July 2021

Transport Modeling: the four step model
SYSTRA

Table of Contents

SYSTRA

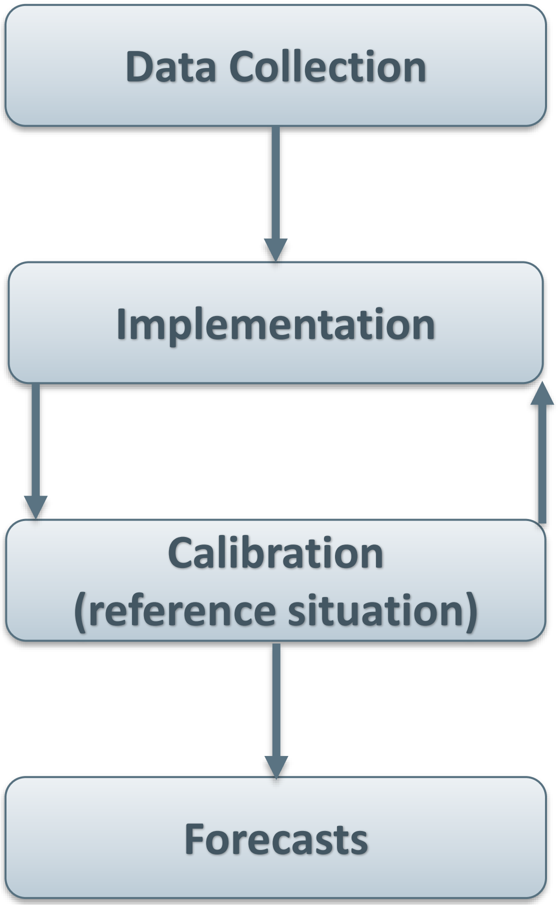
What is the purpose of a model?

- Design of a project:
  - Speed
  - Mode of transport
  - Size of the rolling stock
- Economic and financial appraisal of a project
  - Economic benefits such as time savings
  - Revenues

SYSTRA

A model is a digital copy of the interacting transport systems:

- data collection
  - Input data
  - Calibration data
- Implementation :
  - Modeling of the **digital replicas** of the system
    - **demand system (voyagers)**
    - **supply system (networks)**
  - specification of the interactions between the two systems
- Calibration:
  - Refining the interactions (equations)
  - Parametrization of the equations
- Forecasts

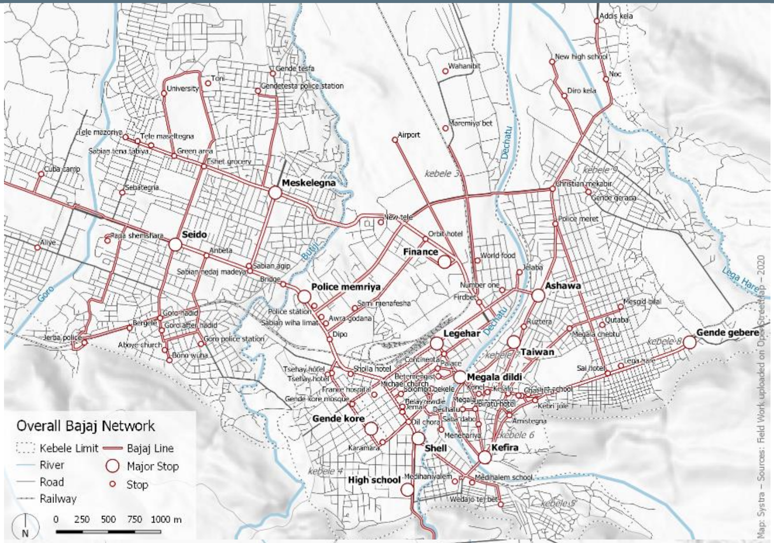


<u>The exogenous data</u> is used to feed the model in the reference scenario. They are stored in organized data structures such as tables and geographical objects

- Networks
  - Road network : speed and capacity
  - Public transport network: speed, stops, capacity and headways
- **Zonal** Socio-economical data
  - Population
  - Motorization rate
  - Jobs, students

<u>The control data</u> is used for the calibration

- Origin destination surveys : based on a **zoning**
- Vehicle counts along the **networks**
- Ticketing data



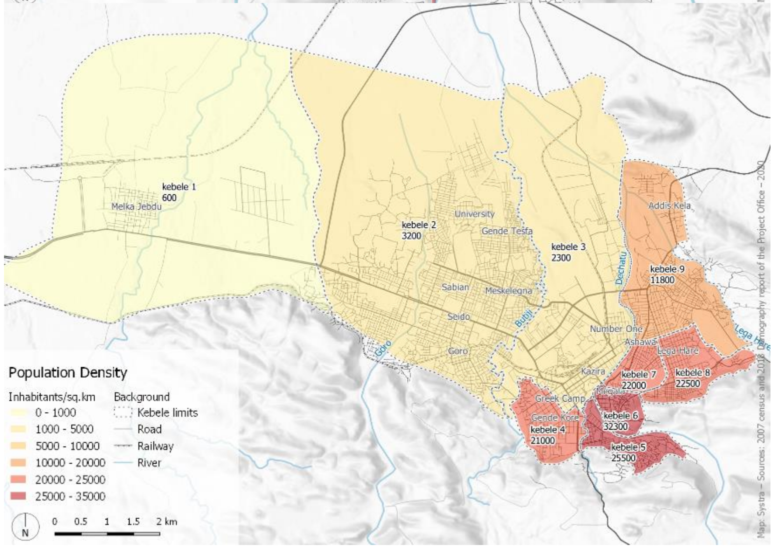
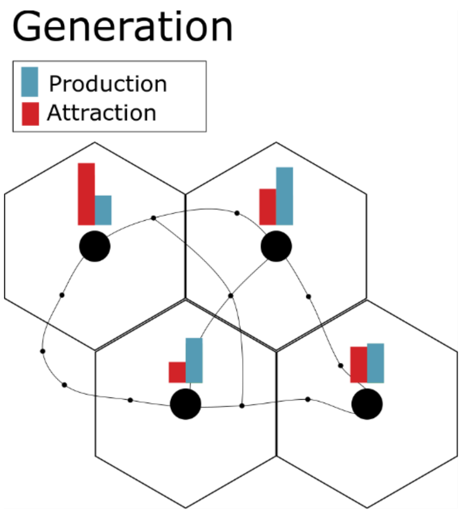


