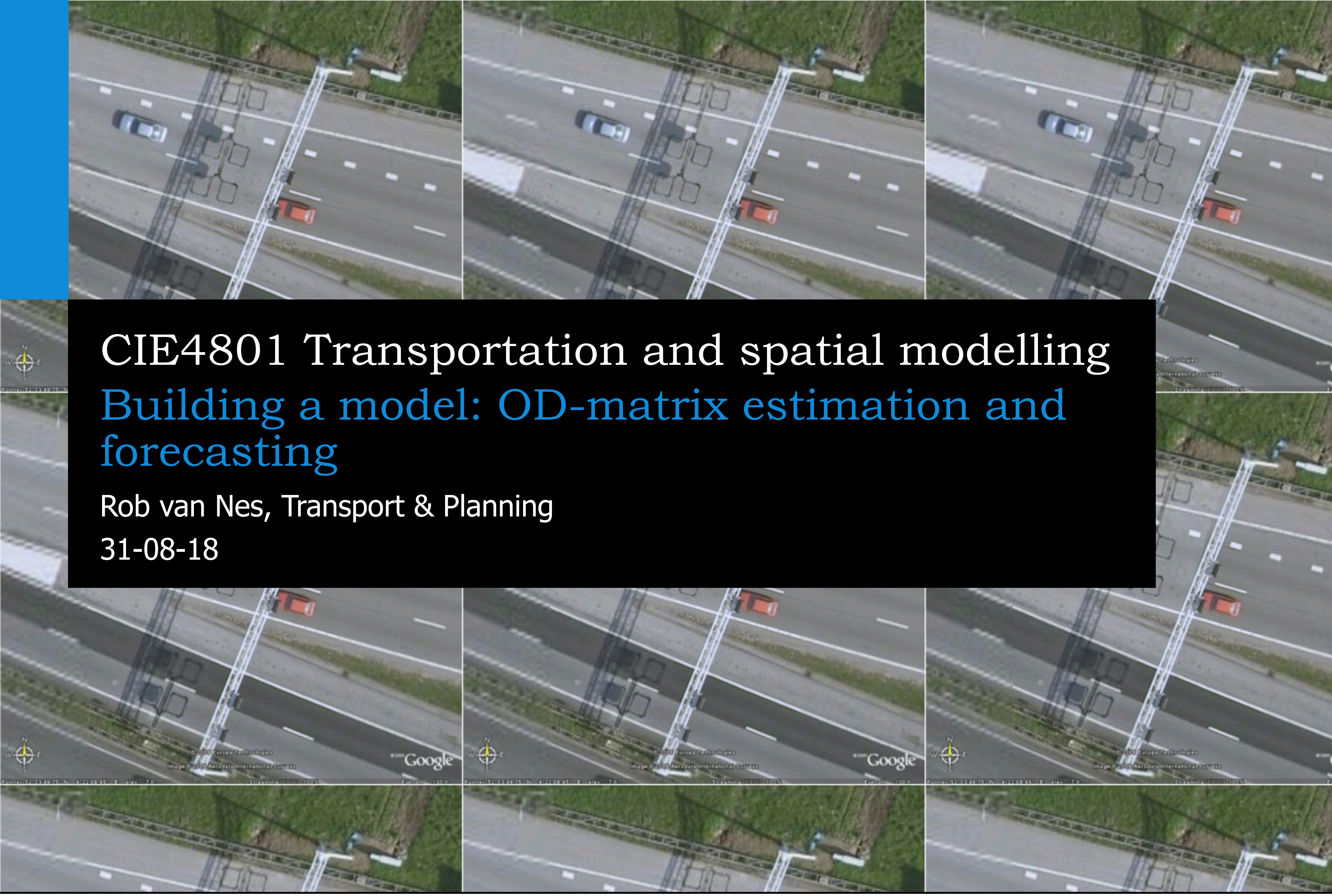


CIE4801 Transportation and spatial modelling

Building a model: OD-matrix estimation and forecasting

Rob van Nes, Transport & Planning

31-08-18

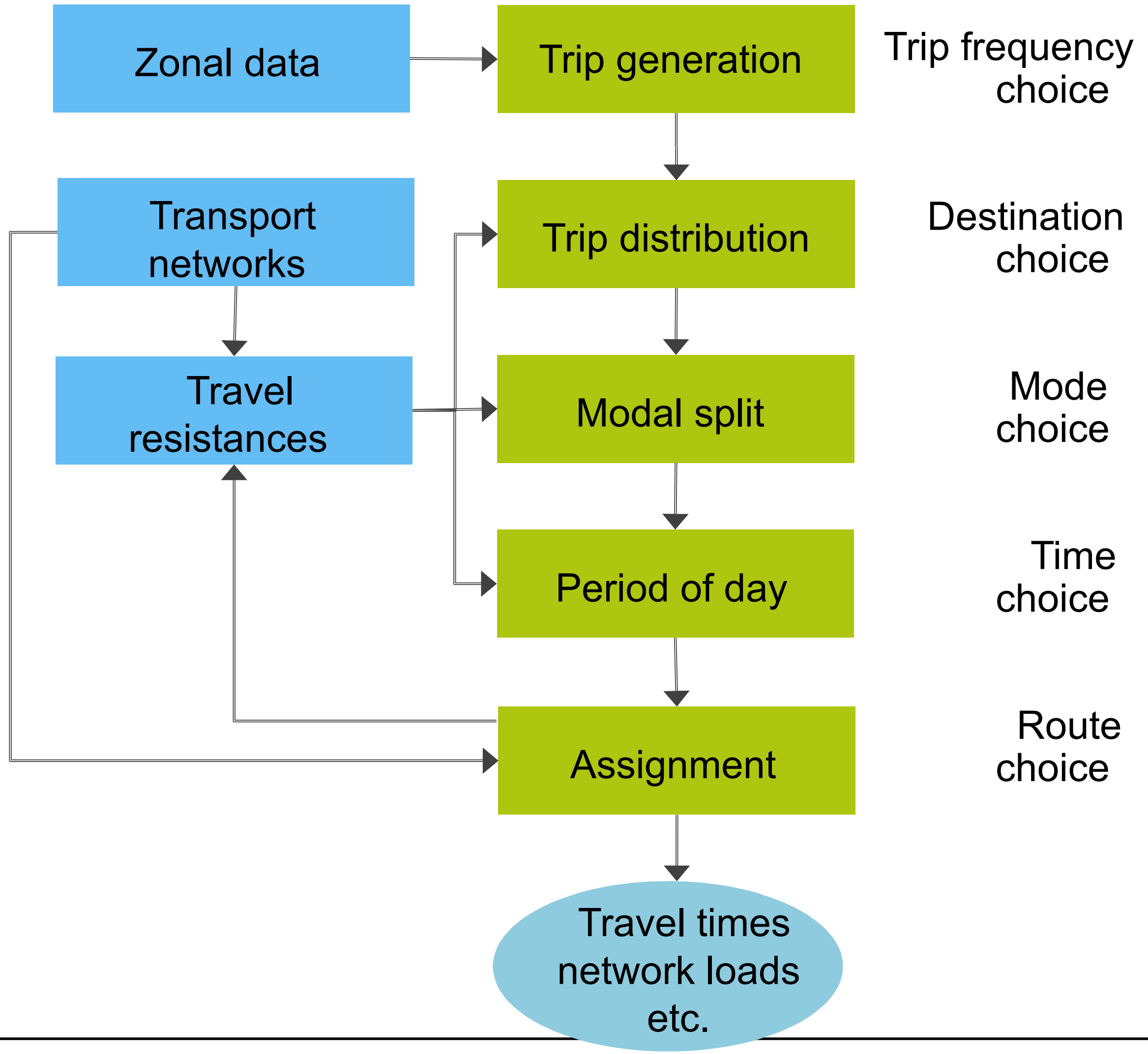
Contents

- Building models:
  - Building a model
  - Estimating OD-matrices using counts
  - Forecasting

2.

Building a model

Overview framework



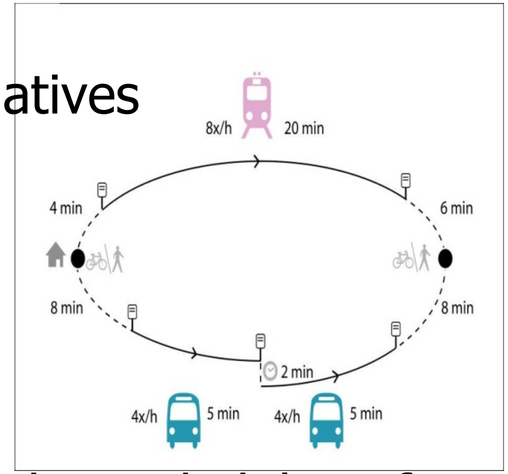
Define your model

- Define study area
  - Cordon model or including influence and external areas?
- Define relevant mode(s), trip purpose(s) and time period(s)
- Define relevant level of detail
  - Network and zoning
- Define model type
  - Descriptive or choice model
  - Determine model formulations

Collect data

- Zone data
  - In case of cordon model Production and attraction for external zones
- Network data
- External data
  - E.g. Through traffic or OD-matrix trucks
- In case of choice models
  - Data on travel behaviour
- Else
  - Data to determine parameters (e.g. travel surveys)
- Note that simply copying functions and parameters is very tricky!

Determine choice functions



- Observe choices as well as the considered alternatives
  - Revealed preference: actual behaviour
  - Stated preference: choice experiments
- Formulate utility functions
- Estimate parameters in utility function such that the probability of the chosen alternative is highest
  - Maximum likelihood method (or log-likelihood)
- Check the quality of the model
  - e.g. $\rho^2 \; \left(\approx R^2\right)$, sign and t-statistics of the parameters
- Dedicated software such as BIOGEME (e.g. see CIE4831, SPM4612)

Determine trip generation functions

- Linear regression
  - Or Cross classification
- Usually at a more aggregate level than zones
  - e.g. Municipality
- Check quality of the model
  - R^2 and t-statistics of the parameters
- See Omnitrans exercise 2

Determine cost functions for the skims

- Specify generalised cost function
  - One single specification
    or
  - Trip purpose specific specifications
    - For instance when time and costs are included (VoT!)

Determine distribution functions

Hyman's method

Given:

- Cost matrix
- Production and attraction
- Mean trip length

Assumed:

- Exponential distribution function

Estimated:

- Q_i, X_j and parameter distribution function

Poisson model

Given:

- Cost matrix
- (partial) observed OD-matrix

Thus also known:

- (partial) production and attraction
- Totals per class for the travel costs

Estimated:

