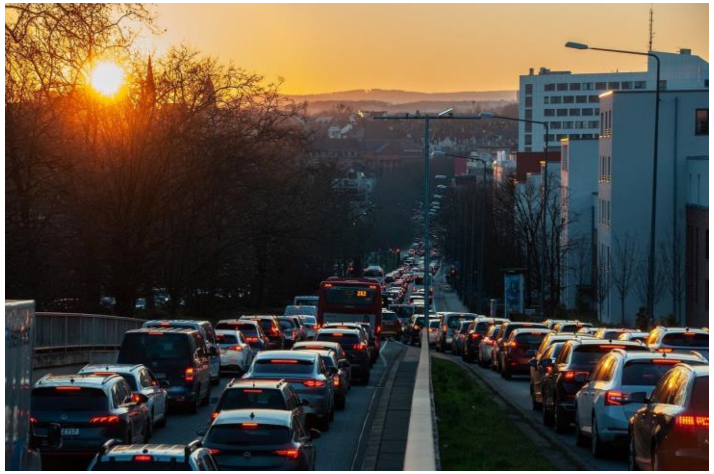



Statistical and Econometric Methods for Transportation Data Analysis
Third Edition
Simon Washington
Matthew Karlaftis
Fred Mannering
Panagiotis Anastasopoulos
CHAPMAN & HALL/CRC INTERDISCIPLINARY STATISTICS
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Setting the background…

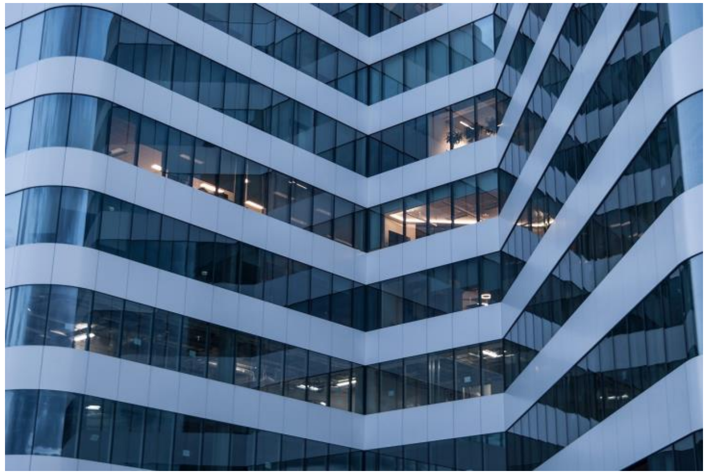
Transportation demand

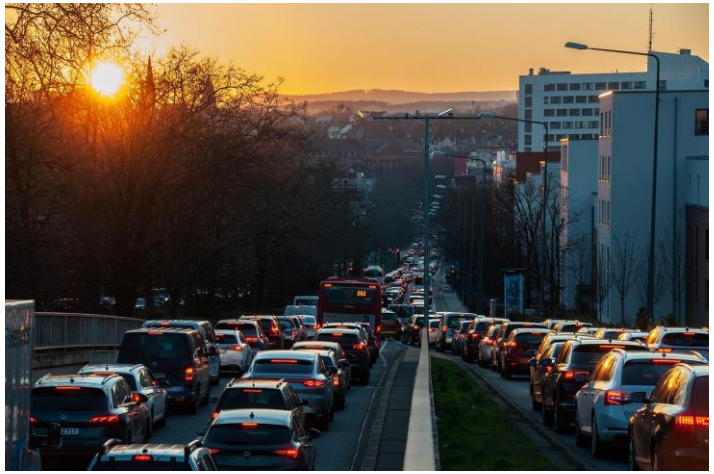
- "The demand for transport is *derived*, it is not an end in itself" (Ortúzar & Willumsen, 2024)
- Transport demand takes place over space
- People travel in order to satisfy a need (work, leisure, health) undertaking an activity at particular locations
- Every trip is associated with some cost (monetary, infrastructure, environmental, safety, travel time)

Transportation demand

- Strong dynamic elements.
  - Demand is mainly concentrated on a few hours of a day, especially in urban areas.
  - A transport system could cope well with the average demand but not during peak periods
- Problem: Demand higher than capacity





Transportation supply

- Service and not a good – it is not possible to stock it e.g. stock and use it in times of higher demand.
- A transport system requires a number of fixed assets, the infrastructure, and a number of mobile units, the vehicles
- Infrastructure and vehicles are not owned nor operated by the same group or company – complex interactions in supply

Transportation supply and demand

Aim of transport planning: ensure the satisfaction of a certain demand D...

- …for movements of person and goods…
- …with different trip purposes…
- …at different times of the day and the year…
- …using various modes…
- …given a transport system with a certain operating capacity.

Transportation supply and demand

The transport system itself can be seen as made up of:

- an infrastructure (e.g. a road network)
- a management system (i.e. a set of rules, for example driving on the right, and control strategies, for example at traffic signals)
- a set of transport modes and their operators.

Transportation planning and modelling

Transportation planning

Transportation planning is:

- The process of defining future policies, goals, investments, and spatial planning designs to prepare for future needs to move people and goods to destinations.
- A collaborative process that incorporates the input of many stakeholders including various government agencies, the public and private businesses

The objectives of transportation planning

The design of transportation systems that will reduce problems in mobility and accessibility, subject to safety, finance, and development constraints.

- Improvement of safety
- Reduction of the (operational) costs
- Reduction of travel time
- More affordable implementation of new projects
- Reduction in the interruption of equilibrium of the system
- Assist the development of the various land uses

Transport modelling

- The main tools of transport planning are transport models
- A model: Simplified representations of the travellers' choices/ behaviour via a series of mathematical procedures
- A model is ranging from a few simple equations to complex and advanced computer software
- An important part of decision-making processes in transport
- Allows users to explore, understand, and estimate the consequences of particular policies, strategies or schemes on a desktop rather than in a real network
- Plays an important role in understanding & interpreting the real world
- Skills of a good transport planner: theoretically sound modelling techniques & with competent implementation in software tools

The scope of transportation models

Forecasting in new situations: What happens if we add a new lane to the motorway or convert a main urban road to a pedestrian street?

Behavioural responses to new situations: How would route choice be affected by the closure of a road?

Relevant transport modes : What the market shares would be if a new public transport mode is introduced in a city?

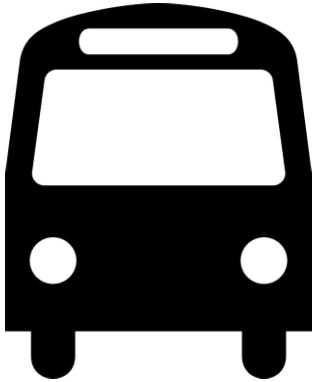
Geographical scope: Understand the range of impact of an intervention

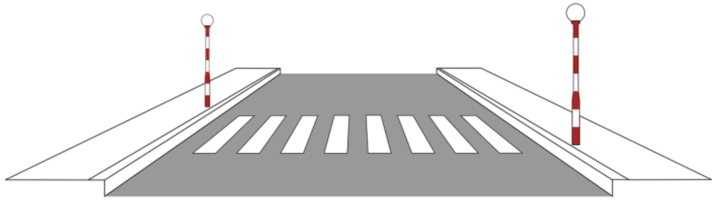
Level of detail: No right or wrong level of detail – it depends on the research question

Two main uses: project design and project appraisal

Use of transportation models: Project design

- New road: how many lanes, expected share of modes, amount of type by vehicle type…
- Traffic calming: times of the day to be applied, effect on traffic…
- New public transport routes: how many services, expected number of passengers…
- Pedestrian – cycling infrastructure: Length of green light cycle, number of cyclists, estimated cycling routes



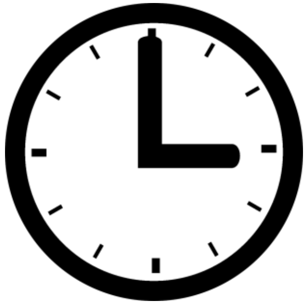


Use of transportation models: Project appraisal

Appraisal: models are the main input when investigating what is the best solution to a transportation problem

- Cost: amount needed for constructions, maintenance, and other expenses
- Income: amount earned if the solution is implemented e.g. fees, tolls, fuel tax etc.
- Benefits: Same as income for private sector investments – additional factors are examined when public money is invested (not purely economic)
- Disbenefits: any negative impact counts as disbenefit

Benefits - Disbenefits

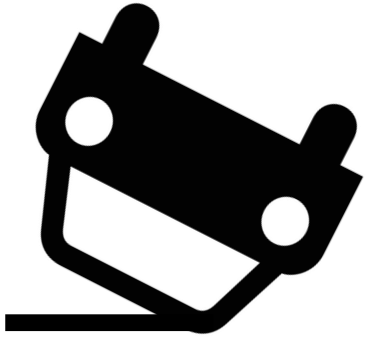


Travel time:
Aim – reduction of travel time.
Both benefit and disbenefit.

Physical activity:
Indirectly captured via shares of walking and cycling trips.

Environmental impacts:
Aim – reduction of air pollution and greenhouse gases.
Estimation of environmental impacts based on number of vehicles, traffic composition, average speed etc.

Noise:
Aim – reduction of noise levels.
Similar to environmental impacts.



Safety:
Number of accidents can be approximated as a function of the number of km travelled.
Safety can be also evaluated with microscopic traffic models

Model levels

Macroscopic models (the main focus of this course)

- Focus on aggregate variables (e.g., flow, density, speed).
- Commonly used for large-scale, regional analysis.

Microscopic models (we will see some elements of these too…)

- Simulate individual vehicles and driver behaviours.
- Capture interactions like car-following and lane-changing.
- Suitable for detailed, small-area studies (e.g., intersections).

Mesoscopic models

- Mix of micro and macro models.
- Model individual vehicles but with simplified behaviour rules.
- Balance detail and computational efficiency.

Agent-based / Activity-based models

- Simulate individual travellers (agents) with daily activity plans.
- Focus on decisions like activity participation, departure times, and travel mode.
- Useful for analysing behavioural responses and policy impacts.