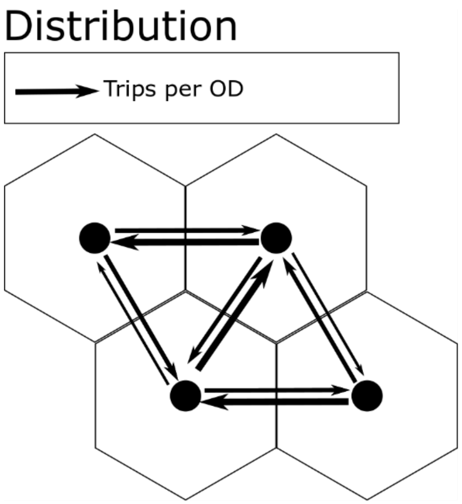
Table of Contents

Four step model
Introduction



1st Step: GENERATION

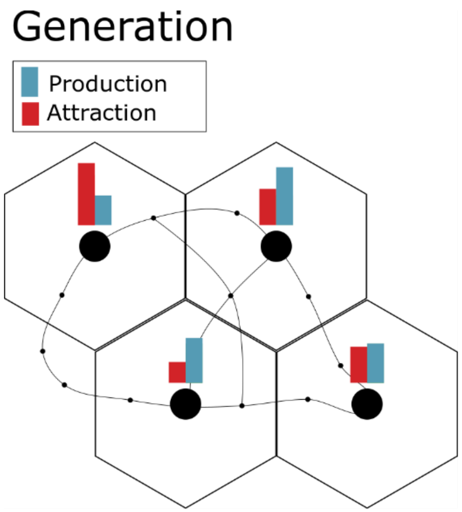
How many trips are produced and attracted by each zone?



2nd Step : DISTRIBUTION

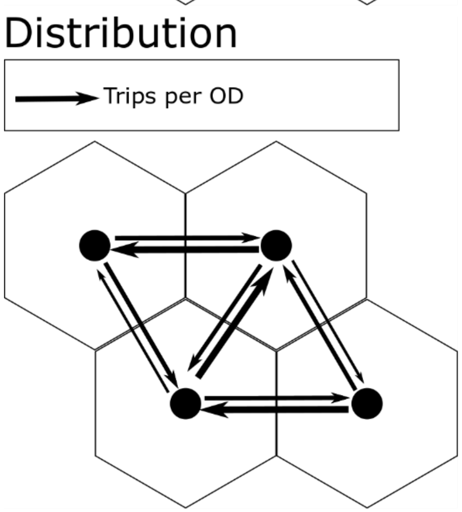
How many trips are they between each pair of Origin and Destination?

Four step model
Introduction



1st Step: GENERATION

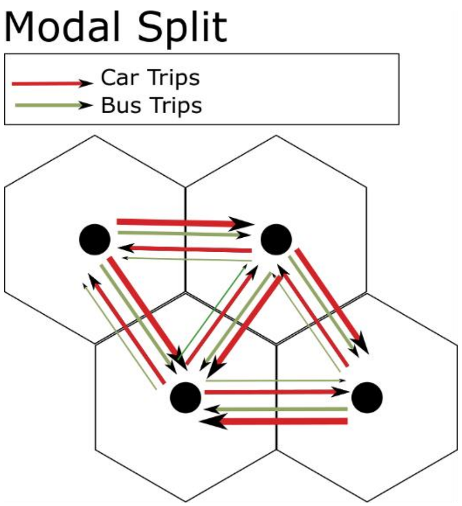
How many trips are produced and attracted by each zone?

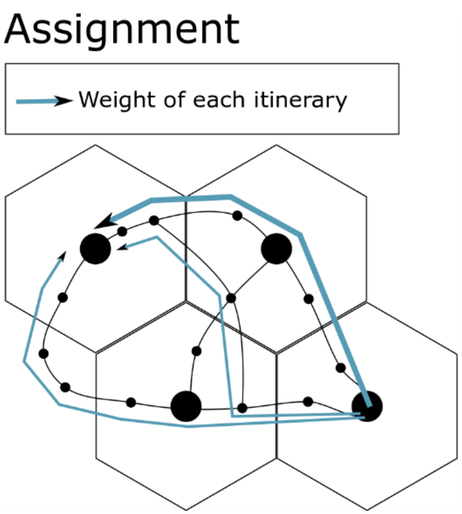


2nd Step : DISTRIBUTION

How many trips are they between each pair of Origin and Destination?