Q_i, X_j and F_k

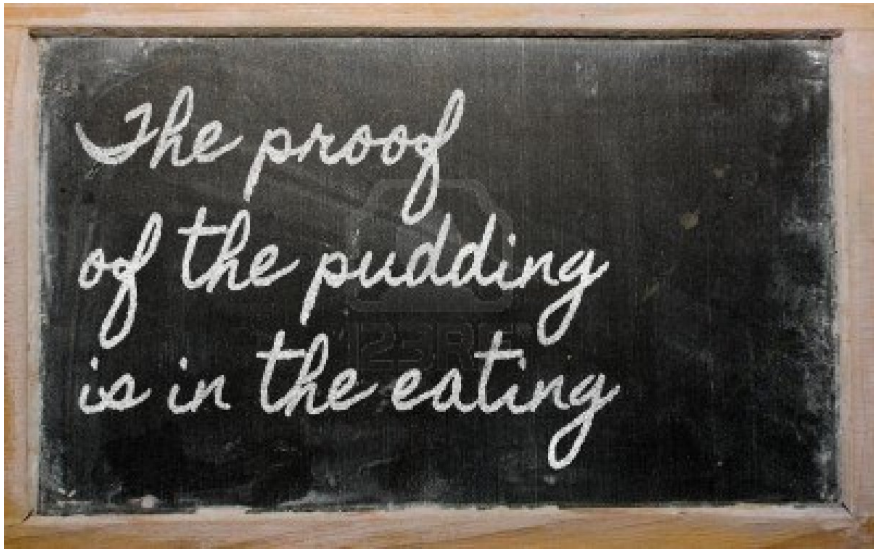
(Latter can be used to estimate distribution functions)

Determine travel time functions

- BPR-function is mostly used
- Different road types have different parameters
  - Especially the value of α
- Literature suggests some other functions based on either theoretical requirements or on computational advantages (or both)

Run the model

- First check and double-check everything
- Run the 4-step model
  - In the right order ;)
- Analyse the results: are they OK?
- When is it OK?



The "eating of the pudding"

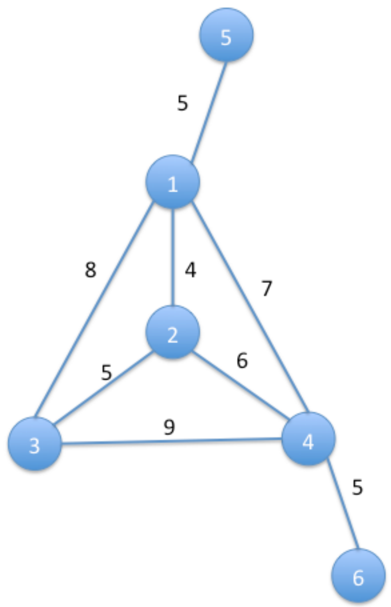
- Key question then is:
    - Does it match the counts?
- Common finding
    - There are quite some differences
- Two options:
    - Check the work done (especially the network!) and build a new (better) matrix
    - Adjust the OD-matrix itself to achieve a better match

3.

OD-estimation

Example

Given a network and an OD-matrix



	1	2	3	4	5	6
1	62	125	62	20	138	91
2	125	121	125	8	223	147
3	62	125	62	20	138	91
4	20	8	20	3	89	59
5	138	223	138	89	0	811
6	91	147	91	59	811	0

What is the flow on the link from node 1 to 4?

=> 20+91+89+811=1011

Let's assume we have a count for that link: 900 vehicles/hour.
How would you adapt the OD-matrix?

And what would you do if a second count but now for for the link from 4 to 6 is available: 1400 instead of 1199 as computed?