Common terms

Common terms

(Transport) Networks (supply side in a transport model)

- Usually contains streets, roads, junctions, bike lanes, bus lanes etc…
- Graphical representations of the transport system usually represented as a set of zones, links and nodes (see next slide)
- Road network: Captures private and potentially public transport.
- Route (transit) network: Captures public transport itineraries
- Both road and route networks are a simplification of the geometry of the real traffic network
- Smaller roads are often ignored in the representation of the transport network

Common terms

Zones

- reduction of the total study area into manageable portions
- size of zones: compromise between accuracy and estimation efficiency
- choice of boundaries: (1) zones for which data is available, e.g. enumeration districts or political units (2) zones used in previous studies
- centroid: represents the location where traffic is generated within a zone

Links

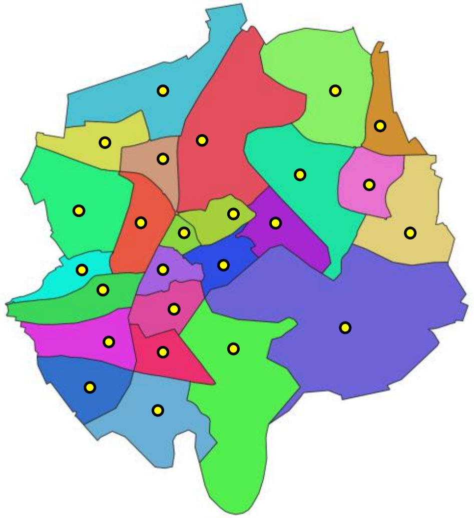
- one-way sections of transport infrastructure or service
- typically assign attributes to links: number of lanes, capacity, speed, direction etc.
- centroid connectors: virtual links that connect to the centroid (do not represent real-life roads, may actually represent a number of real roads)

Nodes

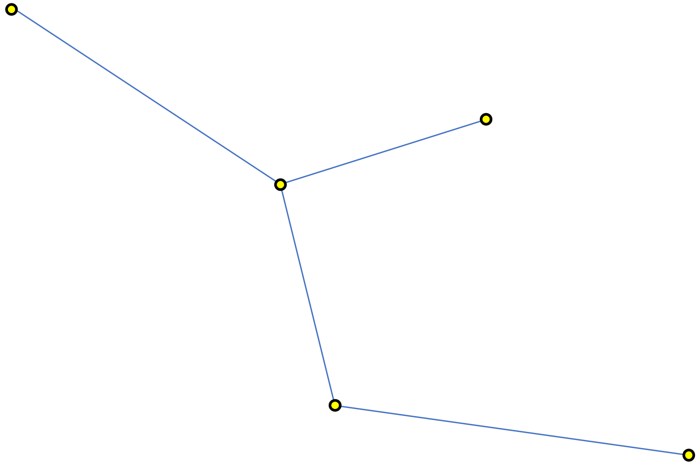
- link end points, typically intersections or points representing changes in link attributes

Common terms

Zones and centroids

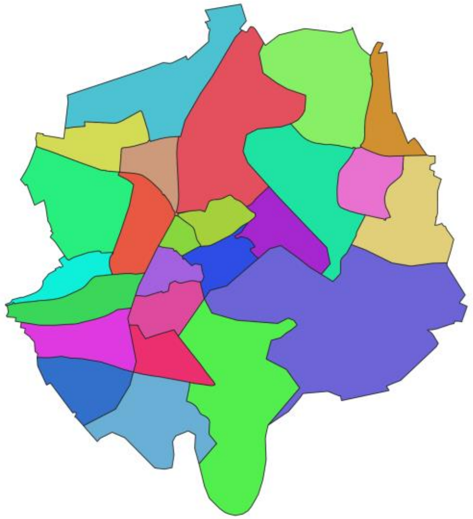


Nodes and links



Zoning - guidelines

- Not too small: otherwise too few trips will be observed in the zone
- Not too large: if a zone generates a lot of demand it is better to further split it to analyse more efficiently
- Typical land use: whenever possible create zones with one dominant land use type (examples: residential, commercial, business…)
- Centroid connector capacity: if a connector received too many trips the model may suggest congestion problems where they don't exist (connector is not a real road)
- Reasonable walking distance: when public transport is involved
- Centred around main transport facilities (e.g. stations)
- Physical barriers (e.g. rivers): so travel is limited between specific barriers while modelling
- Relatively equal size (really case specific!!)



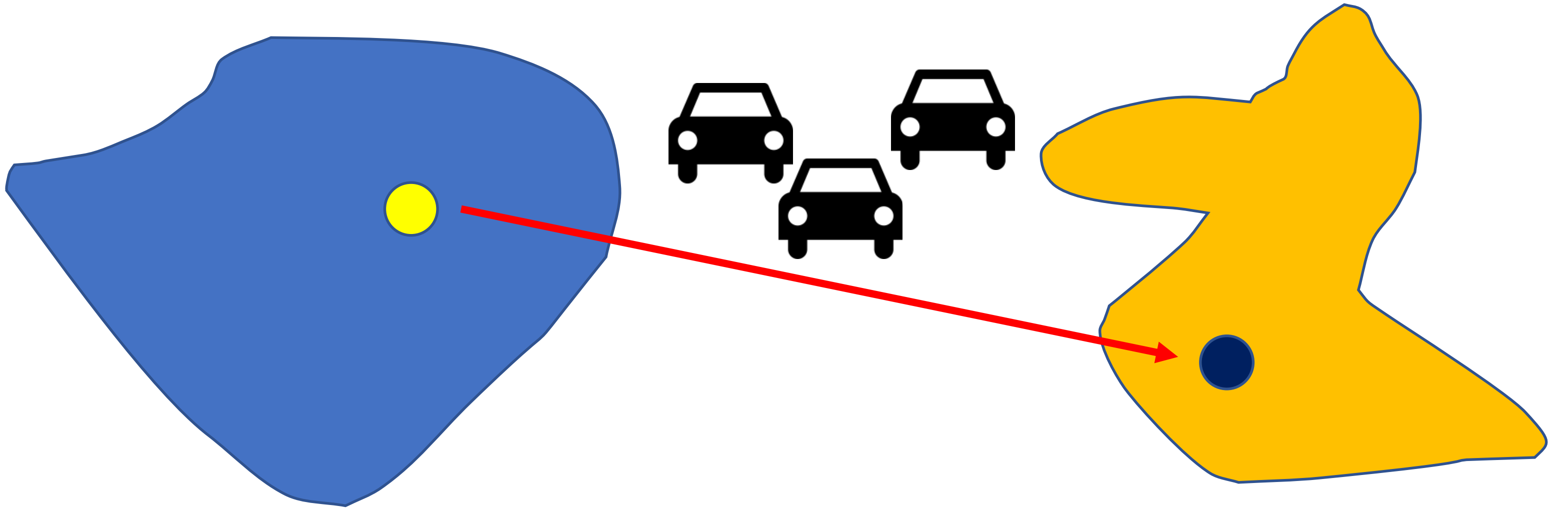
Zoning - guidelines

Small zones	Large zones
Advantages	**Advantages**
More detailed output	Faster estimation times
Easier to ensure pure land use	Less sensitivity to coding network detail
Physical barriers are considered	More data available per zone, results more robust in terms of statistical validity
Model outputs for individual public transport stops	
Fewer problems for trips within the same zone	

Definition of zones

- Centroids:
  - Assuming that all trips from/to a given zone will start /end at its centre
  - Useful simplification when calculating inter-zonal costs

- The number of zones, based on model purpose and data availability, can vary significantly



Definition of zones

- Internal trips: trips within the same zone
    - Difficult to analyse because the origin and destination are the exact same points
    - Some models may completely ignore these internal trips

- External zones:
    - Zones outside the study area
    - Not modelled with a lot of detail to ensure all observed trips have an origin and a destination
    - Still included because traffic from external zones uses the network of the study area

More common terms…

- Journey or tour: a complete excursion (out and back)
- Trip or “Journey leg”: a one way journey
- Origin: The place (zone) where the trip started
- Destination: The place (zone) where the trip ended
- Home-based Trip – trip having origin or destination at the trip-maker’s home
  - Home-based Work (HBW), Home-based Education (HBE), and Home-based Other (HBO)
  - non home-based - all other trips (NHB)
- Mode: Means of transport used for the trip or trip stage
- Trip Purpose - with respect to the destination, e.g. work, business trip, leisure, shopping, education
- Origin-Destination (OD) Matrix: A matrix of trips from particular origins to particular destinations (more on matrices later today)

Modelling considerations

Why zones, links and nodes?

- Lack of data
- Computational complexity
- Coding effort
- Aggregate trip flows are often enough

There are some models that work with exact coordinates (agent-based/activity-based modelling)

Modelling considerations

- Some models ignore the fact different trips same person (there are exceptions)
  - Complex trips of several legs or different trips, depending on the model
- Walking: typically part of another trip – not modelled explicitly
- Level of detail of trips legs depending on the output that we need e.g. do we want separate bus and train trip legs or just consider a public transport trip?

More on models…

- Data input:
  - Socioeconomic / individual characteristics (age, gender, income, etc.)
  - Trip attributes (trip purpose, number of travellers, etc.)
  - Choice attributes (destination choice, mode choice etc.)

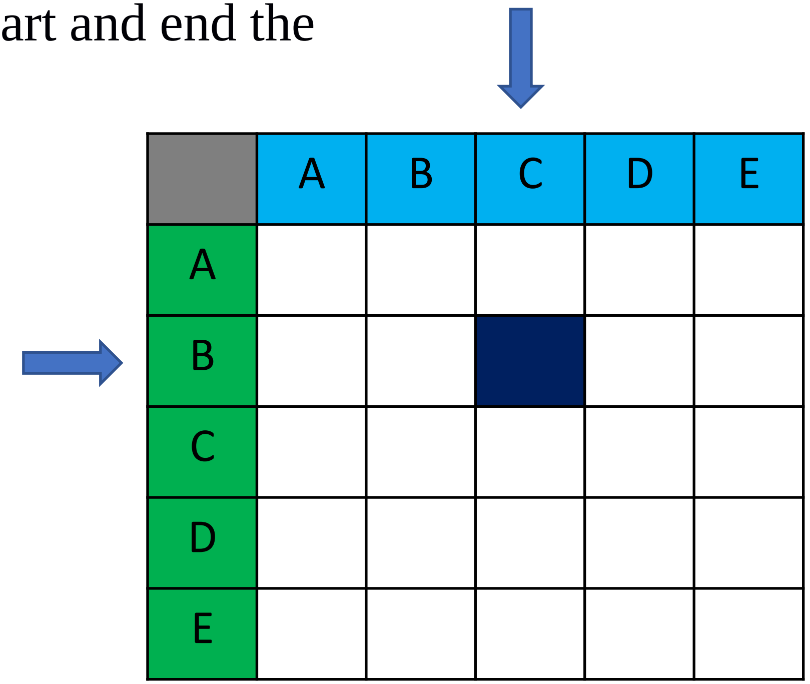
- Models are largely limited to the available data and assumptions that we make
  - Relationships remain constant over time
  - Consider the evolution only of the variables available in the data
  - A model cannot forecast based on the impact of some variable/ attribute that it is not trained

A word on matrices…

Matrices

- Demand matrix (trip matrix): stores information of travel demand
- Rows and columns represent the zones of the study area
- Rows: origin zones
- Columns: destination zones
- Cells: Number or trips from origin (row) to destination (column)
- Number of zones, critical impact on the number of calculations in our model
- Daily OD matrix fairly – but not exactly - symmetrical (most people start and end the day at home)

