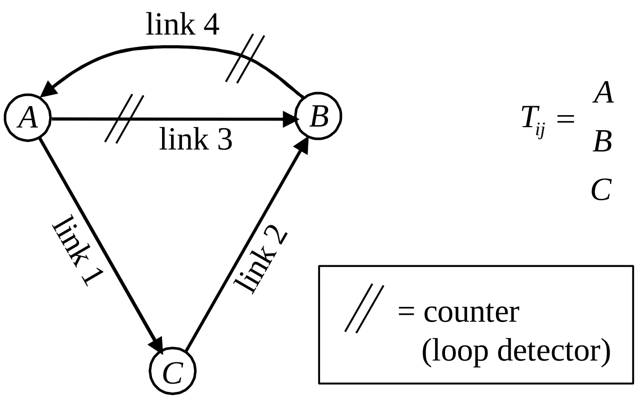
Post lecture questions Modelling in practice 2017 plus answers

OD-estimation

Consider the network below with three origins and three destinations, A, B, and C. The travel demand is given by the OD trip matrix T_{ij} below.



$$T_{ij} = \begin{array}{c|ccc} & A & B & C \\ \hline A & 0 & 1000 & 140 \\ B & 400 & 0 & 160 \\ C & 240 & 200 & 0 \end{array}$$

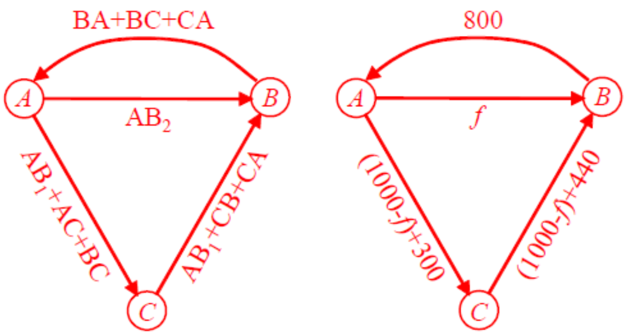
Furthermore, all links are assumed equal, having link travel times t_a depending on the link flows q_a given by the following function:

$$t_a(q_a) = 1 + \left(\frac{q_a}{400}\right)^2, \qquad a = 1, 2, 3, 4.$$

There are two counters in the network, one on link 3 and one on link 4, counting the number of vehicles passing through that link.

(a) Assuming that all travellers are assigned according to a deterministic user equilibrium, how many travellers will be assigned to link 3?

First, we determine which OD flows pass through which link, indicated below. Note that the flow from A to B can take two different routes. All other OD flows only have a single route, so for these routes the user equilibrium is easy to determine. Using the numbers from the OD matrix, only one unknown variable remains, f



Since for OD pair (A,B) also the user-equilibrium should hold, we can write down the following equations (assuming that both routes will be used):

$$t_3(q_3) = t_1(q_1) + t_2(q_2)$$

$$1 + \left(\frac{f}{400}\right)^2 = 1 + \left(\frac{(1000-f)+300}{400}\right)^2 + 1 + \left(\frac{(1000-f)+440}{400}\right)^2$$

$$\left(\frac{f}{400}\right)^2 = 1 + \left(\frac{1300-f}{400}\right)^2 + \left(\frac{1440-f}{400}\right)^2$$

$$f^2 - 400^2 - (1300-f)^2 - (1440-f)^2 = 0$$

$$f^2 - 5480f + \left(400^2 + 1300^2 + 1440^2\right) = 0$$

$$f = \tfrac{1}{2}\cdot 5480 - \tfrac{1}{2}\sqrt{5480^2 - 4\left(400^2 + 1300^2 + 1440^2\right)} \approx 847$$

Hence, 847 vehicles will be assigned to link 3.

(b) Suppose that the counter on link 4 indicates 1000 passing vehicles. Change the OD trip matrix T_{ij} such that it is consistent with this count.