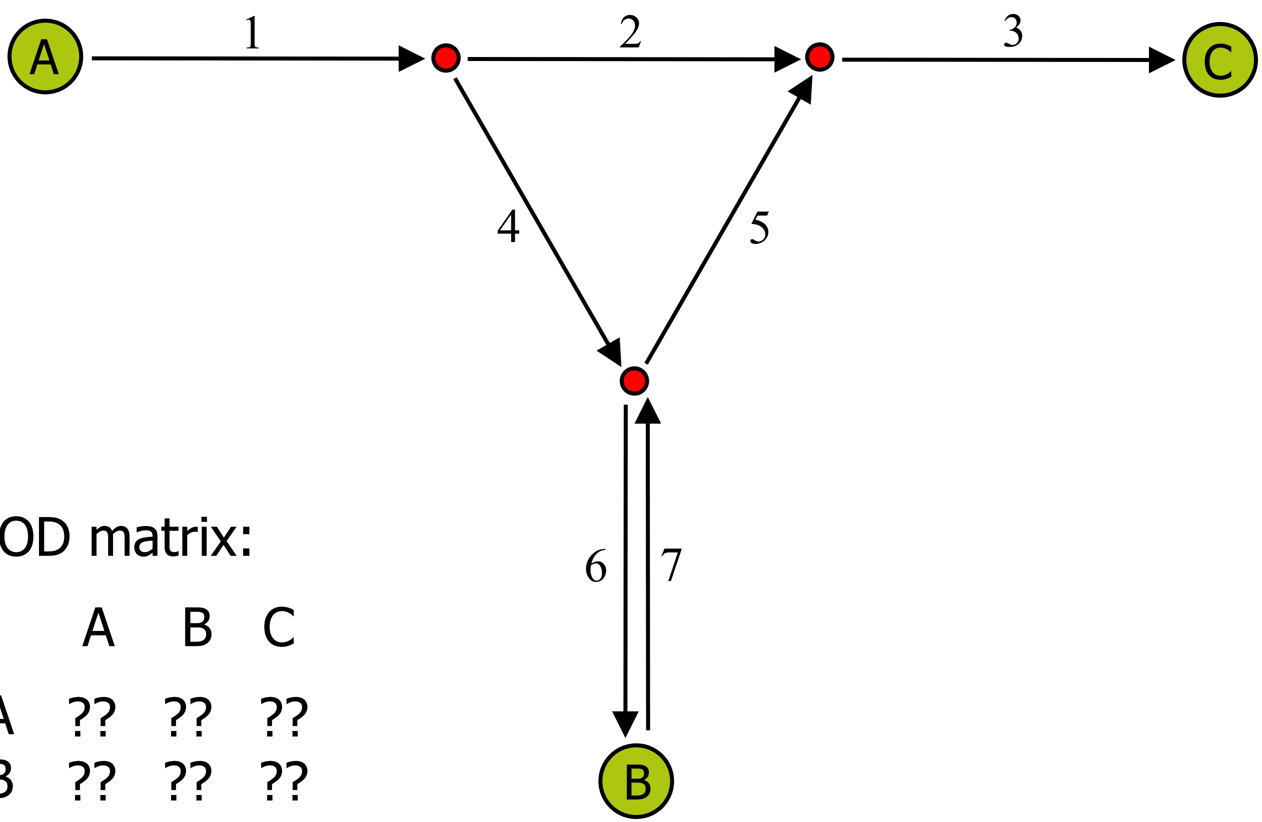
Topics for discussion

- What does this modelling component do? What's its output and what's its input? How does it fit in the framework?
- Important notions
  - Independency of counts, consistency
- Concept of OD-estimation?
  - Estimation of an OD-matrix from counts only?
  - The basic application: multiplier per count
  - Similarity between matrix estimation method (ME2) and the entropy model
  - Concept of Pivot-point modelling
- Practical issues
- Are these models appropriate?

3.1

Estimating OD-matrices
Updating using traffic counts

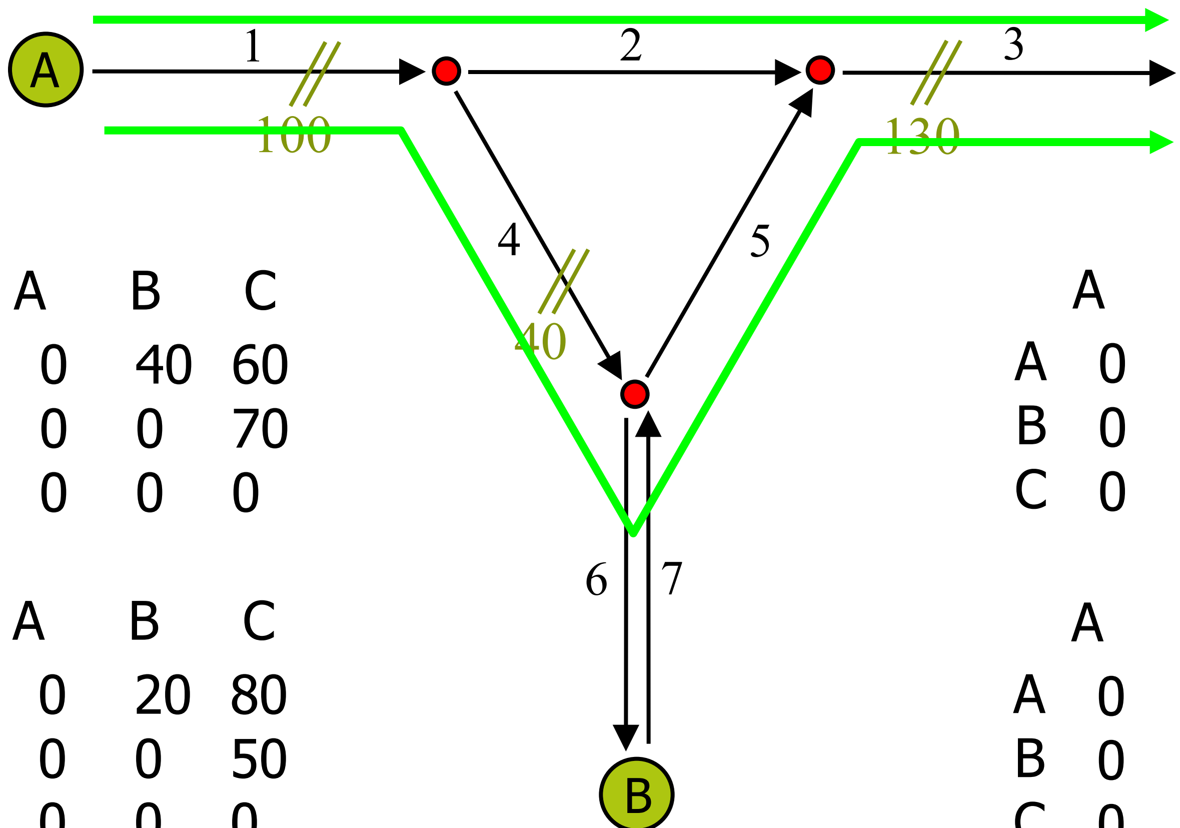
Calibration from link counts only



OD matrix:

	A	B	C
A	??	??	??
B	??	??	??
C	??	??	??

