Aggregation in transport networks for a flexible assignment model

UNIVERSITY OF TWENTE.

dat mobility

Master thesis
Justin van Steijn
October 2016

Aggregation in transport networks for a flexible assignment model

Master thesis for Civil Engineering and Management

Enschede, October 2016

Author: J.A. (Justin) van Steijn
University of Twente, Faculty of Engineering Technology
Student number: s1111825
Email: j.a.vansteijn@student.utwente.nl

Supervisors:

University of Twente

Prof. Dr. E. van Berkum

Dr. Ir. L. Wismans

DAT.Mobility

Dr. Ir. L. Wismans

Ir. L. Brederode

Ir. B. Possel

Summary

Transport modelling is a frequently used tool for transportation planning. This research focusses on the assignment step of the four-step transportation model: we develop a flexible aggregated assignment method for transport networks in order to reduce its computational costs.

Introduction
In the last decades, various developments played an important role in transport modelling. First, transport policies have become more complex which has lead to more complex and time-consuming transport models. Secondly, there is an increasing demand for better consistency between different transport models describing the same geographical area. Thirdly, there is demand for flexible transport model methods that can be used to match with the detail that is required for the output.

More simple and faster models are developed in order to overcome the problem of complexity and computational costs. However, these faster models cannot replace the original full-scale transport models due to lack of detail. Therefore, the goal of this research is to **develop a static transport assignment model that uses smart forms of aggregation to reduce computational costs and maintain accuracy as much as possible**. When aggregation is applied, potential research problems exist in the form of interpretability of results, the definition of zones and the level of aggregation.

The goal that the aggregated model must achieve is two-fold: to produce accurate traffic assignment traffic volumes and travel times, for lower computational costs. To remain interpretable, the aggregated model must use well-known existing low-level zone structures and the output must be consistent with current models.

Aggregation principles
The shortest path algorithm used in this research is Dijkstra’s algorithm. A sub-problem is defined as one Dijkstra’s algorithm with one source and multiple targets. The size and computational cost of this sub-problem depends on the number of nodes or links that is being considered before the algorithm is finished. Dijkstra’s algorithm can stop when all targets have been found.

Different basic aggregation methods have been tested. Basic zone and network aggregation methods result in equally optimal computational costs and accuracy. The rest of this research focusses on zone aggregation, as it is better able to model traffic flows on all roads and using the original zone structure.

Aggregation alternatives
We developed the following two building blocks for aggregated assignment models: the aggregated zone hierarchy and the route reconstruction method.

The zone hierarchy can either be fixed aggregated zones or adaptive zoning. With **fixed zones**, a single layer of aggregated zones is used in the assignment model. With **adaptive zoning**, there are multiple layers of aggregate zones; the idea is that every original zone (centroid) interacts with a different set of aggregated zones: small original or aggregated zones nearby and big aggregated zones at distance: this is the neighbourhood.

In both zone aggregation methods, the route between any origin-destination pair is not determined directly, but must be reconstructed. The route reconstruction method can either be with **first/last mile routes** within aggregated zones and shared routes between aggregated zones, or **two-sided route** reconstruction which means that two routes are generated and reconstructed into one route.

Three feasible combinations of building block options are developed: fixed zones with first/last mile routes, fixed zones with two-sided routes and adaptive zoning with two-sided routes.

Conclusion and recommendation

The aggregation alternatives have been applied on the transport network of The Hague. Fixed aggregated zones with first/last mile routes generate far less accurate link loads and travel times than fixed aggregated zones with two-sided routes, for only slightly better computational time. For the other two methods, the difference in computational cost and accuracy differs much between level of aggregation but less between aggregation method. Therefore, the following recommendations with regards to the application of aggregated transport models are done.

When the goal is to produce accurate skim matrices, adaptive zoning with two-sided routes on a high level of aggregation is advised. The case study showed that adaptive zoning with two-sided routes on the highest aggregation level results in 97% of all origin-destination pairs with a relative travel time error of at most 2%.

Adaptive zoning with two-sided routes is sometimes also the preferred way of aggregation for producing assignment results, especially when the network is big, considered in its totality and when the available time is low. For adaptive zoning on the high level of aggregation, 74% of all links have a relative link load difference of at most 10% for a reduction of 45% in computational costs. When more detail is required, for example in a variant study, fixed zoning with two-sided routes on the low level of aggregation is advised. This method results in 85% of all links within the acceptable link load difference of 10%, for a computational cost gain of almost 30%.