Example: a trip from origin zone B to destination zone C

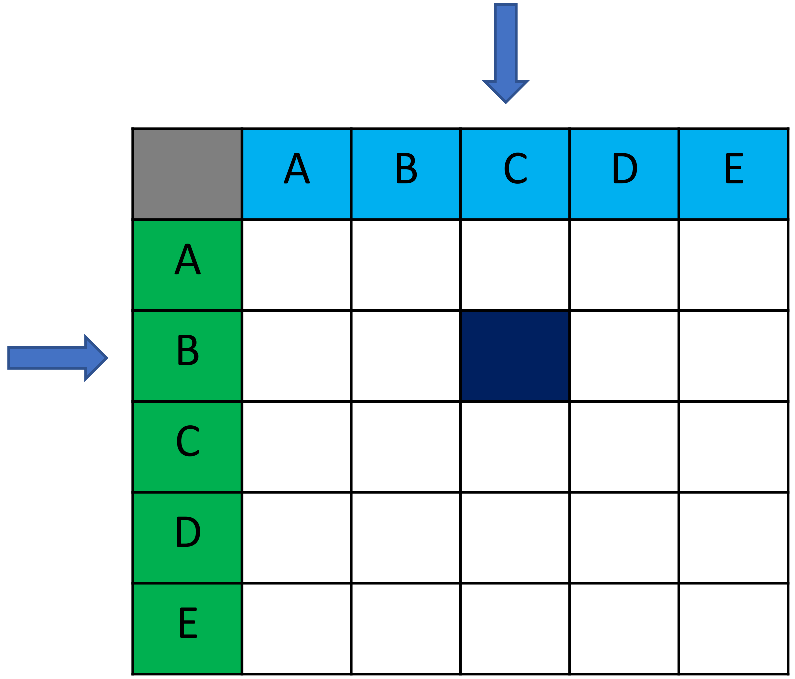


	A	B	C	D	E
A					
B			■		
C					
D					
E					

Matrices

- Rare use of daily OD matrices
- Separate matrices for different time periods (e.g. morning peak or afternoon peak)
- Morning peak OD matrix not symmetric at all!! Clear direction of the demand
- Usually two different exercises
  - We have partial information about our matrix and need to fill the missing cells
  - We already have a filled demand matrix
- We can also use an OD matrix to store travel time or other attributes

Example: a trip from origin zone B to destination zone C



	A	B	C	D	E
A					
B			■		
C					
D					
E					

Understanding a matrix

Let’s assume the matrix shows the trips 7–8 AM

- Not many trips to zone 2? Why?
- Why do zones 4, 5, and 6 receive so many trips?
- Diagonal is overall low, what happens with zone 4??
- How about zone 9? Not much demand there.

Zone	1	2	3	4	5	6	7	8	9	10
1	4	0	2	19	23	33	12	9	3	13
2	17	2	21	32	26	47	28	34	0	32
3	18	20	6	46	61	35	28	25	1	21
4	0	0	5	75	24	29	0	3	6	6
5	1	4	18	12	13	3	4	4	5	6
6	4	9	11	13	31	22	2	4	5	6
7	9	4	43	21	32	3	3	8	6	11
8	10	3	6	14	29	47	7	1	8	14
9	0	0	0	3	2	9	0	1	1	0
10	8	2	3	11	31	29	12	13	1	0

Understanding a matrix

- Not many trips to zone 2? Why?
  - If the matrix represent the morning peak, maybe zone 2 is a residential area.
- Why do zones 4, 5, and 6 receive so many trips?
  - If the matrix represent the morning peak, maybe these are business zones.
- Diagonal is overall low, what happens with zone 4??
  - Mistake or maybe poor zone design? Is it worth splitting to more zones?
- How about zone 9? Not much demand there.
  - What if it is an external zone; not too critical for our analysis

Zone	1	2	3	4	5	6	7	8	9	10
1	4	0	2	19	23	33	12	9	3	13
2	17	2	21	32	26	47	28	34	0	32
3	18	20	6	46	61	35	28	25	1	21
4	0	0	5	75	24	29	0	3	6	6
5	1	4	18	12	13	3	4	4	5	6
6	4	9	11	13	31	22	2	4	5	6
7	9	4	43	21	32	3	3	8	6	11
8	10	3	6	14	29	47	7	1	8	14
9	0	0	0	3	2	9	0	1	1	0
10	8	2	3	11	31	29	12	13	1	0

Costs and externalities

Costs and externalities – generalised cost

- Cost of trips is an integral part of transport models
- Cost is actually generalised cost, not monetary cost
- Generalised cost summarises everything about the travel experience from origin *O* to destination *D*
  - If I have different options each of them will have a different generalised cost (even if monetary cost is the same)
- Typically measured in minutes (or some other time-related unit)
  - Interpretation of time does not change over time
  - Currencies may require conversions across time
- All the elements are added up to on generalised cost function per mode

Costs and externalities – generalised cost

Types of generalised cost

Monetary	Non monetary
Parking fee	In-vehicle travel time
Fuel cost	Walking time to the public transportation (PT) stop
Vehicle maintenance	Headway time (for PT)
Tolls	Time to find parking and associated discomfort
Ticket cost	On board crowding (in PT)
	Crowd noise (in PT)
	Delays in PT
	Road safety and security
	Uncomfortable seats
	Surrounding view

Costs and externalities – weights and penalties

- Boarding penalty: non direct public transport (PT) options (require changes) are less attractive
- Waiting time penalty: 1 minute of waiting time is perceived differently than 1 minute on board
- Walking time penalty: Typically it is perceived negatively to have to walk a long distance to the PT stop
- Crowding penalty: Who likes a crowded metro on a warm summer day?
- Uphill penalty: Physically demanding parts of a trip receive a penalty
- Unreliability factor: Relate to congestion and unreliable PT
- Safety and security: Dangerous junctions or lack of pedestrian/cycling infrastructure can receive a penalty
- Ambience factor: Less common but pleasant or unpleasant journey environment is considered in some models

Costs and externalities – Comments

- Generalised costs capture the average behaviour but do not capture everyone accurately
- Wrong to assume that everyone choses the option with the lowest generalise cost
  - Instead, we assume that the option with the lowest generalised cost attracts more people but some will also choose other options
- “Generalised cost” is sometimes used interchangeably with the term “utility”.
  - Not exactly the same, but very similar in the way they work in transport modelling
  - We will examine further when we talk about modal split

Value of time (VoT)

Definition: the amount of money someone is willing to pay to reduce travel time by one minute

- Different value of time for different demand segments (e.g. higher value of time for commute trips compared to leisure trips)
- When computing generalise cost, we divide all variables that have a price by the value of time to convert them to minutes
- Once we have computed the total generalised cost (in time units), we can multiply by the value of time to convert to monetary units (for economic analysis)
- Drawbacks:
  - Same VoT for a minute saved or a minute lost – a minute lost is perceived as more severe
  - Data reliability: VoT values depend on the data and our assumptions
  - Payment method: may actually affect VoT
  - Uniform behaviour: representation of individual heterogeneity is limited (but possible)
  - Size of the change: VoT may not be the same for smaller or larger changes (e.g. higher value to every minute if the total time saved is larger)

Costs and externalities

Monetary costs

- Paid by the users
- Paid by the operators
- Paid by the tax payers / government

Externality (different than the generalised cost)

- A side effect or consequence that affects other parties without this being reflected in the costs
- Costs (or benefits) imposed on others by a trip that the traveller does not directly pay for or perceive

Costs and externalities

Externalities

- Air / noise pollution
- Greenhouse Gas (GHG) emissions
- Traffic congestion
- Accidents and safety risks
- Infrastructure deterioration
- Public health impact
- Inequal development in the use of the different transport modes

Externalities largely coincide to the main problems that transportation systems face

Demand segmentation

Demand segmentation

- Complicated and time consuming to model each individual behaviour (although it is possible)
- May be too simplistic to assume everyone behaves the same
- Compromise: We model behaviour per some segment and keep a rather small number of different trip purposes (Demand segmentation)
- Summary:
  - Everyone is using the same network
  - The information is stored in different matrices per segment
  - Each segment is using different generalised cost functions

Demand segmentation

More segments	Fewer segments
Realism of outputs	Base year data availability
Realism of appraisal	Patterns may not hold in future
	Model runtime

Demand segmentation

Typical segmentation types:

- Car availability (sometimes car ownership): people with access to car behave differently
- Income: may affect travel choices e.g. willing to pay tolls
- Trip purpose:
  - Commuting: Usually during peak-hours, similar journey every day
  - Business trips: may reflect the policy imposed by the employer
  - Other e.g. leisure or shopping

Steps of a modelling exercise

Steps of a modelling exercise

1. Determine the study area / area of interest
2. Data collection
3. Model specification
4. Model calibration
5. (Model validation)
6. Scenarios
7. Evaluation

Determine the study area / area of interest

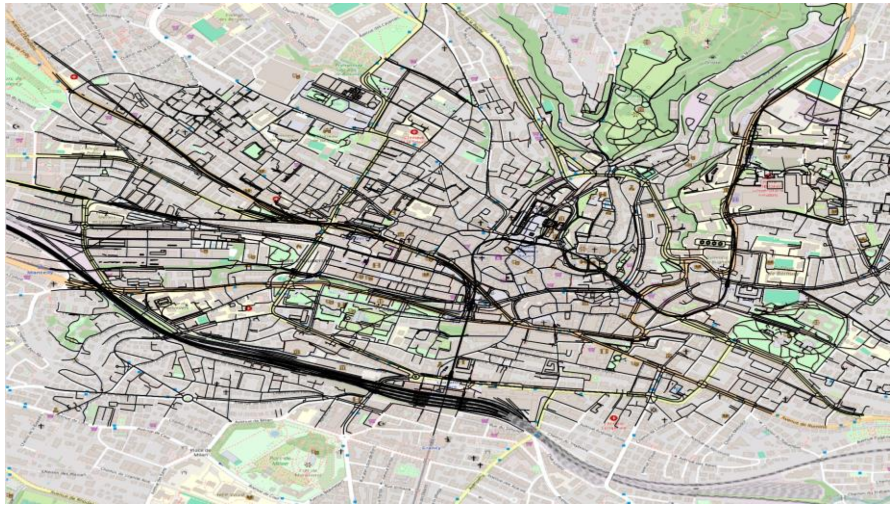
- An area large enough to study all significant impacts and influences of a project or intervention

- Compromise between level of detail and computational efficiency

- Distinguish between “internal zones” in the study area and “external zones” covering the rest of the world

  - External zones serve as a point for trips into, out of, or through the study area
  - Internal zones – part of the study area, much smaller than external zones

- All trips made by all travellers in the study area form the travel demand



Data collection

- Data for model estimation specific
- General data for model validation

- Active – direct response from respondents
  - Assumption: Perfect understanding of the question
  - Assumption: Truthful (social desirability, bad memory issues)
- Passive – observations or extraction from reliable sources
  - Image recognition from videos, traffic counts, Bluetooth signal, GPS data

Data sources

- Household surveys or census survey: questions about travel habits, including trip diary
    - Prons: Powerful for collecting travel information. Key input in many models
    - Cons: Expensive (cost and time). If done online then we may capture only people familiar with technology
- Roadside interviews: Short surveys by stopping drivers. Locations are chosen in away to capture traffic representatively
    - Prons: Good for capturing information unavailable in traffic counts, such as OD and trip purpose
    - Cons: Expensive, cause traffic disruptions and require coordination with bodies like the police