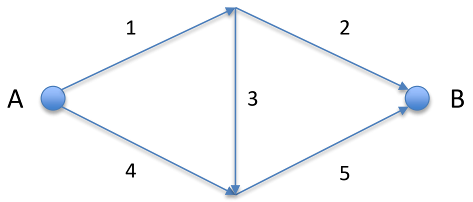
As indicated above, there will be passing 800 vehicles through link 4 according to the OD matrix, while the counter on link 4 now registers 1000 vehicles. This means that the OD flows BA, BC, and CA need to be updated such that BA+BC+CA = 1000 instead of 800. By increasing these OD flows by 200/800 = 25%, they will add up to 1000. Therefore, the new OD flows will become:
BA = 400*1.25 = 500
BC = 160*1.25 = 200
CA = 240*1.25 = 300
Hence, the new updated OD matrix will be:

$$T = \begin{bmatrix} 0 & 1000 & 140 \\ 500 & 0 & 200 \\ 300 & 200 & 0 \end{bmatrix}$$

(c) Suppose that the OD trip matrix is unknown, but that link counts on all links are available. Is it possible to uniquely determine the OD trip matrix using these link counts? Explain your answer.

No, it is not possible to determine the OD trip matrix using all link counts. There are 6 unknowns in the OD matrix. If there are 4 link counts, there are 4 equations:
AB1+AC+BC = count link 1
AB1+CB+CA = count link 2
AB2 = count link 3
BA+BC+CA = count link 4
We cannot solve for six unknowns with four equations, hence there exist many OD matrices that would fit the link counts. A unique OD trip matrix cannot be determined.

OD-estimation based on counts is not always straightforward. For instance OD-pairs could have different routes. Consider for instance the following network with link numbers:

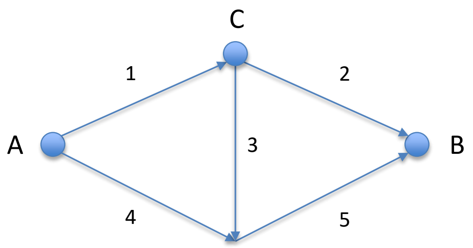


In this network there is only one OD-pair having 3 routes. Assume there's a limited budget for traffic counts.

(a) Specify the minimum number of counts you would need to determine the flow AB based on counts only in the case that you would allocate them strategically, and in the case if you would randomly assign the counts to the links. Explain your line of reasoning. [4]

Two would already be sufficient if all outgoing or all incoming links are considered.
3 routes might suggest that you would need 3 counts. Many combinations of 3 counts are indeed feasible as long as all routes are "covered" but not all: the options 1-2-3 and 3-4-5 lack a specific route and thus make it impossible to determine the total flow. So you would need 4 counts to make sure you don't miss a route.

In this network a new zone C is added:



(b) Indicate how many counts you would need to determine the 3 flows: AB, AC and CB. Explain your answer. [3]

In this case you miss information to determine an OD-matrix. There are 6 unknowns (OD specific route flows: AC, AB: 3 routes, CB: 2 routes) and only 4 independent counts. It is not possible to distinguish the use of route 1-2 or OD-pair AB from the flows AC and CB via link 2.

Up until now the uncongested situation was considered. Determining an OD-matrix in congested networks brings in a new perspective.