Calibration from link counts only



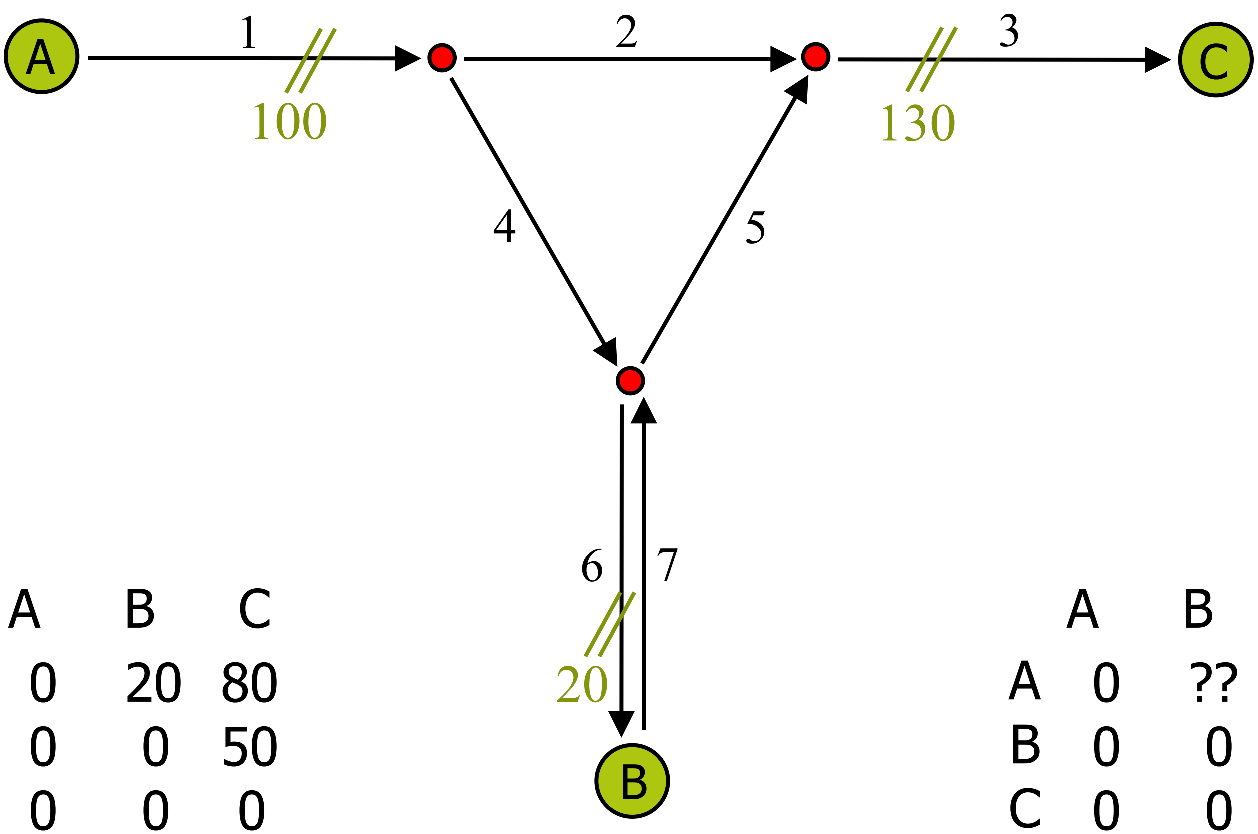
	A	B	C
A	0	40	60
B	0	0	70
C	0	0	0

	A	B	C
A	0	0	100
B	0	0	30
C	0	0	0

	A	B	C
A	0	20	80
B	0	0	50
C	0	0	0

	A	B	C	
A	0	??	??	100
B	0	0	??	
C	0	0	0	
			130	

Calibration from link counts only



	A	B	C
A	0	20	80
B	0	0	50
C	0	0	0

	A	B	C	
A	0	??	??	100
B	0	0	??	
C	0	0	0	
		20	130	

Updating an OD-matrix to traffic counts

- Counts set additional constraints for the OD-matrix

- Previously the constraints were only $\sum_j T_{ij} = P_i$ and $\sum_i T_{ij} = A_j$
- The constraint for a count on link a is $\sum_i \sum_j \sum_r \alpha_{ijr}^a \cdot \phi_{ijr} \cdot T_{ij} = S_a$

- So, one option is to start with a simple matrix ($T_{ij}=1$) and balance the matrix to fit all constraints (tri- or multi-proportional fitting)
  - Note that the distribution function and possible segmentations are omitted
  - Furthermore, this only works if all constraints are consistent…….

- The second (and better) option is to find a matrix that finds a balance between an original (a priori matrix) and the counts

Simple approach: multiplier per count

- Use an a-priori matrix from a model

- For each link having a count the ideal multiplier for the OD-pairs using that link is

$$x_k = \frac{S_k}{\sum_i \sum_j \sum_r \alpha_{ijr}^k \cdot \phi_{ijr} \cdot T_{ij}}$$

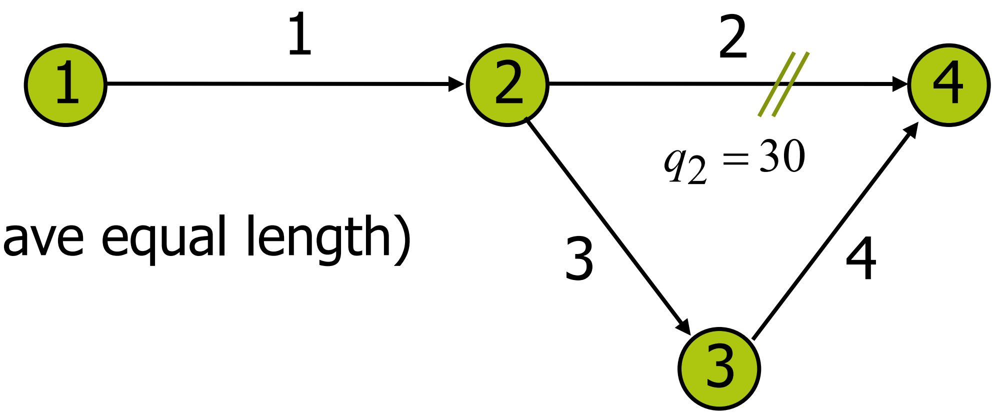
- As an OD-pair might pass multiple counts the resulting multiplier for that OD-pair is

$$y_{ij} = \frac{\sum_k \left(x_k \cdot \sum_r \alpha_{ijr}^k \cdot \phi_{ijr} \cdot T_{ij} \right)}{\sum_k \sum_r \alpha_{ijr}^k \cdot \phi_{ijr} \cdot T_{ij}}$$

i.e. the weighted average of the ideal multipliers for that specific OD-pair

- Obviously, the averaging of the ideal multipliers per count leads to a poorer match than intended, thus (again) an iterative procedure

Example based on AON



(all links have equal length)