Data sources

- Passenger counts: on buses, on trains and at stations.
  - Prons: Effective for validating public transport models
  - Cons: No info about the actual OD and trip purpose
- Ticketing data: Any data that includes information of ticket purchases e.g. barriers or machines on buses
  - Prons: Potentially can have a lot of information especially if there is an identifier of the user…
  - Cons: … may require a lot of data processing. Privacy issues

Data sources

- On board surveys: Short surveys for public transport passengers
  - Pros: Effective for collection OD and trip purpose
  - Cons: Low participation and low representation of short trips
- Station or bus stop surveys: Short surveys for public transport passengers conducted at stops
  - Pros: Effective for collection OD and trip purpose
  - Cons: Low participation rate, can bias the result towards infrequent services (with longer waiting times)

Data sources

- Behavioural surveys: Surveys asking about people's behaviour in hypothetical scenarios (stated preference)
  - Pros: Very useful for estimating parameters (e.g. modal split)
  - Cons: Real behaviour is different than the stated behaviour
- Mobile phone network data: Following signal transmission to detect locations and travel patterns
  - Pros: Cover wide range of locations at all times of the day
  - Cons: Accuracy bias, uncertainty about mode and trip purpose. Data must be purchased, requires a lot of processing and there are data privacy issues

Data sources

- GPS data: Tracking individuals and learn about travel habits, and travel time
  - Pros: High geographical accuracy, large sample without any fieldwork
  - Cons: Data collected separately by different companies, each app has different characteristics that can bias the data. Privacy issues

Model specification

- Typically a transport model is a collection of models
  - The standard example is the 4–step model (later in this presentation)
- Different types of models (e.g. deterministic or stochastic models)
- Different types of model specifications (linear, non–linear etc.)
- Definition of variables / attributes to be considered

Model Specification, Calibration and Validation

- Models use parameters (pre-calculated numbers) to perform calculations
- The number of parameters ranges from a few to thousands

Parameters representing sizes, durations, and dimensions	Parameters representing individual behaviour	Parameters representing collective behaviour	Parameters describing the composition of traffic
Road width	Preferred walking distance to bus stop	Average car occupancy	Passenger car unit (e.g. a bus is 2-3 passenger car units)
Number of lanes	Willingness to use a mode under given circumstances		
Vehicle length			
Green light duration			
Existence of a pedestrian crossing			

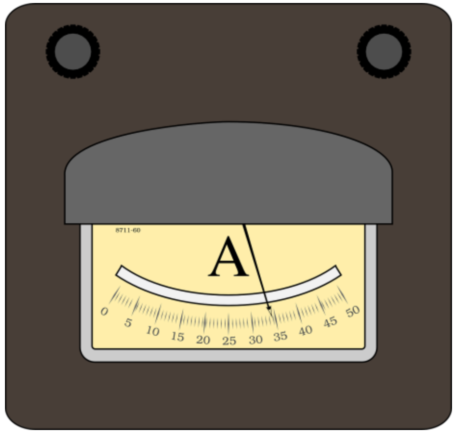
Model Specification, Calibration and Validation

Parameters representing traffic phenomena	Parameters replacing missing parts of the model	Segmentation factors
Typically related to specific traffic behaviour (time for a junction to be cleared)	Hypothetical demand in a future scenario at a time of the day but our model does not consider the time element	We store information regarding the proportion of each segment (e.g. number of trips)

Weights	Penalties	Conversion factors	Scaling factors
Measure the impact of certain variables on behaviour	Parameters that ensure that the model make specific type of behaviour less popular than others	Ensure that all parts of the model are compatible (same units)	Capture people’s sensitivity to small changes on certain choices

The calibration process

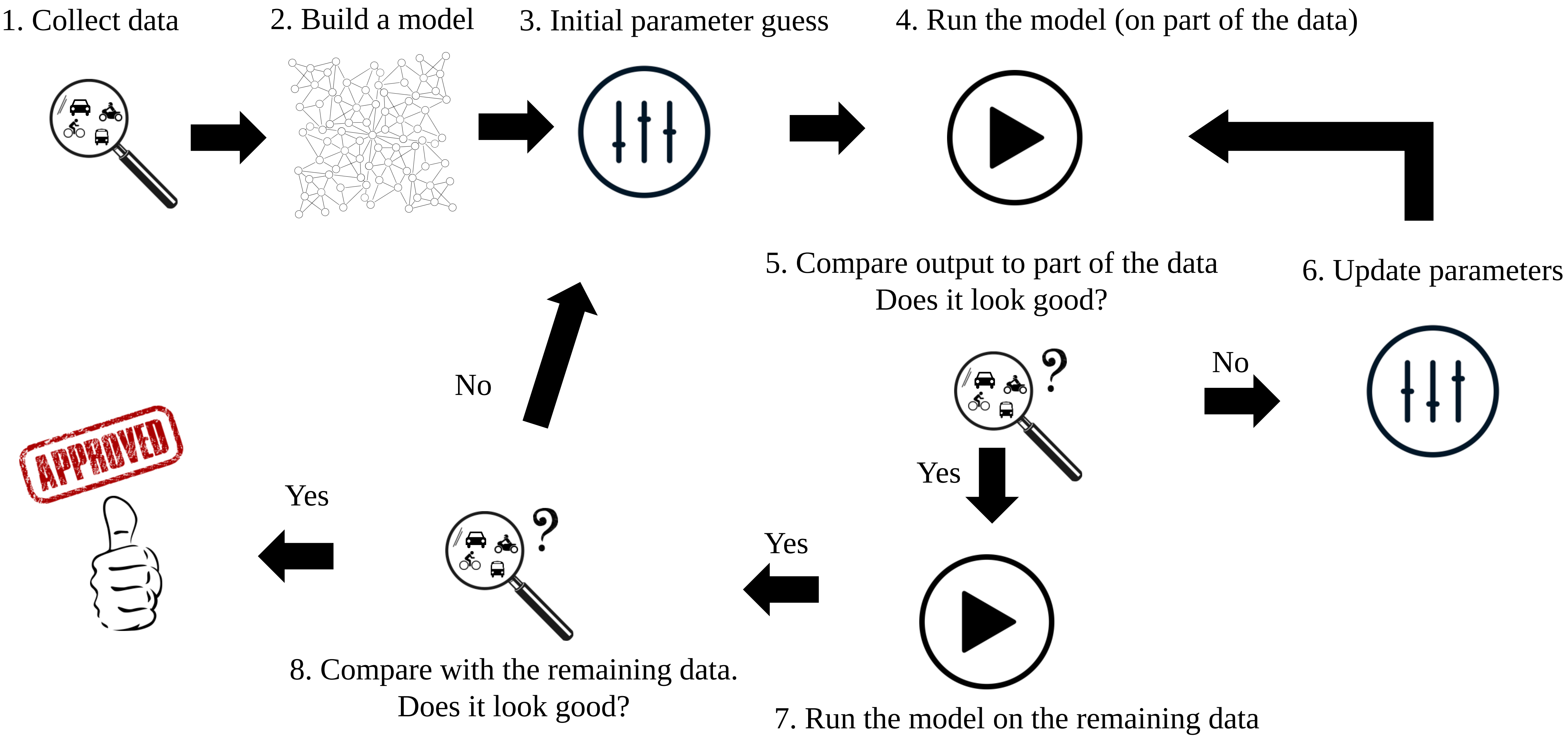
- The process of determining the value of the unknown parameters
- We modify the parameter values until the model output resembles the data
- Can be manual (trial & error) if the number of parameters is low
- Sometimes parameters are calibrated separately for different steps of the model
    - Simpler to separately calibrate different steps
    - We use data from different sources for each model step
- Dependence: generalised costs computed in assignment become the generalised costs of the distribution model in the next iteration



Model Specification, Calibration and Validation

- Parameters are typically part of some simplified theory (e.g. utility theory)
- Parameters representing observed measures are easier to incorporate
- Other parameters we need to “measure” indirectly e.g. change the value until the model output is similar to what happens in a real life transportation system (calibration)
    - E.g. the weights related to sensitivity to travel time and travel cost in a mode choice situation

The calibration-validation process



1. Collect data

2. Build a model

3. Initial parameter guess

4. Run the model (on part of the data)

5. Compare output to part of the data
Does it look good?

No

6. Update parameters

Yes

7. Run the model on the remaining data

Yes

8. Compare with the remaining data.
Does it look good?

No

Yes

The calibration process

- Steps 1 – 6: calibration process
  - Aim: the model reproduces the data used to calibrate it
- Steps 7 – 8: validation process
  - Aim: the model reproduces the data **not** used to calibrate it
- Potential issue: both calibration and validation data suffer from the same issues
- Steps 5 and 8 are evaluated with goodness-of-fit metrics

Calibration limitations

- The model may not reflect well road types that did not exist in our data (e.g. minor roads)
- Are our data still relevant now? Some recent event may have changed travellers’ behaviour
- What if our future investment substantially change travel behaviour and our model is not valid?
  - E.g. we investigate the impact on the transport system by building new luxurious appartements
  - These may (or may not) attract households with different characteristics and different mobility patterns than the current ones.
  - Is our model still relevant?