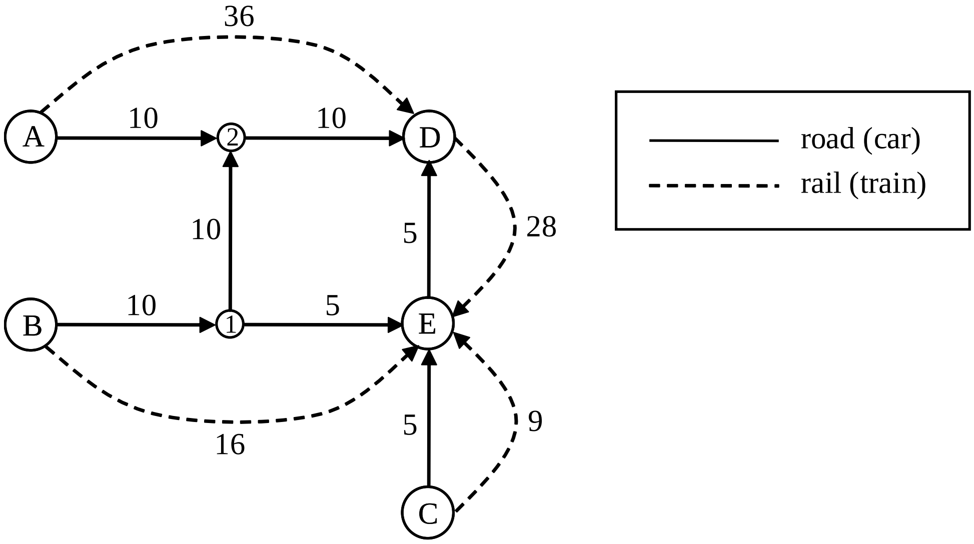
(c) Explain which issue should be dealt with in case of estimating OD matrices in congested networks, compared to OD-estimation in uncongested networks. [3]

In congested network the distribution over the paths is affected by capacity constraints. Therefore changes in the OD-matrix will lead to new route proportions and thus to different matches with individual counts. Recall the iterative process of matrix calibration for LMS/NRM. You could also say that the capacity already works as a kind of "count' when performing the assignment.

Full 4-stage model

Aggregated modelling approach

Consider a transportation network with infrastructure for two modes: car and train. There are three residential zones (A, B, and C) and two employment zones (D and E). The road and rail segments are indicated as directed links in the following figure. The travel time (in minutes) in uncongested situations is indicated for each link. We only consider unimodal trips, no multimodal trips.



The focus is on the commuters from home to work in the morning peak, hence zones A, B, and C are considered origins and zones D and E are considered destinations.

From a cross-classification model the following relationship between income, number of cars, and the average number of trips made by an individual has been determined:

	low income	high income
0 cars	0.6	1.1
1 or more cars	0.8	1.4

For example, an individual with a low income and no car makes on average 0.6 commuting trips in the morning peak.

The following table provides the number of residents for each of these classes in each of the origin zones A, B, and C:

		low income	high income
Zone A	0 cars	402	120
	1 or more cars	241	60
Zone B	0 cars	30	110
	1 or more cars	40	235
Zone C	0 cars	165	104
	1 or more cars	103	103

For example, there are 402 residents with a low income and no car in zone A.

Using employment data on a zonal level in a regression model, the commuting trip attractions in the morning peak have been determined, yielding 750 and 799 arriving trips in zone D and E, respectively.

The residential data is collected with a higher accuracy than the employment data.

(a) Determine the trip production for origins A, B, and C.

P_A = 402*0.6+120*1.1+241*0.8+60*1.4 = 650
P_B = 30*0.6+110*1.1+40*0.8+235*1.4 = 500
P_C = 165*0.6+104*1.1+103*0.8+103*1.4 = 440

(b) Determine the travel time impedance (skim) matrices for car and train (assuming zero transfer delays for trains).

c_car	D	E
A	20	Inf
B	20	15
C	10	5

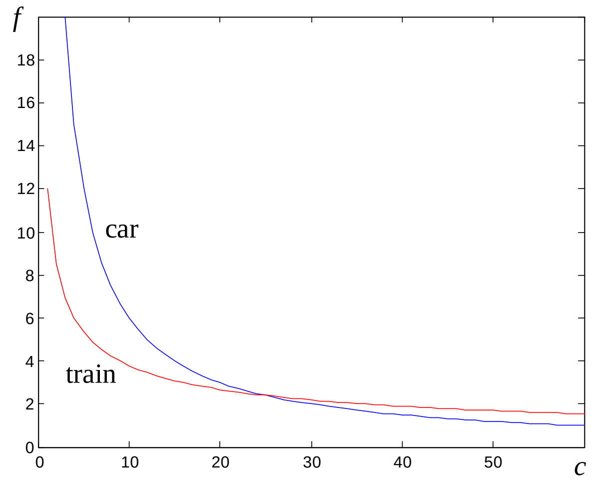
c_train	D	E
A	36	64
B	Inf	16
C	Inf	9

Assume the following power functions as trip distribution functions for car and train:

$$f(c_{ij}^{car}) = \frac{60}{c_{ij}^{car}}, \quad \text{and} \quad f(c_{ij}^{train}) = \frac{12}{\sqrt{c_{ij}^{train}}},$$

where c_{ij}^{car} and c_{ij}^{train} are the car and train travel times (in minutes), respectively.

(c) Sketch the two trip distribution functions in one graph. Assuming equal travel times, for which travel times will the car be the preferred mode for the majority of travellers?



When will $f(c) > f(c)$?

$$\frac{60}{c} > \frac{12}{\sqrt{c}} \quad \Rightarrow \quad 60 > 12\sqrt{c} \quad \Rightarrow \quad c < 25.$$

Hence, for travel times up to 25 minutes the car will be the preferred mode.

(d) Show that the simultaneous trip distribution/modal split based on the doubly constrained gravity model leads to the following car trip matrix:

	D	E
A	300	0
B	150	200
C	120	240

Total production = 650+500+440 = 1590.
Total attraction = 750+799 = 1549.

Since residential data is collected with higher accuracy, the production data is more accurate, therefore the attractions are scaled towards the productions:

Attraction to D = 750 * 1590/1549 = 770
Attraction to E = 799 * 1590/1549 = 820

Matrix with accessibilities:

	D		E		
	car	train	car	train	Prod.
A	3	2	0	1.5	650
B	3	0	4	3	500
C	6	0	12	4	440
Attr.	770		820		

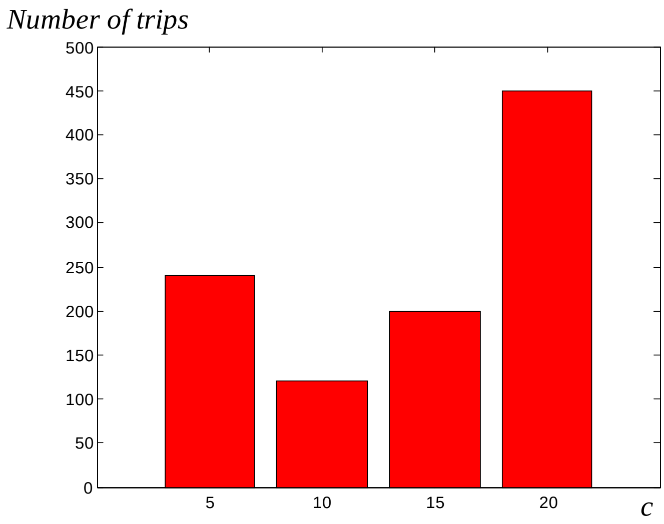
Scaling towards productions:

	D		E		
	car	Train	car	train	Prod.
A	**300**	200	**0**	150	650
B	**150**	0	**200**	150	500
C	**120**	0	**240**	80	440
Attr.	770		820		

Note that scaling towards attractions is not necessary, the trips already add up to the correct attractions.

(e) Plot the trip frequency distribution for the car (use four bins, around 5, 10, 15, and 20 minutes travel time).