Samenvatting

Transportmodellering is een veelgebruikt hulpmiddel bij het plannen van verkeer en vervoer. Dit onderzoek legt zich toe op de toedelingsstap van het vierstaps-verkeersmodel: we ontwikkelen een flexibel geaggregeerd toedelingsmodel voor transportnetwerken om de rekentijd ervan te verkleinen.

Inleiding
In de laatste decennia hebben zich verschillende ontwikkelingen afgespeeld die van belang zijn voor transportmodellering. Ten eerste is transportbeleid complexer geworden, wat ertoe heeft geleid dat transportmodellen ook complexer en tijdsintensiever zijn geworden. Ten tweede is er een toenemende vraag naar meer consistentie tussen verschillende transportmodellen die hetzelfde geografische gebied beschrijven. Ten derde is er behoefte aan flexibele transportmodelmethoden die modeltechnisch overeenkomen met het gevraagde detailniveau in de uitvoer.

Om het probleem van complexiteit en rekentijd te overbruggen zijn er eenvoudigere en snellere modellen ontwikkeld. Echter kunnen deze snellere modellen het originele model op volledige schaal niet vervangen, vanwege een gebrek aan detail. Daarom is het doel van dit onderzoek om een **statisch transport-toedelingsmodel te ontwikkelen die slimme vormen van aggregatie gebruikt om de rekentijd in te korten terwijl de nauwkeurigheid zoveel mogelijk behouden blijft**. Echter, indien aggregatie wordt toegepast, ontstaan er problemen in de vorm van de mate waarin het model interpreteerbaar blijft, welke zonedefinities moeten worden gebruikt en welk niveau van aggregatie wordt gebruikt.

Het doel dat het geaggregeerde model moet bereiken is daarom twee-ledig: het produceren van nauwkeurige toedelingsresulaten in de vorm van linkvolumes en reistijden, gemodelleerd in minder rekentijd. Om interpreteerbaar te blijven moet het model de bestaande zone-structuren gebruiken, en de uitvoer moet consistent zijn met huidige modellen.

Aggregatie-principes
Het kortstepad-algorithme dat in dit onderzoek wordt gebruikt is het Dijkstra-algoritme. Hierbij is een sub-probleem gedefinieerd als één Dijkstra-algoritme met één bron en meerdere doelen. De grootte en rekentijd van dit sub-probleem hangt af van het aantal links dat wordt beschouwd voordat het algoritme is voltooid. Het Dijkstra-algoritme kan stoppen wanneer alle doelen zijn gevonden.

Verschillende basis-aggregatiemethoden zijn getest: de basisvormen van zone- en netwerkaggregatie resulteren in een rekentijd en nauwkeurigheid die vergelijkbaar zijn. De rest van dit onderzoek zal zich richten op zone-aggregatie, om dat die methode beter in staat is om verkeersstromen te modelleren op alle wegen, waarbij gebruik wordt gemaakt van de originele zone-structuur.

Aggregatie-alternatieven
De volgende bouwblokken zijn te onderscheiden in bij het toepassen van geaggregeerde toedelingsmodellen: de hiërarchie van geaggregeerde zones en de methode van route-reconstructie.

De hiërarchie van geaggregeerde zones is ofwel vaste geaggregeerde zones of adaptieve zonering. Bij het gebruik van **vaste zones** is er een enkele laag geaggregeerde zones die wordt gebruikt in het toedelingsmodel. Bij **adaptieve zonering** zijn er meerdere lagen van geaggregeerde zones; het idee hierachter is dat elke originele zone een interactie heeft met een andere verzameling van geaggregeerde zones; kleine originele of geaggregeerde zones dichtbij en grote geaggregeerde zones ver weg: dit wordt de *omgeving* van de originele zone genoemd.

In beide zone-aggregatiemethoden wordt de route tussen herkomst-bestemmingsparen niet direct bepaald, maar gereconstrueerd. De route-reconstructiemethode is ofwel eerste/laatste-kilometer-routes binnen

geaggregeerde zones en gedeelte routes tussen geaggregeerde zones, of **twee-zijdige** route-reconstructie, wat betekent dat er twee routes worden gegenereerd en gereconstrueerd in één.