SYSTRA





3rd step: MODAL SPLIT

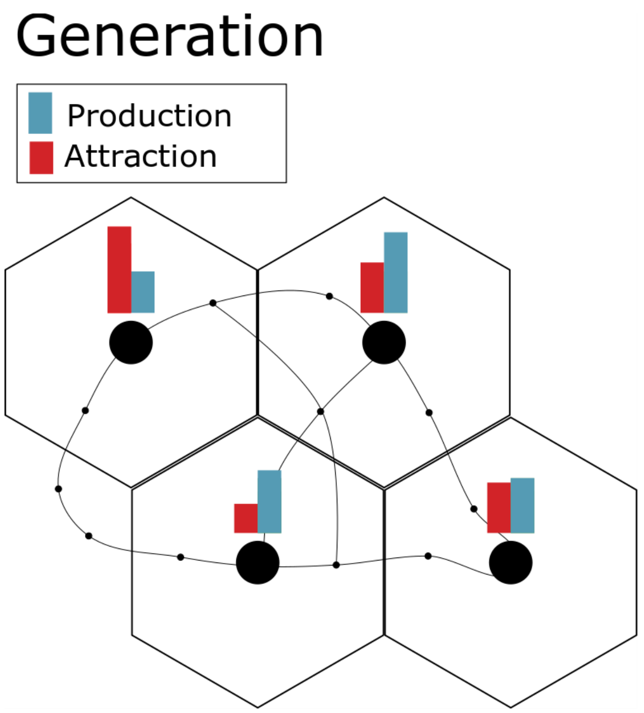
What is the **modal share** (%) of the various modes of transport for each **origin destination** ?

4th step: ASSIGNMENT (itinerary choice)

¿What are the itinerary chosen by the passengers for each mode of transport and each origin destination?

Generation: Mathematical concepts

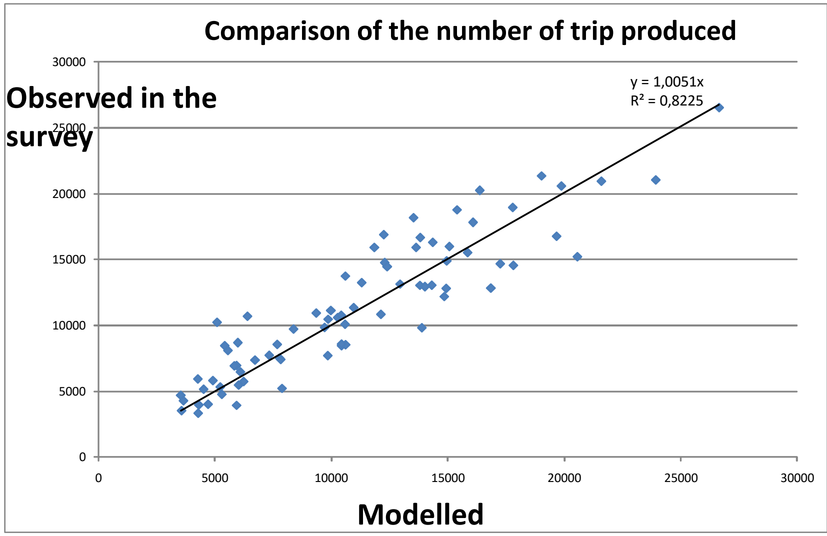
- We try to explain (to model) the number of trips of **any mode produced** and **attracted** by **each zone** .
- Usually, the productions and attractions are explained by **social** and **economical** variables.
- The linear combination is the most common formula:
  - $\text{Emisiones}_i = \sum_j \alpha_j \cdot X_{ij}$
  - $\text{Emisiones}_i = \alpha \cdot Population_i + \beta \cdot \text{Jobs}_i$
- The impact of the accessibility of a zone on its trip generation is called induction. It is hard to quantify and most of the time, we assume that the accessibility of a zone does not impact its productions or attractions.



Generation: calibration

- The calibration data is the **productions and attractions** of the **household survey**
- Sometime, the data is not statistically significant at the detailed zoning level
- When the productions and attractions are not significant at a low level, we work with an aggregated zoning (where each zone contains at least 70 households for instance)
- We compare the calibration data i.e., the number of trips produced and attracted by each zone in the survey with the socio economical values in order to assess the lineal constants α_j

Motif	% PPS	Expression Attractions	R^2 Attraction
Secondaire	23,4%	0,138*EMP + 0,146*POP + 0,0189*COM	0,55
Loi/SanDém-Domicile	11,5%	0,113*POP	0,71
Travail-Domicile	11,1%	0,109*POP	0,77
Achats-Domicile	10,3%	0,104*POP	0,72
Scolaire-Domicile	10,0%	0,097*POP	0,67
Accompagnement-Domicile	9,1%	0,0885*POP	0,56
Domicile-Loi/SanDém	8,6%	0,0634 * POP + 0,0412 * EMP	0,42
Domicile-Accompagnement	6,3%	0,061*POP	0,51
Domicile-Achats	5,1%	0,00983*COM + 0,0262*POP + 0,0363 *EMP	0,44
Travail-Accompagnement	2,9%	0,027*POP + 0,002*COM	0,41