There are 240 cars with c = 5.
There are 120 cars with c = 10
There are 200 cars with c = 15
There are 450 cars with c = 20



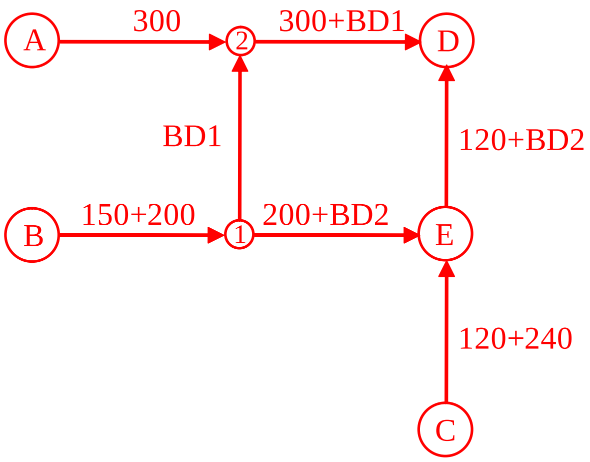
Assume that all links are uncongested, except for the link from node 1 to the destination node E. The travel time on this link is given by the following function:

$$t(q) = 5 + \frac{q}{30},$$

where t is the travel time (in minutes) and q is the car flow in the morning peak.

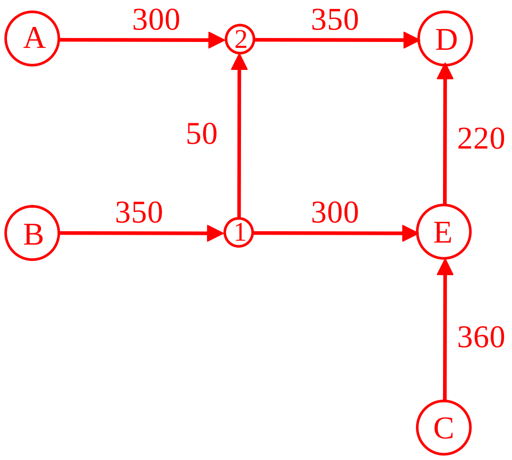
(f) Determine all the link flows for cars in a deterministic user equilibrium assignment.

All OD pairs only have a single route, except for BD that has two available routes (with route flows denoted by BD1 and BD2):



If both routes from B to D are used, they should have equal travel time.
Travel time route 1 = 10 + 10 + 10 = 30
Travel time route 2 = 10 + (5+q/30) + 5 = 20 + q/30, where q = 200+BD2.

Both routes have equal travel time if 30 = 20 + (200+BD2)/30, which yields BD2 = 100. Hence, BD1 = 150 – 100 = 50. The link flows therefore become:



(g) Will the car flow on the link from node 1 to node 2 increase, decrease, or stay the same if the transfer delays for trains are taken into account? Explain your answer.

Transfer delays affect the travel time for train trip AE. If this travel time increases, the train accessibility for AE decreases, and therefore the trip matrix for the car will also change (see also question (d)). Essentially, there will be more car trips on AD, BE, and CE and less on BD and CD. Due to more car trips on BE, some cars from B to D will divert to the first route along the link from node 1 to node 2. Therefore, the flow on that link will increase. Note that the decrease of the flow BD might limit the increase. However, due to the travel time function of link 1-E it is likely that there will be a net increase.

Disaggregated modelling approach

In a disaggregated approach the choices for individuals are determined by computing the probabilities for each choice alternative given the utilities of the alternatives.

For *trip choice*, let the probability of making an additional trip be given by the following binary logit model,

$$p = \frac{1}{1+\exp(-V)}, \quad \text{with } V = 0.3 + 0.2 \cdot C + 0.005 \cdot I,$$

where V is the utility of a traveller of making an additional trip, C equals 1 if the traveller owns a car and zero otherwise, and I is the traveller's income (in thousands of euros).

(a) Using the stop-repeat model, determine the expected number of trips made by a traveller who owns a car and has an income of 20 thousand euros.

Since C = 1 and I = 20, the utility of making an additional trip is
V = 0.3 + 0.2 * 1 + 0.005 * 20 = 0.6.