Validation

- Our model may reproduce well the data used for calibration but may not perform well when other data is used
- Validation is used to check the performance of our model on different data (e.g. some specific cases may have not been captured in the calibration data)
- Validation not a perfect solution – It is very likely that both calibration and validation data suffer from the same issues (especially if they are part of the same data collection)

Implementation - Typical scenarios

- Do-nothing: no intervention – situation may worsen over time
- Do-minimum: do not implement the full investment we had in mind but smaller investments to avoid things getting worse
- Do-something
    - Compare alternatives
    - Factors beyond our own control
    - Factors beyond any control
    - Sensitivity testing

<u>Time considerations</u>

- Base year scenario (road network and demand of the do-nothing scenario)
- Future do-minimum (reference case)
- Different do-something with different time horizons in the future

<u>Assumptions about future demand</u>

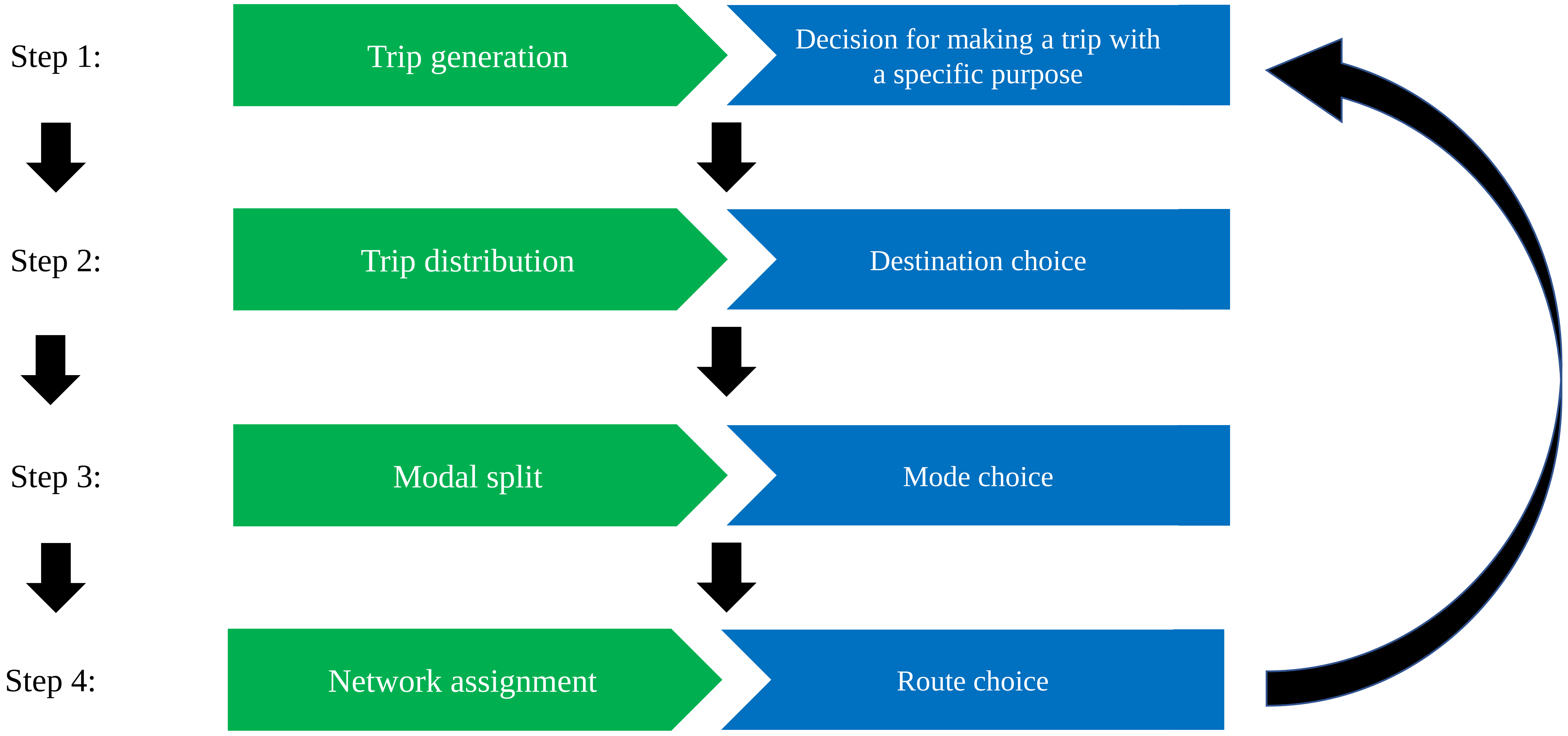
- Future scenarios reflect our interpretation of how future demand will develop
- These assumptions usually involve the engagement of several stakeholders

The 4 – step model

The 4-step model

- Main concept: different parts of the model form a hierarchy
- Iterative process: run several loops of the hierarchy (same as the calibration process we examined earlier)
- Hierarchy is implemented via the 4-step model process
- In each step people make a choice
  - Mathematical implementation typically from the field of choice modelling

The 4-step model



Step 1: Trip generation → Decision for making a trip with a specific purpose

Step 2: Trip distribution → Destination choice

Step 3: Modal split → Mode choice

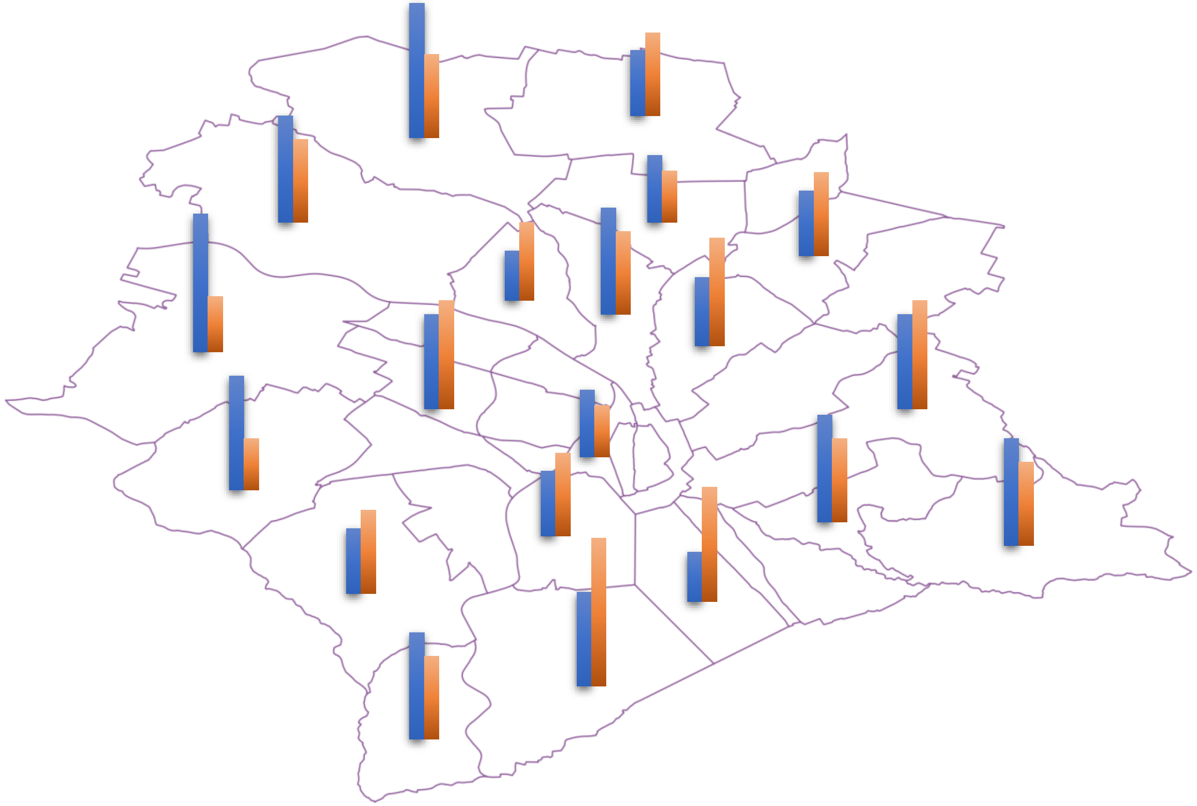
Step 4: Network assignment → Route choice

The 4-step model

- Base year: The year that the majority of data collection takes place
- Typical data input:
  - The origin-destination (O-D) matrix: The number of trips between each origin-destination pair of zones
  - Socioeconomic characteristics segmented population per zone
  - Network characteristics: nodes, links and their attributes (e.g. number of lanes, speed, capacity)
  - Traffic counts for private vehicles and public transportation
- Alternative hierarchy is possible e.g. generation-modal split-trip distribution-assignment
  - Clearer when we talk about the modal split step

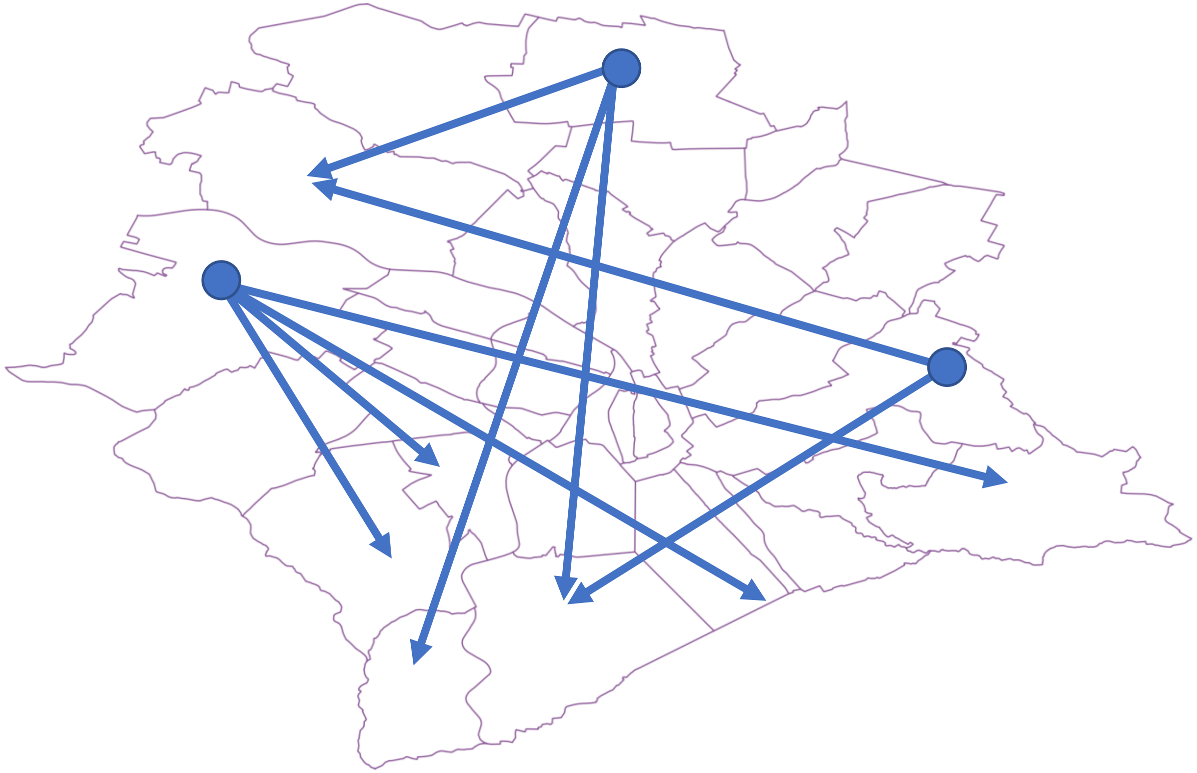
Step 1: Trip generation

- Trip generation step:
  - The number of trips generated in each zone
  - The number of trips attracted in each zone
- Generated trips are typically a function of socioeconomic characteristics and land use
- Attracted trips are typically a function of land use characteristics
- Output: The number of trips generated in and attracted to each zone