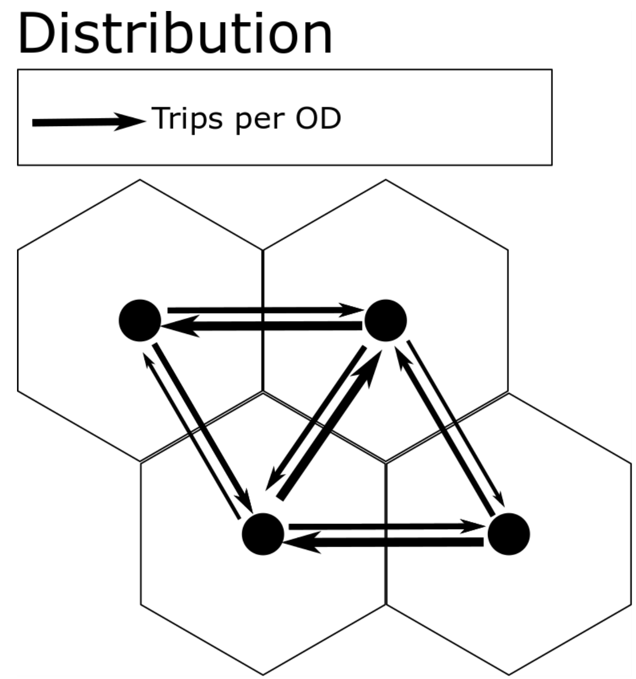
SYSTRA

Four step model

Generation | **Distribution** | Modal Split | Assignment

Distribution: Mathematical Concepts

- We try to explain the flows between each origin and each destination
- The gravitational distribution is the most common formula
  - $T_{ij} = K_i \frac{E_i \cdot A_j}{\mathrm{f}(c_{ij})}$ for a simply constrained model
  - f : deterrence function, depends on the time and price for example
    - $f(cij) = cij^{\alpha}$
    - $f(cij) = e^{\beta cij}$
    - $f(cij) = cij^{\alpha}\, e^{\beta cij}$
  - K_i : adjustment parameter to meet with the constraints. It enables for the sum of the trips originating from a zone to be equal to the productions calculated in the generation step
    - $\sum_{\mathrm{j}} T_{ij} = E_i$
    - $K_i = {}^{1}\!\Big/_{\sum_{\mathrm{j}} \frac{A_j}{\mathrm{f}(c_{ij})}}$
  - A doubly constrained distribution can be used when we need to meet with the attraction constraints.



	Z_1	**Z_2**	**…**	**Z_n**
Z_1	5	8	…	14
Z_2	7	10	…	1
…	…	…	…	…
Z_n	7	2	…	1

SYSTRA

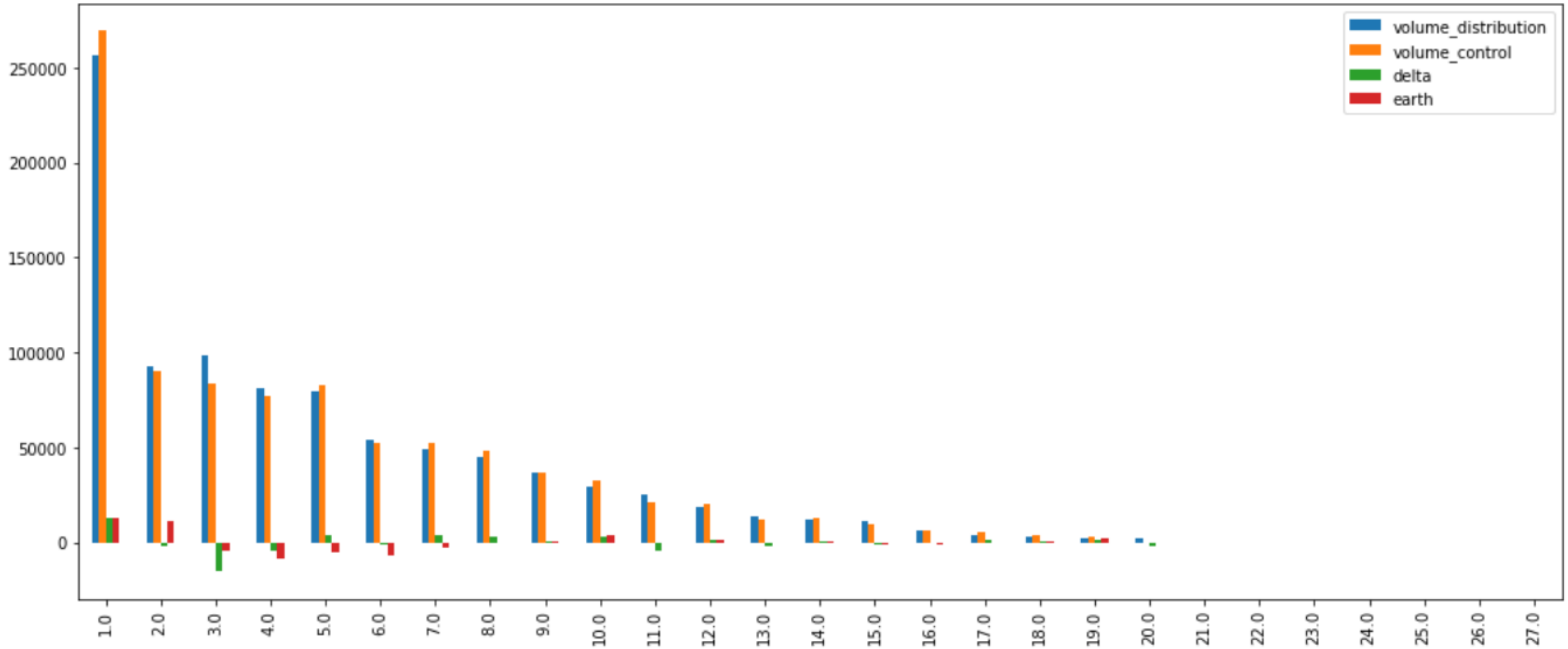
Distribution: Calibration

- **The final aim is to compare the distributed matrix with a reference matrix**
- Usually, the reference matrix is calculated from the household survey
- The sampling of the household survey does not enable the statistical significance of each origin destination cell of the matrix
- For example, with a household survey in 1000 households:
  - 4000 inhabitants surveyed in 200 zones
  - With an average of two trips per inhabitant, we have 8 000 trips
  - And we have 200 origins * 200 destinations = 40 000 OD
  - We do not even have 1 trip per OD and many cells have 0 trips