Er zijn drie mogelijke combinaties van bouwblokken ontwikkeld: vaste zones met eerste/laatste-kilometer-routes, vaste zones met twee-zijdige routes en adaptieve zonering met twee-zijdige routes.

Conclusie en aanbeveling

Deze aggregatie-alternatieven zijn toegepast op het transportnetwerk van Den Haag. Vaste geaggregeerde zones met eerste/laatste-kilpometer-routes genereert verreweg de minst nauwkeurige link-volumes en reistijden dan vaste geaggregeerde zones met twee-zijdige routes, voor slechts een minimale mindere rekentijd. Voor de andere twee methoden zijn de rekentijd en nauwkeurigheid sterk afhankelijk van het aggregatieniveau, maar minder van de aggregatie-methode. Daarom worden de volgende aanbevelingen gedaan met betrekking tot de toepassing van geaggregeerde transportmodellen.

Wanneer het doel is om nauwkeurige skim matrices te produceren wordt geadviseerd om adaptieve zonering met twee-zijdige routes te gebruiken op een hoog aggregatieniveau. De modeltoepassing laat zien dat adaptieve zonering met twee-zijdige routes op het hoogste aggregatieniveau leidt tot 97% van alle herkomst-bestemmingsparen met een relatieve reistijdfout van ten hoogste 2%.

Adaptieve zonering met twee-zijdige routes heeft ook vaak de voorkeur voor het produceren van toedelingsresultaten, vooral wanneer het netwerk groot is, in zijn totaliteit wordt bekeken en beschikbare tijd laag is. Adaptieve zonering met hoog aggregatieniveau resulteert in 74% van alle links met een relatieve foutenmarge van ten hoogste 10%, met een rekentijdbesparing van 45%. Wanneer meer detail benodigd is, bijvoorbeeld in een verkeers-variantenstudie, wordt vaste zonering met twee-zijdige routes aanbevolen. Deze methode resulteert in 85% van alle links binnen de acceptabele foutenmarge van 10%, waarbij een besparing in rekentijd van bijna 30% wordt gerealiseerd.

Preface

This report marks the end of my master thesis and at the same time my study Civil Engineering and Management at the University of Twente. Although the start was difficult, I enjoyed working on the assignment. Working full-time for six months at a Dutch consultancy company gave me many new experiences. Having the freedom to try implement new ideas them is enjoyable, especially when the results turned out to be satisfying. The topic also interested me in new activities like computer programming, of which I learned much during the research.

I would like to thank my supervisors at DAT.Mobility and the university for making this research possible and who always helped me out when needed. Luc, thank you for the organizational advices. Luuk, thanks for the questions that made me thinking about new possibilities in the research. Bastiaan, thanks for your practical insights and advices. And finally, Eric, thank you for the mathematical foundations and reminding me to look now and then from distance at the subject.

Furthermore, I would like to thank my family and especially my parents for the support and faith that they gave me throughout my study. Finally, Iis, thank you so much for being always there for me and your interest and support in everything I did.

Justin van Steijn
October 2016

Table of contents

List of figures

List of tables

1 Introduction

This chapter gives an introduction into the research subject. Paragraph 1.1 describes the context of the subject. Paragraph 1.2 gives a problem definition based on the context. Paragraph 1.3 presents the research objective, in order to overcome the research problem. Paragraph 1.4 presents the research questions in order to fulfil the research objective. Finally, paragraph 1.5 describes the methods of this research.

1.1 Context

Transport modelling is a frequently used tool for transportation planning. Governments, transportation planners and scientists aim to estimate the influence of an investment or absence of new infrastructure, traffic management or the impact of spatial developments on the use of transportation services. Increasing sustainability is a major goal in the improvement of the transport system. Transportation sustainability should, in order to be effective, at least include impacts on economy, environment and social well-being (Mihyeon Jeon & Amekudzi, 2005). Examples are to reduce the number of traffic accidents and thus injuries and mortalities, increase accessibility and travel time reliability, reduce congestion, maintain the economic power of urban areas and decrease air pollution. Transportation models are an instrument to provide more insights into future situations in different proposed variants.