OD-matrix

	1	2	3	4
1	0	10	15	20
2	0	0	10	5
3	0	0	0	20
4	0	0	0	0

$$\frac{30}{20+5} \cdot 20 = 24$$

$$\frac{30}{20+5} \cdot 5 = 6$$

Quality of constraints

- Not all counts have the same quality, e.g. permanent loop detectors versus manual counts between 07.00 and 19.00
- Trick: assign a weight (or "elasticity") e_k to each constraint ranging between 0 (no influence at all) and 1 (maximum adjustment)
- The multiplier for each constraint then becomes

$$x_k = \left(\frac{S_k}{\sum_i \sum_j \sum_r \alpha_{ijr}^k \cdot \phi_{ijr} \cdot T_{ij}} \right)^{e_k}$$

- Consequence: not all constraints are met at all costs

3.2

Estimating OD-matrices Formulation as optimisation problem

Criteria for differences between OD-matrices

- Least squares between estimated number of trips and the a priori number of trips

$$\min : \sum_{ij} \left(\hat{T}_{ij} - T_{ij}^{ap} \right)^2$$

- Minimum distance criterion based on information minimisation

$$\min : \sum_{ij} \left(\hat{T}_{ij} \cdot \ln \left(\frac{\hat{T}_{ij}}{T_{ij}^{ap}} \right) - \hat{T}_{ij} + T_{ij}^{ap} \right)$$

Formulation as optimisation problem

- Given the objective function (criterion) the constraints can be included using Lagrange multipliers, e.g.

$$L = \sum_{ij}\left(\hat{T}_{ij}\cdot \ln\left(\frac{\hat{T}_{ij}}{T_{ij}^{ap}}\right) - \hat{T}_{ij} + T_{ij}^{ap}\right) - \sum_{k}\lambda_k \cdot \left(S_k - \sum_i\sum_j\sum_r \alpha_{ijr}^{k}\cdot \phi_{ijr}\cdot \hat{T}_{ij}\right)$$

- Setting the derivatives equal to zero yields a set of equations that can be solved using dedicated software

NB Note the similarity with the entropy maximising approach discussed for the trip distribution model

Three comments

- Note that this formulation can also be used for other types of constraints such as production, attraction or classes of observed trip length distribution
- These methods can also deal with probability distributions for the constraints
  - Including covariance between observations
- These optimisation methods are very (and perhaps too) powerful!

3.3

Practical topics

Practical topics

- Ratio between variables and constraints?
- Which adjustments are acceptable?
- What about OD-pairs that are not affected by constraints?
- Matrix estimation in congested networks?
- What about a forecast?

Ratio between variables and constraints

- Number of variables: n^2
- Number of constraints:
  - Production and attraction: $2n$
  - Counts: depends on the size of model
    - Municipality: a few dozen
    - Region: a few hundred
  - In order to reduce impacts of “route choice", counts are often aggregated into screenlines => less constraints
- Problem is heavily underspecified

Which adjustments are acceptable?

- In many cases there are OD-pairs that relate to a single count
- Consequence is that a perfect match is always possible
- However, is this really perfect?

What about OD-pairs that are not affected by constraints?

- Suppose that 60% of the trips from a zone pass a count
- Suppose that these 60% show an overall increase of 10%
- What is then realistic for the other 40% of the trips from that zone:
  - Should be decreased to match the original production?
  - Should be increased by 10% as well?
  - Change nothing?

Matrix estimation in congested networks?

- In OD-estimation you try to match the modelled flow with counts
- In congested networks the flows are related to capacities
- In fact, the assignment already tries to limit the flows to the capacities
- So what is estimated in matrix estimation: demand or capacities?
- Furthermore: if the OD-matrix changes the assignment changes as well…………….
  - Iterative process?

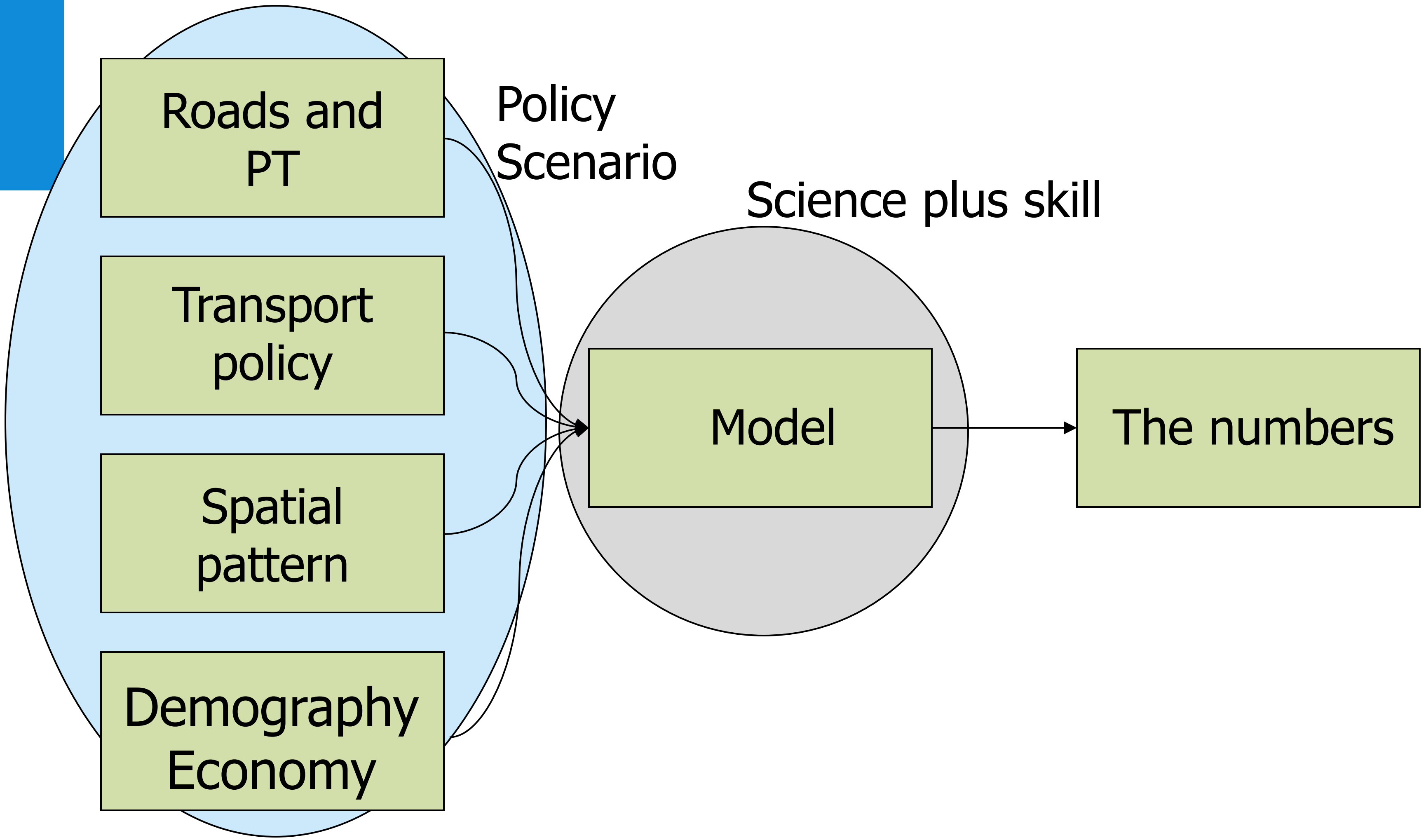
What about a forecast?