	Z_1	Z_2	...	Z_n
Z_1	5	8	...	14
Z_2	7	10	...	1
...	...	...	...	...
Z_n	7	2	...	1

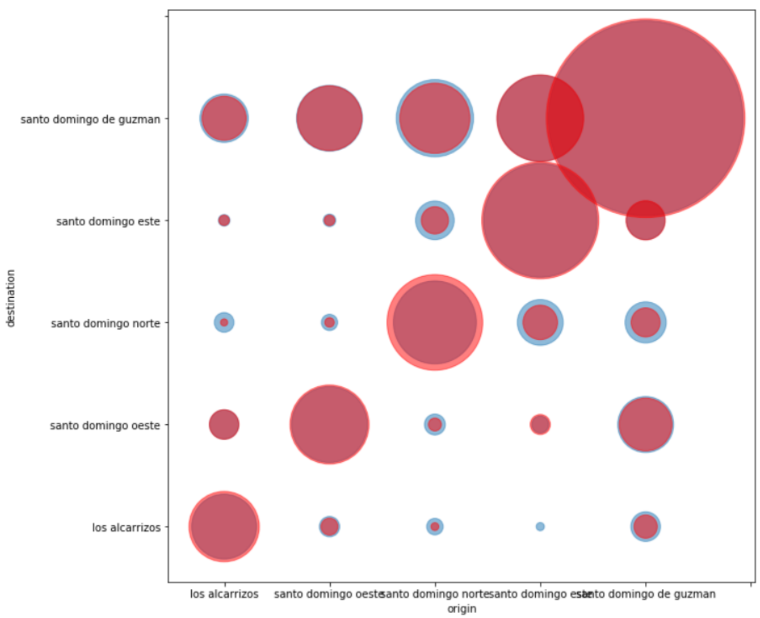
Distribution: Calibration

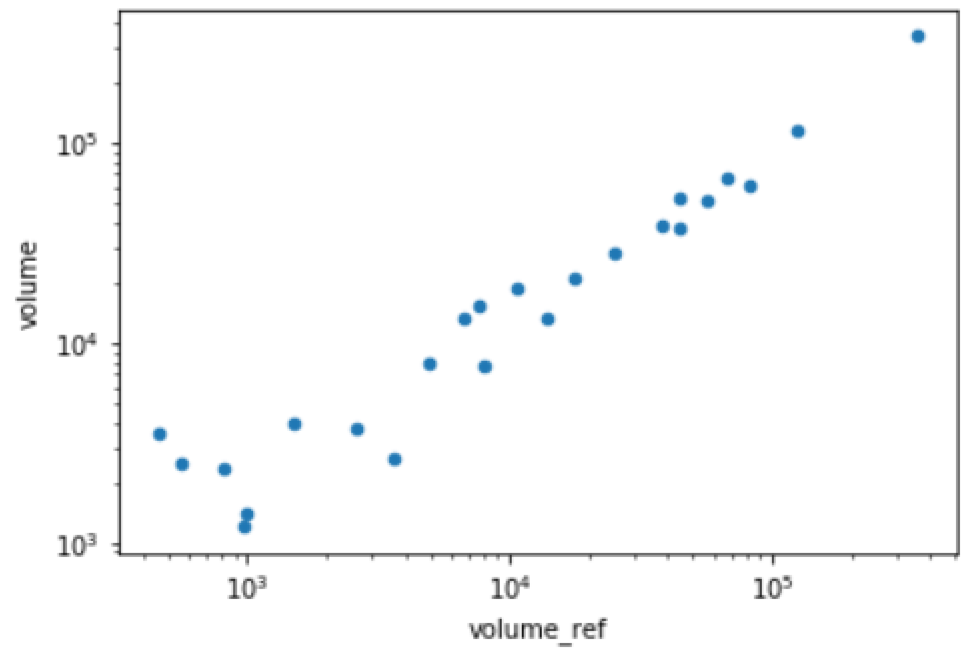
- We cannot compare the flows OD by OD and we must work with an aggregated indicator.
  - We often use the distribution curve of the number of trips by distance class (or time class)
  - We chose the parameters of the deterrence function $f(cij) = cij^{\alpha} e^{\beta cij}$ to calibrate the distribution
  - Additionally, we can compare reference and modeled flows on an aggregated zoning (~10 macro zones)



Distribution: Calibration

- We cannot compare the flows OD by OD and we must work with an aggregated indicator.
  - We often use the distribution curve of the number of trips by distance class (or time class)
  - We chose the parameters of the deterrence function $f(cij) = cij^{\alpha}\, e^{\beta cij}$ to calibrate the distribution
  - Additionally, we can compare reference and modeled flows on an aggregated zoning (~10 macro zones)





Four step model

Generation | **Distribution** | Modal Split | Assignment

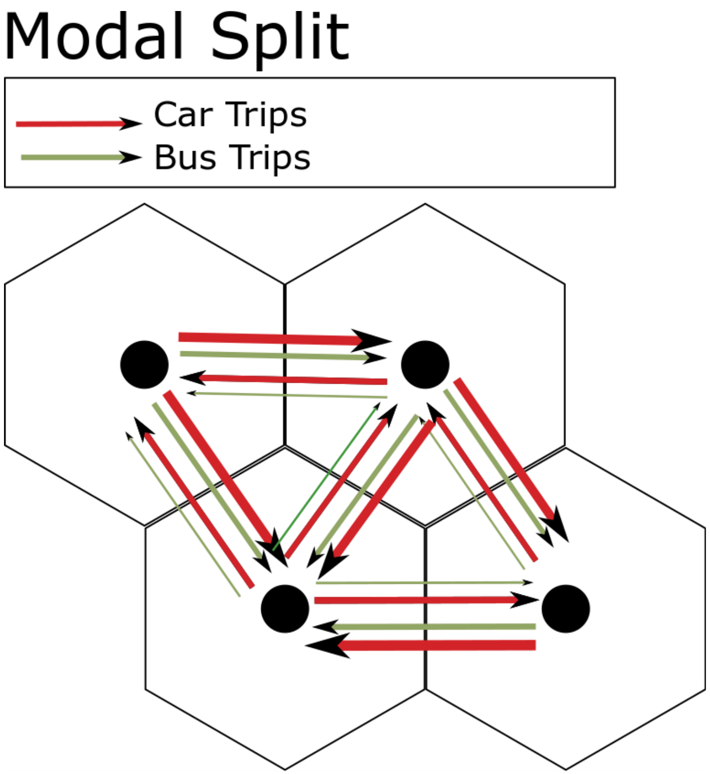
The FRATAR algorithm computes the growth of an OD matrix consistently with the growth of the productions and attractions modelled in the generation step