In the last decades, two developments played an important role in transport modelling. First, transport policies have become more complex: next to policy questions about new infrastructure, came new questions about behaviour influencing and better utilization of existing infrastructure (Martens & Jong, 2009). The effects of these policies are generally smaller than traditional policies, so models are needed on more detailed network level. Secondly, the pressure on transport networks has increased, resulting in larger differences between free flow and congested situations. More feedback loops to user choices are necessary in order to keep modelling accurate, user equilibrium (see section 2.2.1 in the theoretical framework) takes longer to reach and, therefore, computational time and complexity increase.

The result of those developments is that computation methods of complex models are often regarded as “black box” models and the results are therefore difficult to interpret. This means that transport models have become more versatile but also much more complex and time-consuming. However, these detailed and complex models are not always necessary, especially on a strategic level, where different transport investments or measures are compared. Therefore, more simple and faster models are developed in order to overcome the problems of complexity and interpretability. Less complex models have a shorter calculation time, reduce the chance of input or calculation errors and have more transparency as an effect (Annema & Jong, 2012).

The result is that there exists a range of models between detailed, complex and time-consuming models and simple and fast models. Differences between these models include the zone scale and model methods. Because the use of transport zones implies a trade-off between computational costs and accuracy (Hagen-Zanker & Jin, 2015), often different transport models are used to answer questions at different geographical scales. Inconsistency between model results regarding the same geographical area raise questions about the quality of the models. Therefore, there is an increasing demand for better consistency between different transport models describing the same geographical area and between areas (Martens & Jong, 2009). Next to this, there is demand for flexible transport model methods that can be used to match with the detail that is used in the model. This desired level of detail can differ between steps of the four-step model.

1.2 Problem definition

Transport modellers face different problems with regards to scale and aggregation while building a transport model. This section discusses the problems that arise with aggregation in the assignment step of the four-step model, which are the interpretability of results, the definition of zones and the level of aggregation.

Transport modelling requires spatial data aggregation in transport analysis zones. The definition of transport analysis zones is a common problem in transport modelling and also known as an instance of the modifiable area unit problem (MAUP). Aggregation of spatial data in these zones is necessary for transport modelling, because disaggregate information is often not available and not feasible to process in transport models. The transparency and interpretability of transport models for users depends on the form and scale of these transport analysis zones (Guo & Bhat 2004).

The definition of zones and their connection to the network normally does not get as much attention in transport modelling as data collection or estimation of parameters. Often the administrative units in the study area or "common sense" are used to define these zones. However, the method to determine zones can have its consequences for the generation of statistical and geographical errors (Martínez, Viegas, & Silva, 2009). Instead, transport modellers should define zones in such a way that there is a high trip generation/attraction homogeneity, the shape of zones is compact and that these zones are compatible with other administrative divisions (de Dios Ortuzar & Willumsen, 2011). Thus, a problem is that interpretability and accuracy may contradict with each other with regards to the definition of transport analysis zones.

One source of inaccuracy in transport models is that trips can start and end in every place in the transport zone, while trips are modelled to and from the centres of these zones (Martínez et al., 2009). When aggregating zones into bigger ones, these inaccuracies are likely to increase due to higher spreading of trip origins and destinations in bigger zones. On the other hand, aggregation means that less zones and/or less network links are used in the transport model which leads to lower computational costs. Performance of transport models highly depends on the number of transport analysis zones and the level of detail that is used in the network representation. Furthermore, when aggregation is applied, it becomes much harder to fulfil the zone creation criteria. Here, the problem is a lack of knowledge to what extent zone and network scale affects computational costs.

1.3 Research objective

The objective of this research is to **develop a static transport assignment model** that can solve the traditional Traffic Assignment Problem by **using smart forms of aggregation to reduce computational costs and maintain accuracy as much as possible**. It focusses on the assignment step of the four-step transport model. This research objective defines specific goals with regards to the aggregated model, which is constrained by boundary conditions.

The **goal** that the aggregated model must achieve is two-fold: to produce accurate traffic volumes and to generate accurate skim matrices. Skim matrices are used to estimate the traffic demand between origins and destinations.