- Estimated OD-matrix yields a better match with counts
- For a future year you do not have counts
- So what do you do?
- Stick to the corrections of the OD-matrix
    - See discussion on K-factors (Chapter 5)
- Or ignore corrections for a future year
- Or……..

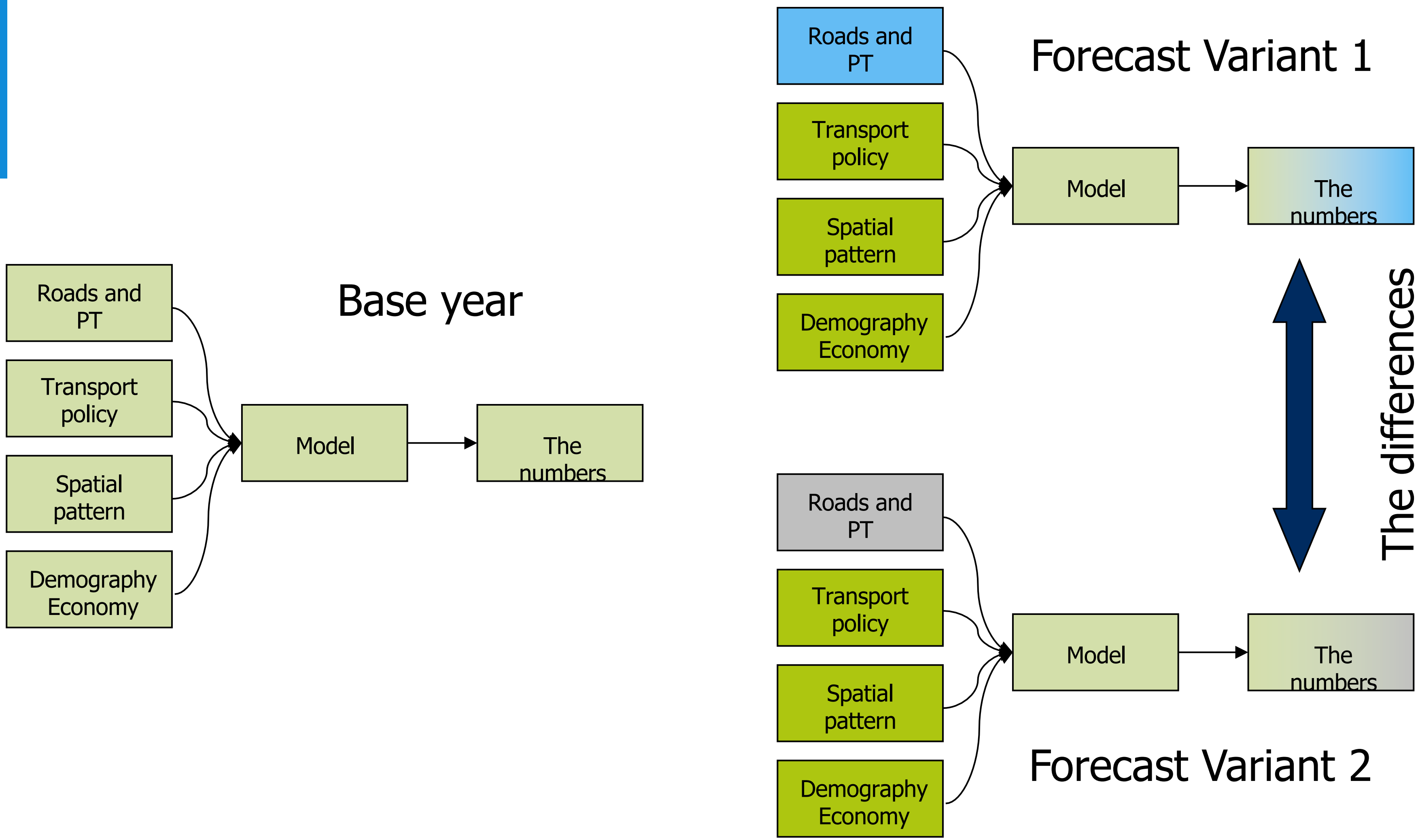
4.

Applying a model for forecasting

How is a model used?



Forecasting



How to deal with the limitations of your model when making a forecast?