The FRATAR algorithm is an iterative process, at each iteration, the celles of the matrix are multiplicated by the growth rate of the productions of the origin zone and the growth rate of the attraction of the destination zone

Modal split: Mathematical Concepts

- We want to explain the mode choice for each OD
- We use a logit model based on the utility which measures the satisfaction of the user of each mode
- We compare the utilities of the modes to compute its share
- The higher the share of a mode (compared to the other mode), the higher its' share
- **Logit** :

$$\% \, bus = \frac{e^{U_{bus}}}{e^{U_{bus}} + e^{walk}} = \frac{1}{1 + e^{(U_{walk} - U_{bus})}}$$



Modal Split: the utility function

- What is in the utility function?
  - It must show the user satisfaction
  - The price of a trip
  - The duration of the trip that can be broken down between
    - The waiting time
    - The access and egress time
    - The in vehicle time
  - Everything that is relevant to the user's choice
  - The comfort of the mode
  - We compute an aggregated average utility for each origin destination pair:

$$U_{mode,ij} = \alpha_{mode} + \beta_{mode} \cdot Time_{mode,ij} + \gamma_{mode} \cdot Fare_{mode,ij} + \ldots$$

  - We suppose that the users chose the mode with the highest utility

Modal Split: the Logit Model

- **Strong hypothesis** « The user are rational and chose the mode with the highest utility»
- But if we assume that all the users have the same utility:
  - All the users are going to chose the same mode
  - In the real world, when there is an alternative, there are always a part of the voyager that chose it even if it seems worse
  - Theis is a need for some variability in the perception of the users

- Therefore, for each user traveling from zone i to zone j we consider :
  - $U_{mode,\text{user},ij} = U_{mode,ij} + \varepsilon_{mode,\text{user},ij}$
  - Then:

$$P(user\ chose\ \text{train}\ \ i\ \text{to}\ j) = P\left(U_{\text{train,user},ij} > U_{\text{car,user},ij}\right) = \text{P}(\varepsilon_{\text{train,user},ij}\text{-}\varepsilon_{\text{car,usen},ij} > U_{\text{car},ij} - U_{\text{train},ij})$$

  - The logit formulation comes from a specific probability distribution term : ε

Modal Split: Calibration

- In order to calibrate i.e. chose the coefficients of the utility function we need:
  - The fares, travel times and headways of the modes for each OD in reference situation (exogenous data)
  - The current choice of the users
    - The modal share of the users per origin destination (aggregated calibration)
    - Or surveys where the choice of the mode is asked for a large sample of users with different trip patterns
- The surveys can deal with
  - The current behavior in terms of mode choice: **revealed preference**
  - Expected choices (for instance when there is a new mode of transport) : **stated preference**

- The calibration is based on the maximum likelihood calculated by a statistical software such as Biogeme or R

Four step model

What do we have right after the modal split?

Car matrix

	Z_1	Z_2	…	Z_n
Z_1	5	1	…	0
Z_2	0	9	…	0
…	…	…	…	…
Z_n	0	0	…	1

Public transport matrix

	Z_1	Z_2	…	Z_n
Z_1	0	4	…	10
Z_2	3	1	…	1
…	…	…	…	…
Z_n	5	2	…	0

	Z_1	Z_2	…	Z_n
Z_1	5	8	…	14
Z_2	7	10	…	1
…	…	…	…	…
Z_n	7	2	…	1

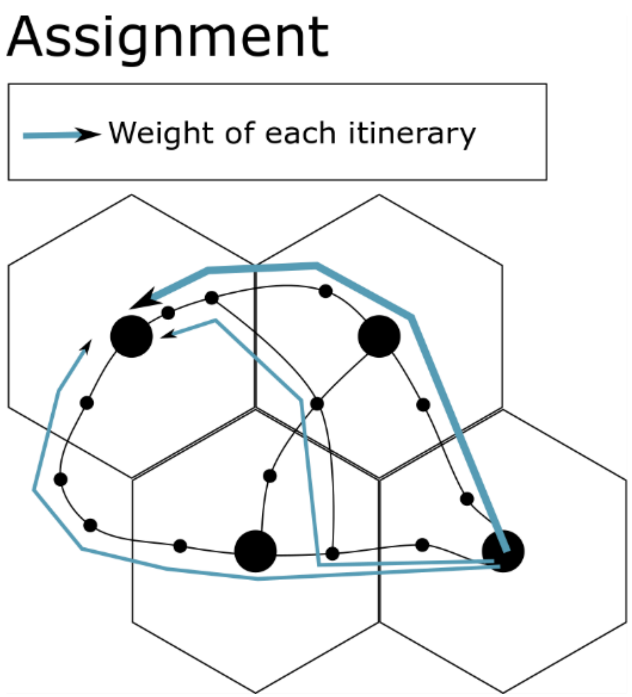
all modes matrix
(= after distribution)