Therefore, the probability of making an additional trip is
p = 1/(1+exp(-0.6)) = 0.65.

The probability of making no trips is (1-p) = 0.35.
The probability of making 1 or more trips is 0.65. The probability of making no more additional trip is 0.35, so the probability of making exactly 1 trip is 0.65 * 0.35.
Therefore, the probability of making exactly n trips is $p_n = 0.65^n * 0.35$.
The expected number of trips is $\sum_n n \cdot p_n = p/(1-p) = 1.82$.

Destination choice and mode choice are assumed to be taken simultaneously and therefore modelled using the following logit model:

$$p_m^d = \frac{\exp(V_m^d)}{\sum_{m'}\sum_{d'}\exp(V_{m'}^{d'})} \text{ (= probability of choosing destination } d \text{ and mode } m\text{),}$$

where the utility of traveling to destination d using mode m is given as

$$V_m^d = 0.01A^d - 0.05T_m^d,$$

with A^d the attraction of destination d and T_m^d the travel time (in minutes) needed for traveling to destination d using mode m.

(b) Considering the same car and train network and the same attractions (750 and 799) as in Question 2, how many car trips will be made from zone A to zone D if the trip production from zone A is given by 650 travellers?

From zone A we can reach:
1. zone D by car
2. zone D by train
3. zone E by train

Therefore, there are three alternatives with the following utilities:
V_car_D = 0.01 * 750 – 0.05 * 20 = 6.5.
V_train_D = 0.01 * 750 – 0.05 * 36 = 5.7.
V_train_E = 0.01 * 799 – 0.05 * 64 = 4.79.

The probability of taking the car to D is therefore
exp(6.5)/(exp(6.5)+exp(5.7)+exp(4.79)) = 0.61.
Hence, the number of trips from A to D by car is 0.61 * 650 = 399.

Route choice is assumed to be determined by a logit model based on the uncongested route travel times,

$$p_{mr}^{od} = \frac{\exp(V_{mr}^{od})}{\sum_{r'} \exp(V_{mr'}^{od})} \text{ (= probability of choosing route } r\text{)},$$

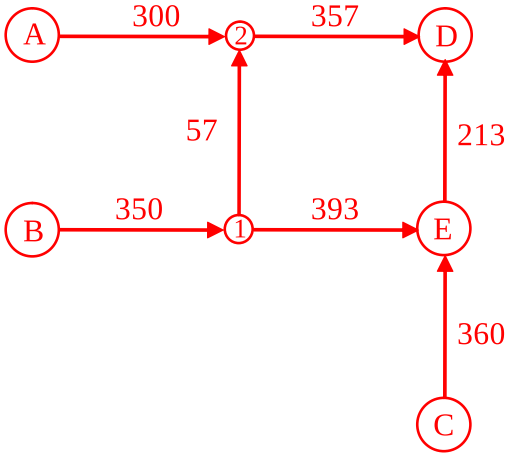
where the utility of taking route r (for a specific given origin-destination pair od and a specific mode m) is given as

$$V_{mr}^{od} = -0.05 T_{mr}^{od},$$

with T_{mr}^{od} the travel time (in minutes) on route r from o to d with mode m.

(c) Considering the same car and train network and the same trip matrix as in Question 2(d), determine the link flows for cars in a stochastic assignment based on the above logit model.

As before, relation BD is the only one that has two alternative routes. The utilities for both routes are (using uncongested travel times):
V1 = -0.05 * 30 = -1.5
V2 = -0.05 * 20 = -1.0
The number of travellers from B to D over route 1 is
exp(-1.5)/(exp(-1.5)+exp(-1.0)) * 150 = 57,
while the number of travellers from B to D over route 2 is 93.
The resulting link flows are:



Also the trains can be assigned to the network based on a logit model with a certain utility function for each route/line. In-vehicle travel time could be one of the attributes included in this utility function.

(d) Name two other attributes that could be included in the train utility functions for route/line choice.