- Base year model
  - Justification of your model

- However, model is a simplification of reality. How to deal with that when making a forecast?

- 3 options
  - Rely on base year matrix
  - Rely on model
  - Combination of both

- In all cases: proper interpretation of the results

Rely on the base year matrix

- Make sure that the base year is of high quality
  - Good fit with counts using a high quality a priori matrix
- Use a growth factor method for the forecast
  - Especially changes in production and attraction
- Limitations
  - How to deal with new developments?
  - Network effects on distribution and modal split are not included
  - Mostly suitable for short term forecasts only

Rely on the model

- Make sure that the model is of high quality
  - Proper functions and parameters
  - Decent match with counts etc.
- Run the model using the input for the forecast year
- Check the model for those results where the base year model had some limitations

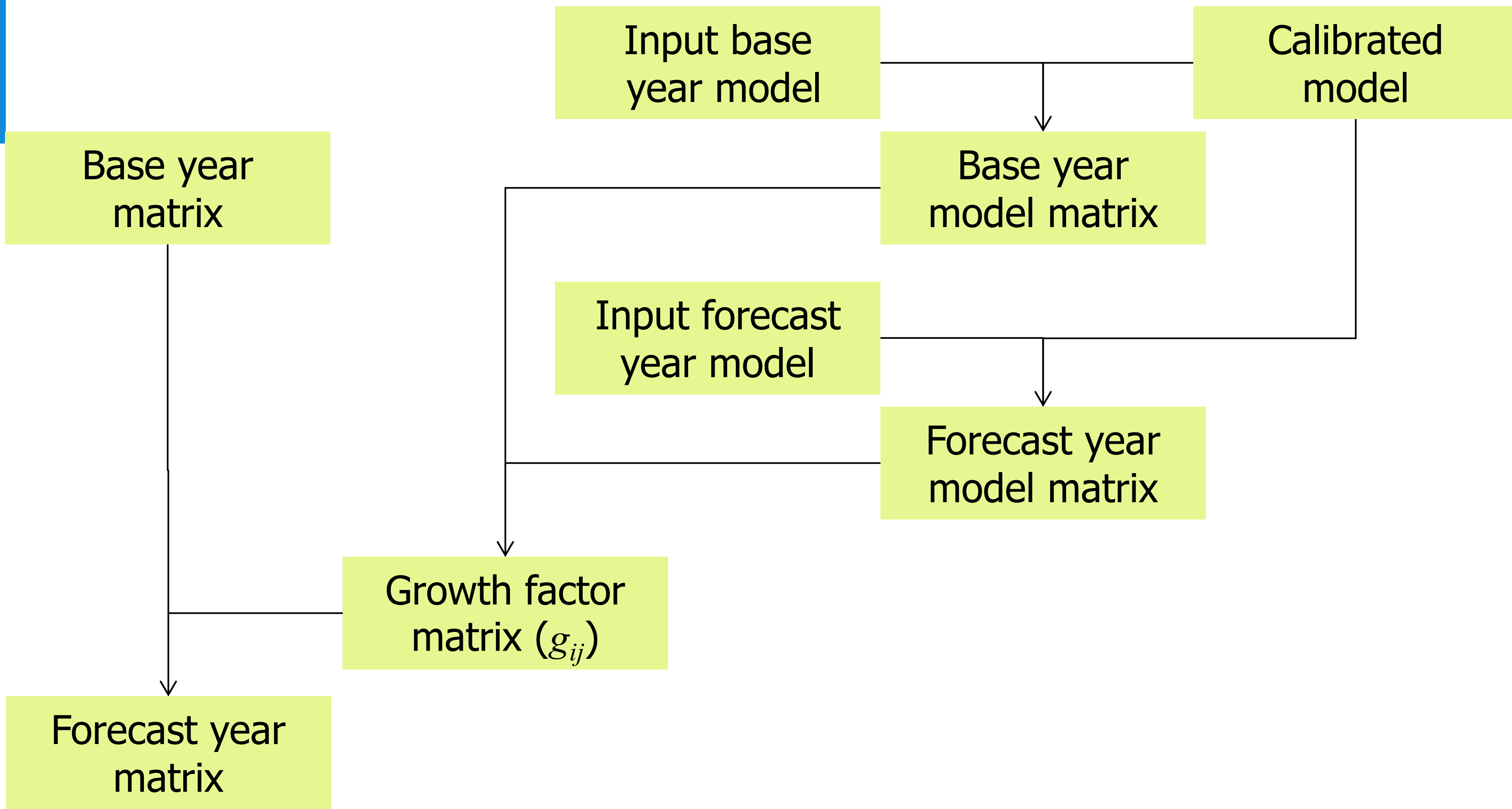
Combination of both worlds

- Key philosophy:
  - Model is the best representation of the mechanisms or sensitivities in the transport system
  - Base year matrix is best representation of the OD pattern

- Thus use the model to determine the relative changes between base year and forecast year (synthetic model or shadow model)

- Next, use these relative changes to update the base year matrix (pivot)

Pivot model approach



Points of attention for the pivot method

- Relationship between base year matrix and base year model?

- Growth factor method extrapolates “errors” in the base year matrix
  => Demand patterns should be consistent
  (e.g. use base year model matrix as input for OD-estimation)

- How to account for changes in spatial structure (empty cell problem)
  - There’s nothing to multiply with

Substitute cell value with results synthetic OD-matrix
Or,
Use matrix manipulation to account for changes in spatial structure and use pivot to account for network and policy changes only

In all cases

- Take care of the relevant accuracy
  - Within the model there are partial persons and partial cars!
  - Exact values tend to become a fact

- Use your common sense!
  - What's the logic?
  - Do you believe the changes and their magnitudes?
  - Can you explain it in laymen's terms as well?