	Z_1	Z_2	…	Z_n
Z_1	0	3	…	4
Z_2	4	0	…	0
…	…	…	…	…
Z_n	2	0	…	0

Walk matrix

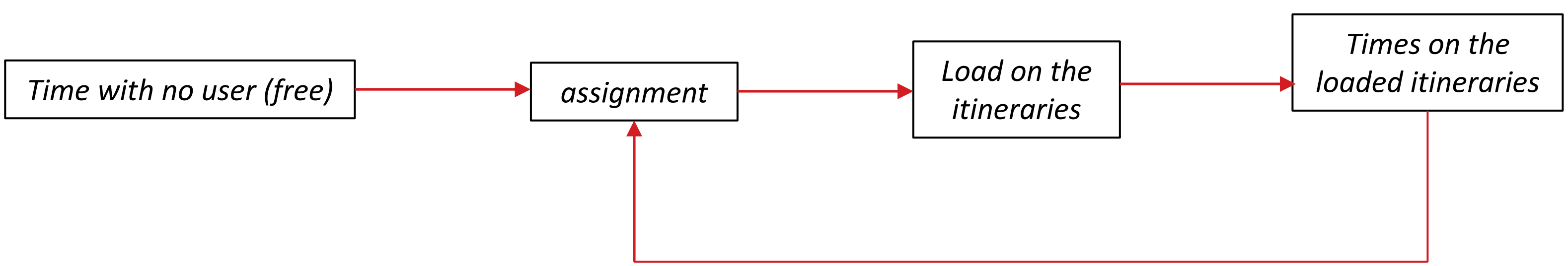
Assignment: Mathematical Concepts

- We want to know the itinerary of the user
  - From every origin zone to every destination zone
  - For each mode of transport
- Every user choses between various options
  - We can use a logit function like in the modal split
  - Or we can compute the probabilities of the paths with the optimal strategy algorithm



Assignment : capacity constraints

- We need to consider the capacity constraints of the roads and transport lines:
  - When an itinerary is attractive, its share rises
  - Consequently, many users are going to use it and it will create jams or discomfort in the public transport
  - The itinerary is going to be less attractive
  - It share is going to go down
  - The jams are going to disappear, and it will become attractive again … etc
- We must run a converging iterative process to assess the equilibrium share of the itineraries
  - The process stops when every itinerary from i to j has the same utility (or time):
  - This state is called the first Wardrop principle (or equilibrium)



Assignment: Calibration

- We want to calibrate the pathfinder parameters and the utility function of the itineraries
  - Transfer penalty
  - Perception of the access or waiting time compared to the in vehicle time
  - Speed of the public transport
  - Speed and capacity of the roads

- The calibration is based on aggregated and disaggregated data
  - Vehicle and voyager counts at the public transport stops and along the road
  - The average number of transfer for the public transport user
  - The daily passengers of the various lines

- The choice of the utility and pathfinder parameters is done empirically. In order to equalize the reference values with the model results.

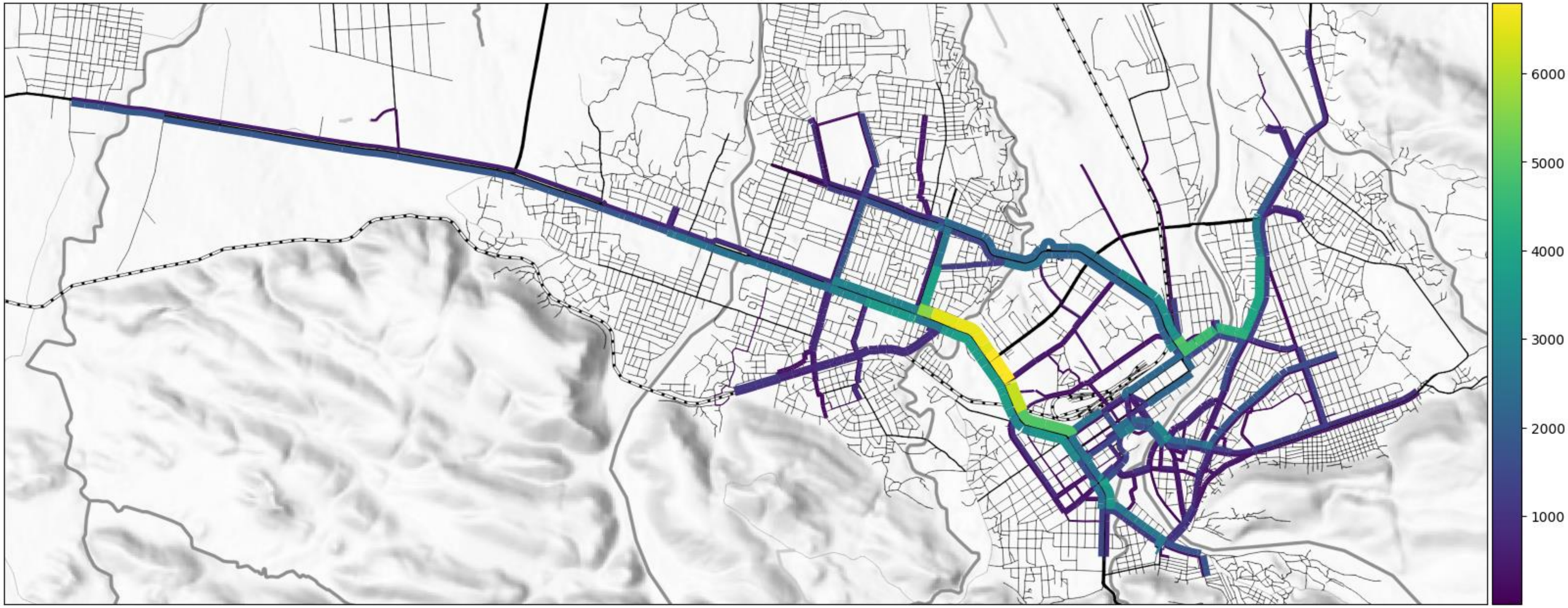
Four step model
Results

- Boarding passengers and flows
- Time an cost perceived by the users
- Saturation maps
- Modal share