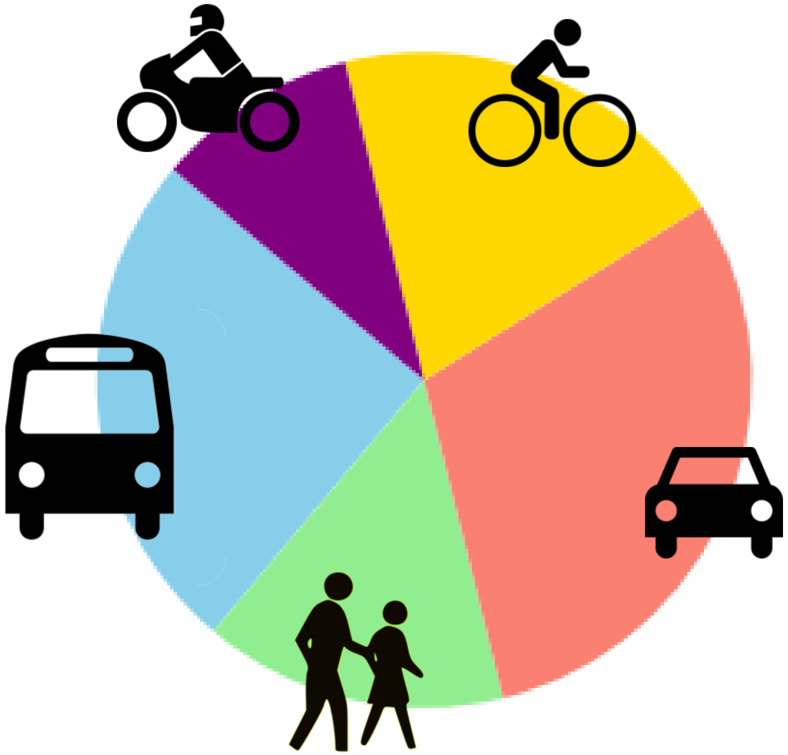
Step 2: Trip distribution

- Trip distribution step:
  - The number of trips between each origin-destination pair
  - The number of trips typically depends on the productivity of the origin zone and the attractiveness of the destination zone
  - Some typical factors that affect trip distribution are the size of a zone, the land use, and the trip cost between the origin and destination zones
  - Input: Trip production/attraction (from step 1), travel cost matrix
  - Output: Origin-destination matrix (typically by trip purpose)



Step 3: Modal split (or modal share)

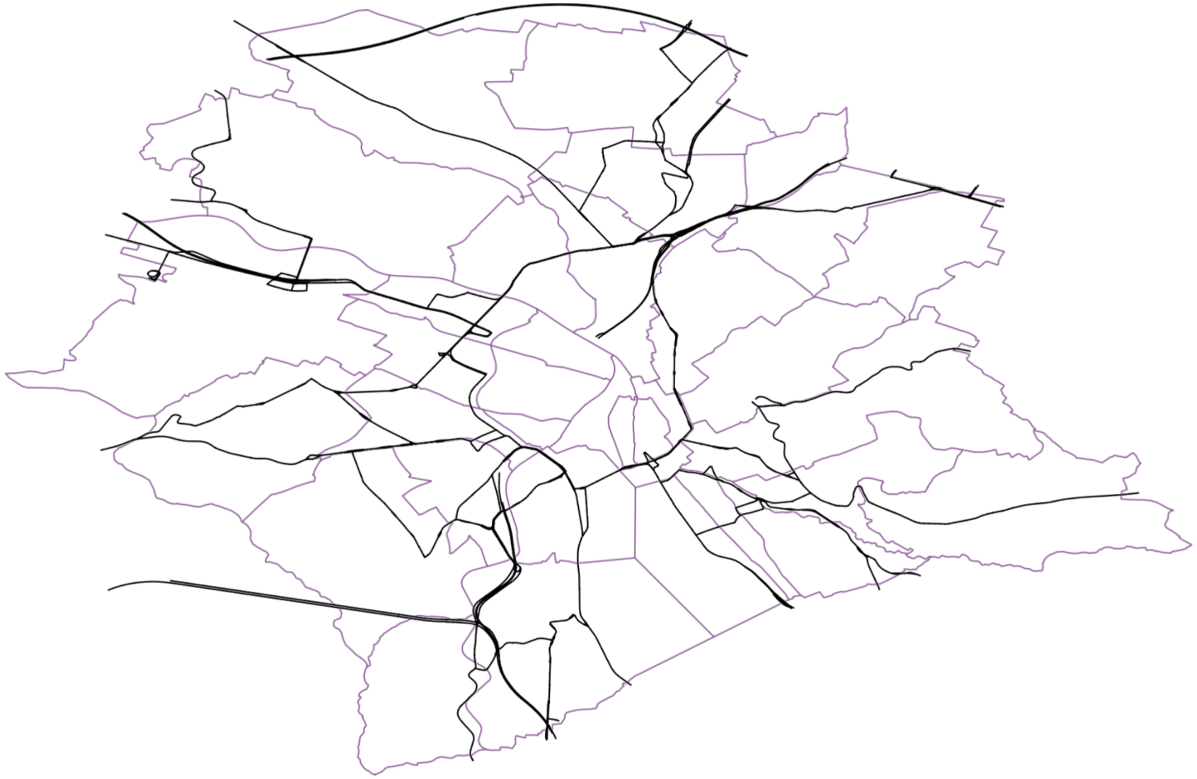
- Modal split step:
  - The number of trips per transport mode between each origin-destination pair
- Main factors are:
  - Mode attributes
  - Socioeconomic characteristics
  - Trip purpose
  - Availability of public transport
- Output: Proportion of each mode used by travellers
- Different OD matrices by mode



Step 4: Network assignment

- Data
  - Representation of the road network with links and nodes
  - Travel time functions per link
  - O-D matrix

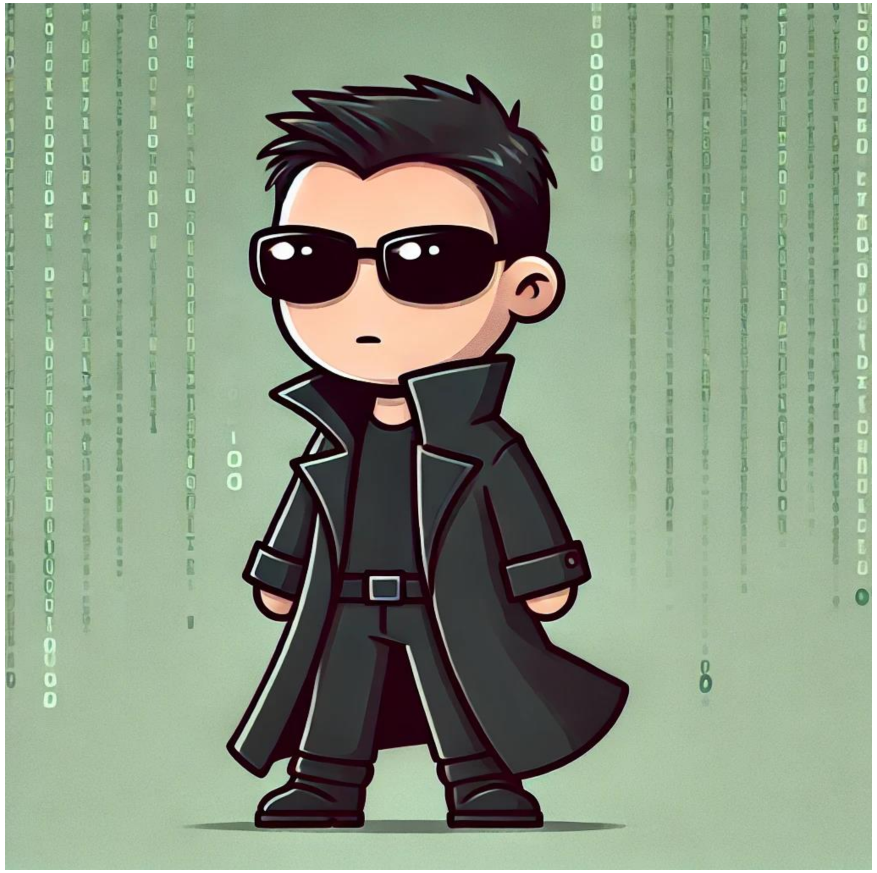
- Output
  - Traffic volume
  - Travel time per link



Some general considerations

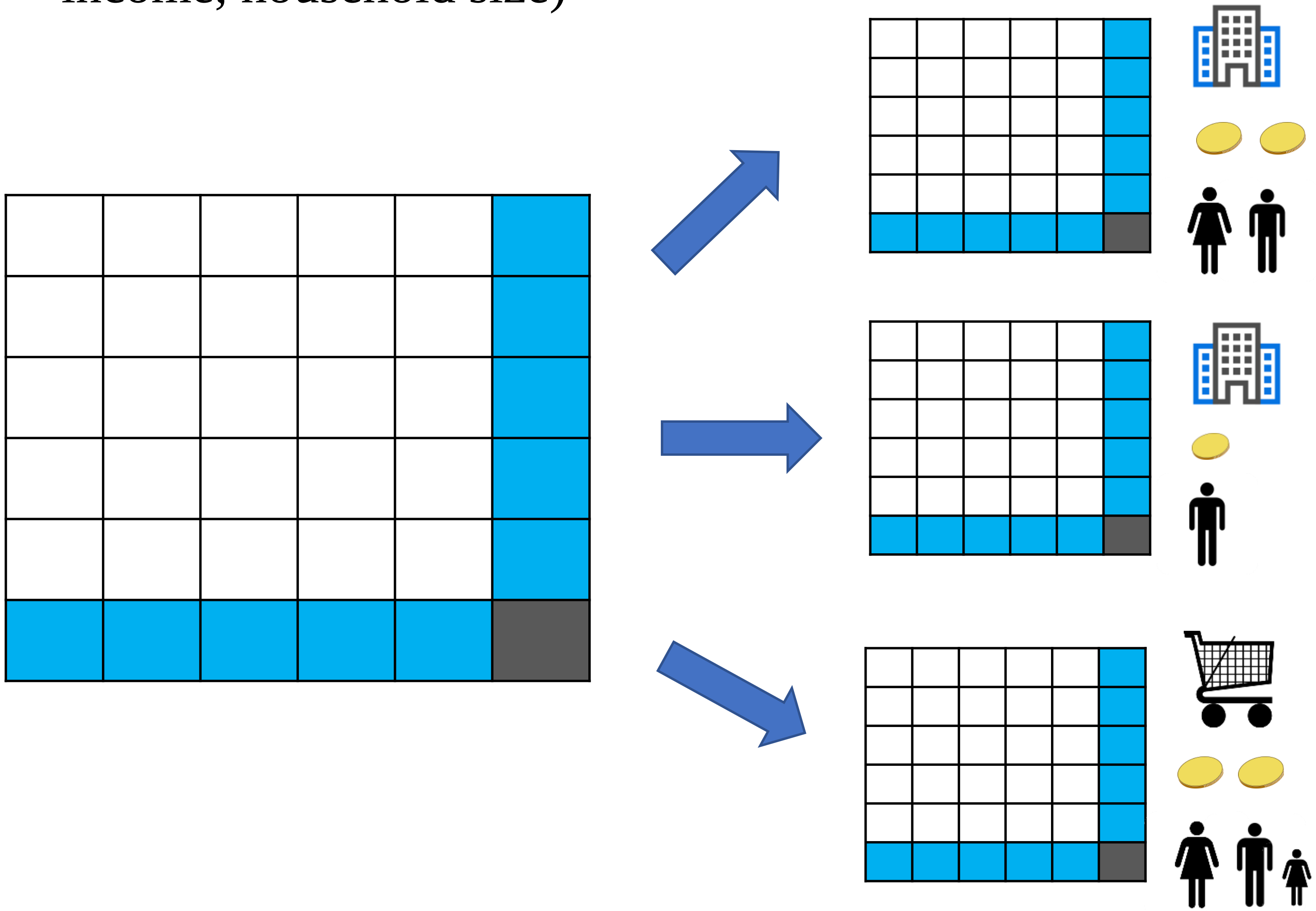
- Resources available are almost always less than the analyst would like.
- The time and resources spent to forecasting/analysis should be analogous to the costs of making a wrong decision.
- Benefits obtained from extra investment in time and resources gradually lessens
  - We will never have a perfect model no matter the amount of effort
- The overall judgement is related to the attitude of the decision makers towards taking risks