- In order for the aggregated model to produce assignment results, both link loads and travel times between OD pairs must be accurate with respect to the regular assignment model. Because the aggregated model can replace multiple models of different scales in the same geographical location, it must produce link loads on every link as in the regular model, where emphasis is put on accuracy on the links that represent higher road levels. It is important that lower level links also contain loads, but the error tolerance is higher.
- For the model to produce skim matrices, travel times between OD pairs must be close to the travel times produced by the regular assignment.

These goals are further elaborated in the evaluation framework.

To maintain accuracy of the assignment model is only useful when the results are interpretable and transparent. Therefore, the following **boundary conditions** state what must be achieved in order to fulfil the research objective.

- Using existing well-known low-level zone structures as a basis preserves both interpretability and the criteria of compact zone formation.

- In order to maintain transparency, the model must be consistent with the regular traffic model: all traffic is assigned and the total junction traffic inflow and outflow should be zero (flow conservation), so that the results show no inconsistencies and junction modelling remains possible.
- The aggregated model requires, apart from the normally required transport network and OD matrix, a limited number of aditional data and manual configuration in order to be used as a generic model.
- Whether a certain gain in computational cost is justified with a certain loss of accuracy depends on model and output requirements. Therefore, the aggregation that the model uses must be flexible, which enables modellers to choose the aggregation level that is appropriate to their needs.

1.4 Research questions

The main research question, in order to fulfil the research objective, is:
How and to what extent can the balance between accuracy and computational costs of the transport modelling network assignment be improved, by using aggregation?

This main research question can be answered by considering all aspects of it. This is done by answering the sub research questions, which are:

1. What types of aggregation are known in transport modelling literature?
2. Which principles explain the computational cost efficiency of aggregation?
3. Which main aggregation alternatives are feasible in the context of traffic assignment?
4. To which extent does the level of aggregation influence the accuracy and computational costs?
5. Which aggregation methods are best suited for generating skim matrices?
6. Which aggregation methods are best suited for accurately estimating traffic flows?

1.5 Methodology

This section describes the steps that are taken during the research. Each step indicates its relation with the research questions. The research starts with a literature review, determination of the main aggregation principles and alternatives. After that, a design process is started that further develops these initial alternatives using the circle of alternative description, development and evaluation. At the end, the level and method of aggregation are evaluated and form the basis for recommendations. When all research steps are performed, the research objective should be achieved and the main research question can be answered. The steps of the methodology are summarized in Figure 1.