Day Average

indicator	day average
time (min)	42.0
distance (km as the crow flies)	4.7
expenses (birr)	9.0
avg # of correspondences	1.5
total generated trips	480510.0

Flows in base (mode=pt)



SYSTRA

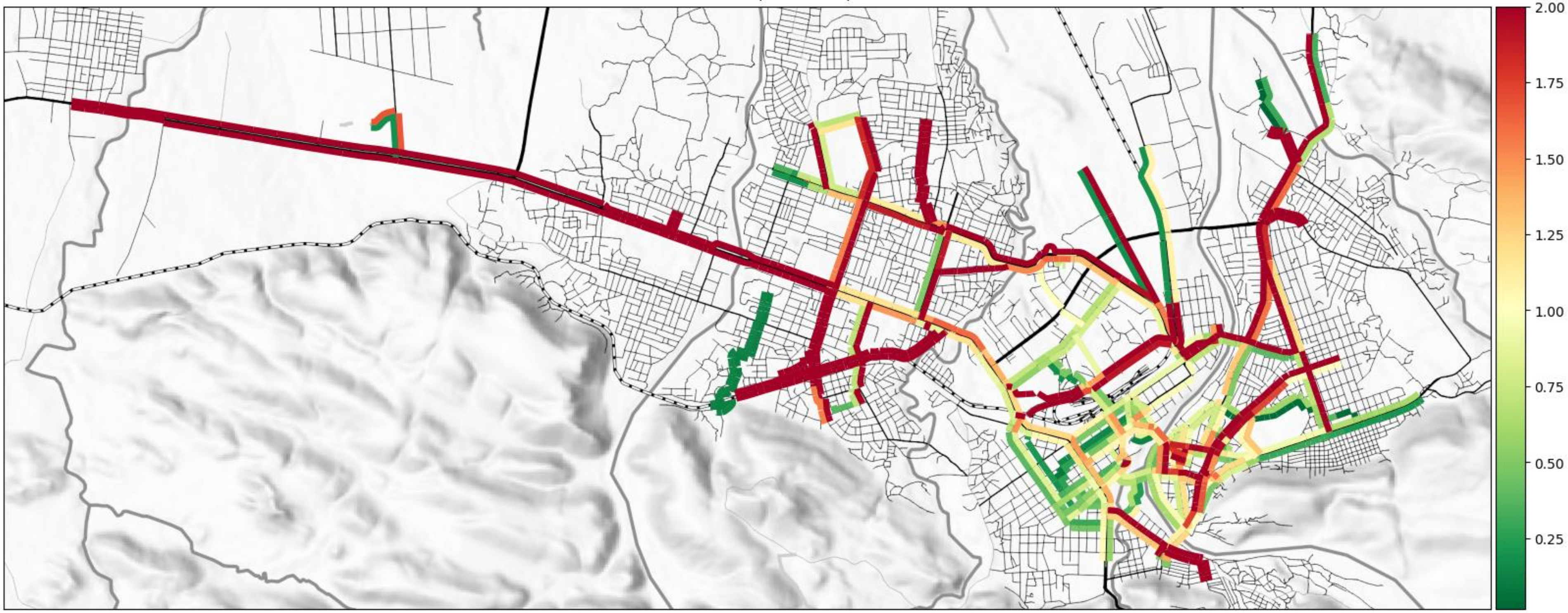
Four step model
Results

- Boarding passengers and flows
- Time an cost perceived by the users
- Saturation maps
- Modal share

Day KPIs by mode

mode	km	vehicles	vehicle_km	voy_km	seat_km	max_load	boardings
bajaj	70	1062	80900	12223000	8086000	1800	284000
total	70	1062	80900	12223000	8086000	1800	284000

Flows in base (saturation)

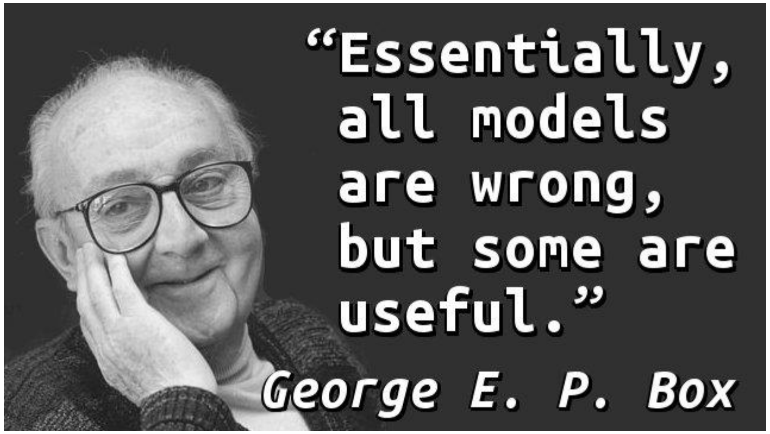


SYSTRA

Table of Contents

SYSTRA

Limits of the models



Type	Error	Example
Relevancy	Choice of the explicative variables	Variables of the generation and distribution functions for example
Relevancy / Consistency	Formulation of the relations	The choice of the deterrence function in the distribution step for example.
Consistency	Calibration (Hypothesis not respected)	The modelled and measured flows along an axis are not equal
Measure	Input data	Sampling error,

SYSTRA