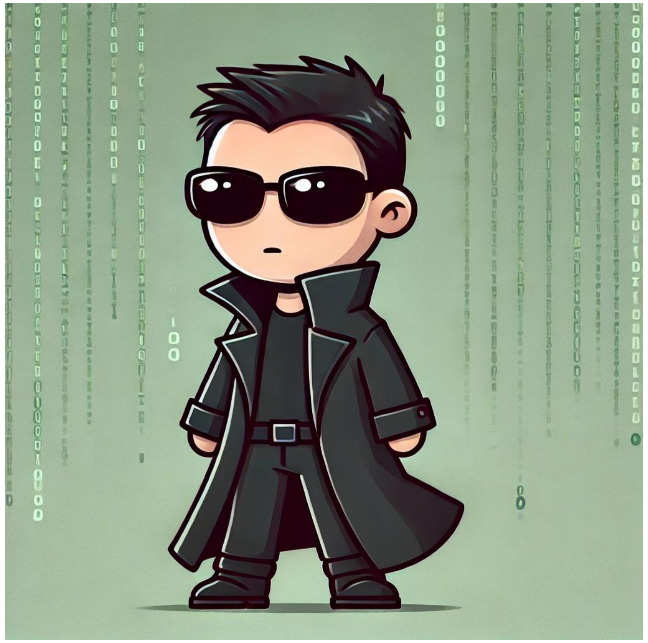
The 4–step model – Matrix representation



The 4-step model – Matrix representation

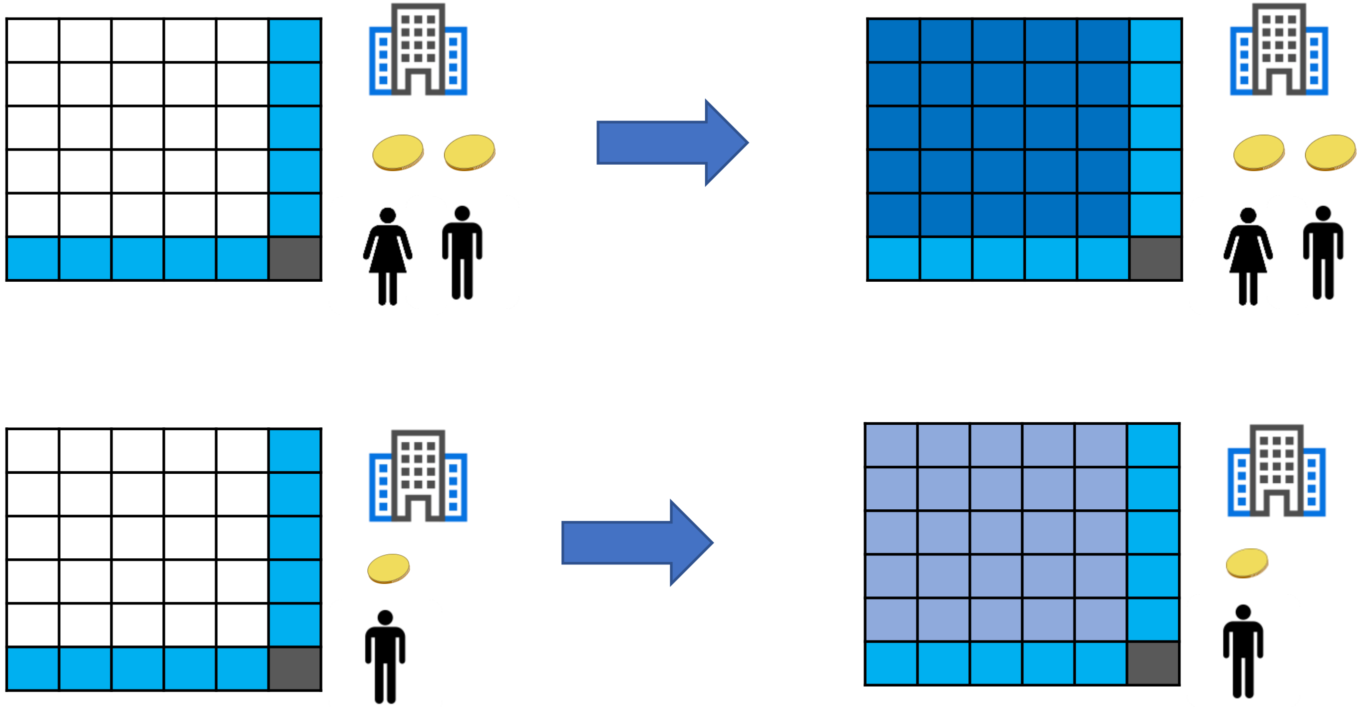
- Step 1: we only know the row and column totals per segment (e.g. trip purpose, income, household size)





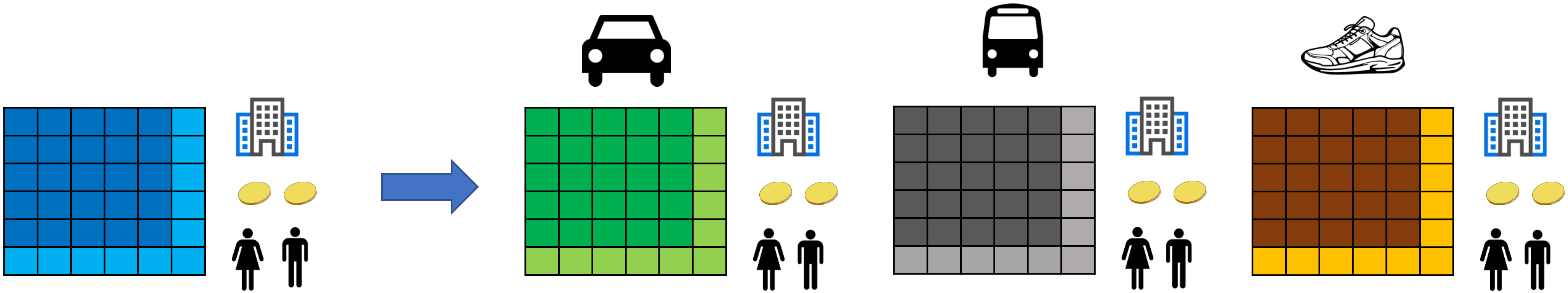
The 4-step model – Matrix representation

- Step 2: For every segment, we fill in the blanks of the OD matrix:
  - The number in each cell represents the number of trips from origin O to destination D
  - The values in the cells should (in theory) sum up to the rows and columns total values from the previous steps



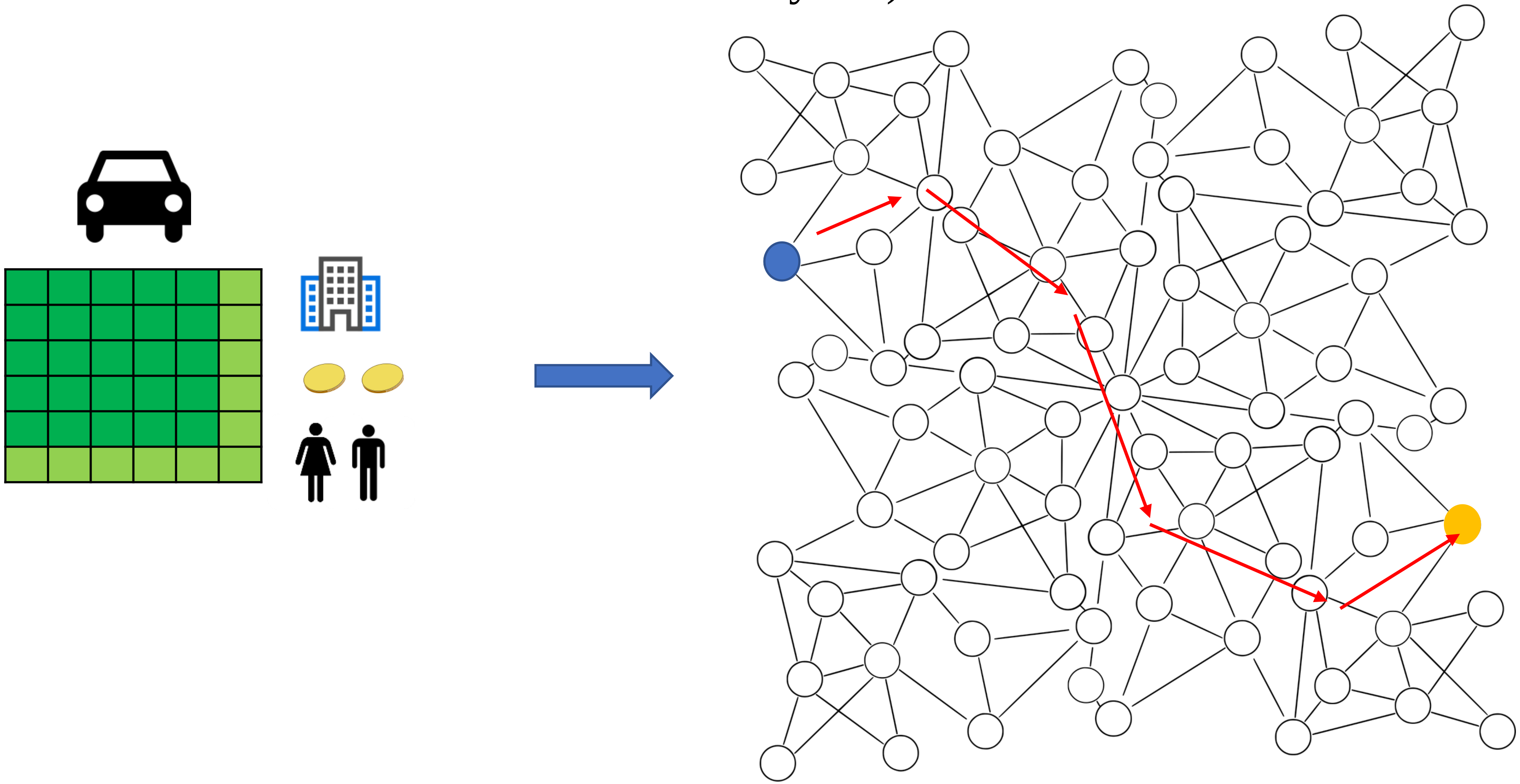
The 4-step model – Matrix representation

- Step 3: For every segment, we split the row and column totals by transport mode



The 4-step model – Matrix representation

- Step 4: We assign each segment for each mode and each OD pair to the network (all segments use the same network – some PT may not)



The time element

- Typically we know the trip generation for the whole 24h period
- Assignment typically focuses on peak hours – help identify the weak points
- Typically use data from specific times to calculate the proportion of daily trips during peak hours
- Some people may shift their departure time to avoid peak hours – ignored in the traditional 4-step model
    - Some models include this “fifth” step, usually just before the assignment step (departure time choice)

The 4-step is not the answer to everything

- Microsimulation: Similar to the assignment model but examine traffic interactions and the individual level. Useful for conducting road safety analysis or understanding congestion better
- Operational models: Allow for accurate representation and modification of the network operators e.g. changing the signal of traffic lights
- Uni-modal models: models that focus on the demand of a single mode of transport.