- transfer time
- access time
- egress time
- number of transfers
- … etc.

LMS/NRM and 4-stage model

In the Dutch national transport model, LMS, some methods and techniques are applied that are different in traditional transport models. For example, LMS considers tours instead of trips. In computing the future origin-destination (OD) matrix, LMS uses the so-called pivot point method. For traffic assignment, QBLOK is used instead of more traditional assignment techniques.

(a) Explain the main advantage of using tours in LMS as opposed to using trips.

When using tours, one can guarantee that same person who travels from zone A to zone B is the same person that travels back from B to A. When using trips, two different matrices are determined independent of each other. Tours are particularly useful with respect to mode choice. If a person travels from home to work by car, then the person will also use the car to travel back home. This dependency is made explicit within the tours.

(b) Explain how the pivot point method works.

The pivot point method determines growth factors on a base matrix. This goes as follows. A model run for the base year is conducted, and a model run for a future year is done. Based on these model runs, growth factors per OD pair are determined. These growth factors are then applied to the base matrix. The future OD matrix is therefore pivoted around the base matrix with growth factors.

(c) What is a major disadvantage of such a pivot point method?

The pivot point method has the same disadvantage as a growth factor model. If a zone did not exist yet in the base year, but new developments exist in a future year, then the pivot point method is not able to determine this growth factor (as the value for this value in the OD matrix is zero).

(d) Explain the main difference between QBLOK and traditional traffic assignment models.

Traditional traffic assignment models assign all the flow to the network, even when the road capacity is exceeded. This leads to a higher travel time, but not to lower traffic flows. In QBLOK, the capacity is a hard constraint, such that the traffic flows do not exceed the road capacity. Travel time functions are not used in QBLOK, instead travel times are implicit outcomes using delays at bottlenecks.

(e) The estimated OD matrix in the LMS model should only be assigned with QBLOK and not with another assignment method. Explain this dependence between the OD matrix and the assignment technique.

The OD matrix in LMS has been calibrated on link counts using QBLOK. The OD matrix is loaded on the network using QBLOK and the link counts are compared with the link flows predicted by QBLOK. Hence, the resulting calibrated OD matrix fits the data with this specific assignment model. However, using a different assignment model will lead to different link flows and therefore a differently calibrated OD matrix. Hence, the OD matrix and the assignment technique should be used consistently.

A classical transportation model consists of four sub models.

(a) Which are these four sub models? And what travel choices are being modelled in each of these sub models?

1. Trip generation – trip choice
2. Trip distribution – destination choice
3. Modal split – mode choice
4. Trip assignment – route choice

In the Dutch transportation models LMS and NRM, two of these sub models have been combined.

(b) Which two sub models have been combined? And give two reasons why these sub models are usually combined.

The trip distribution and modal split steps have been combined.
Reason 1 (behavioural): destination and mode choice is often a simultaneous choice
Reason 2 (computational): an input for trip distribution is the distance or travel time to each destination. The question is, which distance or time to use when multiple modes exist? The logit analogy offers a solution, but a simultaneous trip distribution / modal split overcomes this problem entirely by directly using the mode specific distance and/or travel time.

For each sub model, an aggregate (on a zonal level) or disaggregate (on a household or individual level) approach can be taken.

(c) Name the main advantage and the main disadvantage of the disaggregate approach.

Advantage: More flexible to include travel behaviour and include individual or household specific attributes
Disadvantage: Requires more data

Often, a fifth sub model is added to the four-step model.

(d) What travel choice behaviour does this fifth sub model describe?

Time of day choice

The model systems LMS and NRM use tours instead of trips.

(e) What the advantage is of using tours in a model system?

A tour describes a sequence of trips from a certain origin to one or more destinations, and back. The main advantage is, that by using a tour as the unit of modelling, one can ensure that the same mode is used for each trip in the tour. For example, it is unlikely that the trip from origin to destination is done by bike, while traveling back is done by car.