A word on sampling…

Sampling – statistical considerations

- Data: sample of observations taken from a certain population of interest
- Not economically (or perhaps even technically) feasible to observe the whole population
- Observations made of one or more attributes (say income) of each member of the population

Sampling – statistical considerations

- Sample design: ensuring that the data provide the greatest amount of information about the population, at the lowest possible cost
- The problem: how to use the sample data to make correct inferences about the population

Difficulties:

- how to ensure a representative sample?
- how to extract valid conclusions from a sample?

Data collection - Sample

Sample: A collection of units to represent a larger population with certain attributes of interest (i.e. age, income, etc.).

Considerations:

- Which one is our population?
- What do we mean by ‘especially selected’?
- How large the sample should be?

Data collection - Which one is our population?

Population of Interest: The complete group about which information is sought

- It is composed of individual elements
- The sample is usually selected on the basis of sampling units which may not be equivalent to these individual elements
- Example: a frequently used sampling unit is the household while the elements of interest are individuals residing in it

Data collection - What do we mean by ‘especially selected’?

Sampling Methods

- *Simple random sampling*
    - Enumerate all units in the population and then select numbers at random to obtain the sample
    - Problem: far too large samples may be required to ensure sufficient data about minority options
    - Example: Sampling households at random might provide little information on multiple car ownership

- *Stratified random sampling*
    - *A priori* information is first used to subdivide the population into homogeneous strata
    - Apply simple random sampling inside each stratum using the same sampling rate for all strata
        - The correct proportions for each stratum in the sample is obtained
    - Important for relatively small subgroups in the population as they could lack representation

Sampling – Sampling Error and Sampling Bias

Sampling Error: We use a sample and not the total population; it cannot be avoided due to random effects

- It does not affect the expected values of the means of the estimated parameters but it affects their variability
- It determines the degree of confidence that may be associated with the estimated means
- It is a function of sample size and of the inherent variability of the parameter under investigation

Sampling bias: Mistakes when defining the population of interest, selecting the sampling method, the data collection technique or any other part of the process.

Differences from the sampling error:

- It can affect both the mean and the variability around estimated parameters
- It may be reduced or eliminated by taking extra care during the various stages of sampling design and data collection.

Sampling Error and **Sampling bias** contribute to the measurement error of the data

Sampling – How large the sample should be?

Sample Size: No straightforward answers…

… although sample size calculations are based on statistical formulae:

- Many of their inputs (effect size, confidence levels, variables) are relatively subjective and uncertain
- Must be produced by the analyst after careful consideration of the problem

Trade-offs:

- too large sample requires expensive data-collection and analysis process given objective and required degree of accuracy
- too small sample may imply results which are subject to an unacceptably high degree of uncertainty

Sampling – How large the sample should be?

Factors affecting sample size:

- Variability of the parameters in the population under study,
- Degree of accuracy required for each
- Population size

Sampling – Confidence intervals

- The statistics we estimate from a sample (like the mean or standard deviation of a variable of interest) can vary from sample to sample, even when drawn from the same population.

- In practice, we estimate a sample statistic (like the mean) and use it to infer the population mean, along with a confidence interval that reflects the variability (i.e., the spread of the sampling distribution)

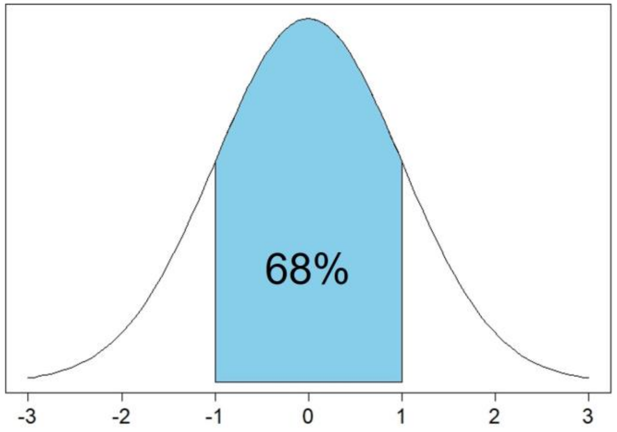
- A common confidence level is **95%**, meaning that if we repeated the sampling many times, about 95% of the resulting confidence intervals would contain the true population mean

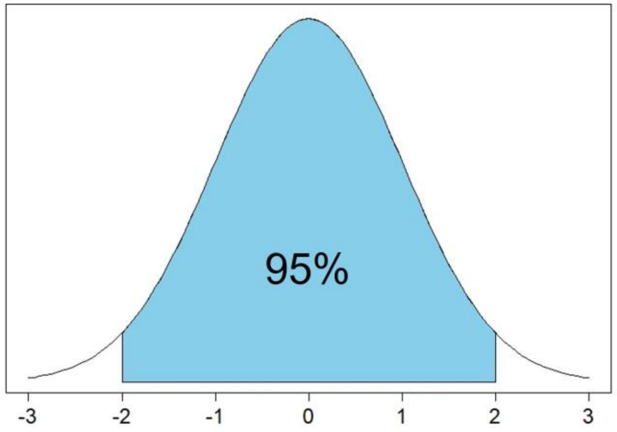
Sampling – Confidence intervals

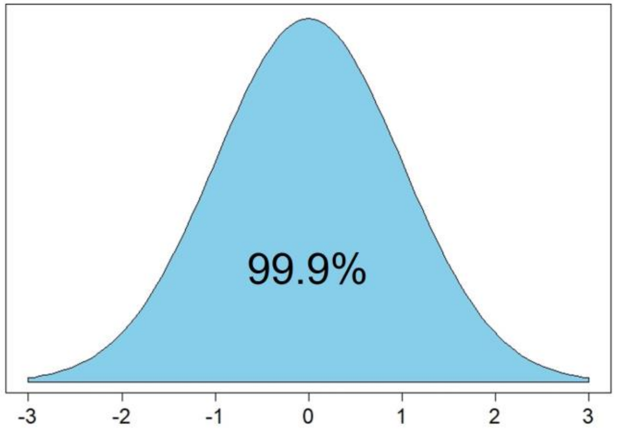
First, let’s remember some properties of the standard normal distribution

N ~ (0,1):

- 68% of the observations are between -1 and 1 standard deviations of the mean
- 95% of the observations are between -2 and 2 standard deviations (-1.96 and 1.96 to be precise)
- 99.9% of the observations are between -3 and 3 standard deviations







Sampling – Confidence intervals

- Central limit theorem (CLT): When a ***sufficiently large random sample*** of size n is drawn from a population of size N with mean μ and standard deviation σ, the sample mean $\bar{x}$ is approximately normally distributed with mean μ and standard error $\sigma/\sqrt{n}$.

- The CLT holds for any population distribution if $n \geq 30$

- The CLT holds in the case of smaller samples, if the original population has a Normal-like distribution.

- **Be careful!** The rule of thumb $n \geq 30$ applies to estimating the mean of a single variable in a population with homogeneous characteristics. E.g. if you are estimating a regression or controlling for several exogenous variables, a much larger sample size is required.

Sampling – Confidence intervals

- Brief reminder: A standard normal variable Z is defined as:

$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

- A standard normal variable Z is with 0.95 probability between the range [-1.96, 1.96] (from the previous slide); then:

$$0.95 = P\left(-1.96 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

- The confidence interval that captures μ with a probability of 0.95 can be rewritten as:

$$\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

Sampling – Confidence intervals

For smaller (or any sample size):

- The standard error is a combination of the formula we saw previously and the finite population correction (FPC) factor (no need for more details about FPC right now)

$$\text{se}(\bar{x}) = \sqrt{\frac{(N-n)\sigma^2}{n(\mathrm{N}-1)}}$$

- If only one sample then (s^2 is the sample variance):

$$\text{se}(\bar{x}) = \sqrt{\frac{(N-n)s^2}{n(\mathrm{N})}}$$

Sampling – Confidence intervals

For smaller (or any sample size):

- For large populations and small sample sizes (the most frequent case) the factor (N - n)/N is very close to 1. Then...

$$\mathrm{se}(\bar{x}) = \sqrt{\frac{(N-n)s^2}{n(\mathrm{N})}} \approx \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$$

... which is the formula that we used earlier

Sampling – Sample size

Using the previous formula, our required sample size (for infinite population) is:

$$n' = \frac{s^2}{se(\bar{x})^2}$$

The sample size for finite population size:

$$n = \frac{n'}{1 + \frac{n'}{\mathrm{N}}}$$

Sampling – Sample size

Issues:

- The sample variance s^2 is only known after we obtain the sample
- Desired degree of confidence using the sample mean as an estimate of the population mean; not a specific standard error value, but an interval around the mean for a given confidence level
    - A confidence level for the interval must be chosen (95% that we mentioned previously); expresses how frequently the analyst is willing to make a mistake by accepting the sample mean as a measure of the population mean
    - Specify the limits of the confidence interval around the mean, either in absolute or relative terms; as the interval is expressed as a proportion of the mean, an estimate of this is required to calculate the absolute values of the interval. Typically we express the sample size as a function of the expected coefficient of variation ($CV = s/\mu$) of the data.

To conclude…

To conclude…

- The definition of transportation modelling
- The purpose of transportation modelling
- Terminology
- Data – Sampling
- Model specification, calibration, and validation
- The 4–step model

What’s next (tentative schedule)

Date	Lecture	Lab session
Tue, 08.04	Introduction to transport modelling	Set up of a traffic network in QGIS
Tue, 15.04	Trip generation models – Trip distribution models	Trip generation models – Trip distribution models
Tue, 29.04	Mode choice models – Traffic assignment	Mode choice models – Traffic assignment in QGIS
Tue, 06.05	Car–following models	Car–following models
Tue, 13.05	Wrap–up, (Lane–changing models) – Project	Guest lecture: Dr. Matthias Hellwig
Tue, 20.05	Project Q&A	
Tue, 27.05	Exam	