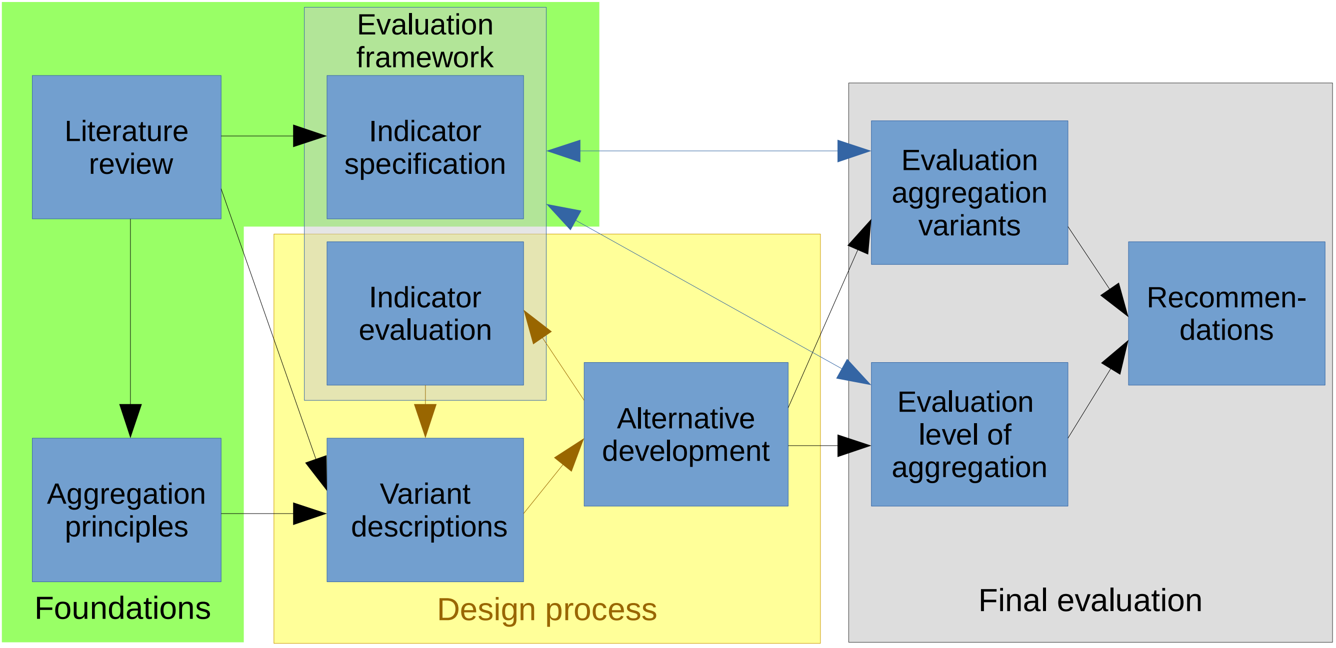


Figure 1: Methodology scheme

1.5.1 Foundations

Literature review

The research starts with a review of the literature in the field of transport modelling, the assignment phase of it, and known aggregation methods. Literature in the field of transport modelling, specifically the four-step transport model, is reviewed in order to gain a broad insight in the current modelling approaches. This is important, as this research does a partial redesign of the assignment phase. Therefore, most focus is on the literature on the assignment phase of the four-step model. Known aggregation methods in the literature are collected as a starting point for the research. The literature review is used as a starting point for the aggregation principles and the aggregation alternative descriptions. Furthermore, it answers research question 1.

Aggregation principles

In this step, a further investigation on aggregation in network assignment is done based the literature review. The principles that explain the computational advantage of aggregation, while at the same time maintaining a large part of the accuracy of the regular assignment, are explored. This step should answer research question 2 and forms the basis for the development of aggregation alternatives.

Evaluation framework specification

This step specifies the evaluation framework in order to objectively measure the results of various aggregation alternatives. The evaluation framework is based on the main objectives of this research as described in section 1.3, so the main evaluation criteria are accuracy of the assignment results and computational costs. The accuracy is further divided in link loads and travel times. Based on these main research objectives and indicators found in the literature, the performance indicators for this research are specified. The indicators are used in the final evaluation of aggregation alternatives.

1.5.2 Design process

Alternative descriptions

The description and definition of aggregation alternatives is performed multiple times in this research and forms the starting point of the design process. Initially, aggregation alternatives are determined based on the literature review and the knowledge gained by the aggregation principles. As the research progresses, some alternatives are implemented, tested and evaluated. This leads to ideas regarding the modification of existing alternatives or creating new aggregation alternatives that are expected to improve the balance between accuracy and computational costs. In the end, the aggregation alternatives that have the most principal difference remain and are used for the final evaluation. Research question 3 is answered using the final list of aggregation alternatives.

Alternative development

The development of aggregation alternatives forms the main activity of this research and is done iteratively in the design process. In the first instance of this step, the initial basic alternative is implemented as part of a transport model. In later instances, this step improves the implementation according to the modified description of aggregation alternatives.

Evaluation framework indicator evaluation

The developed aggregation methods in the previous step are evaluated using the evaluation framework, so that the effect of the improvements can be measured. Depending on the outcomes of this evaluation, it can be decided that a certain alternative is suitable or not. Emphasis is put on the expected cause of inaccuracies in the outcomes, so that improvements can be done in this area. The evaluation generates a new iteration in the design process, starting with the refinement of the aggregation alternative descriptions.

1.5.3 Final evaluation

When the design process is done several times and it is expected that the research goals can be sufficiently met, the final evaluation can be performed. Its goal is to give recommendations for usage of the aggregated assignment models.

Evaluation of aggregation level

In this research step, the level of aggregation in the various aggregation methods is investigated based on the final versions of the aggregated assignment models. The indicators in the evaluation framework are used for this. After this step is performed, research question 4 can be answered.

Evaluation of aggregation alternatives

The aggregation alternatives that result from the design process are evaluated and compared with each other. This is done using the indicators in the evaluation framework.

Recommendations

The final step consists of giving recommendations with respect to usage of the aggregated assignment models. This includes both the recommended alternative and the aggregation level for both goals of the research. When this step is performed, research questions 5 and 6 can be answered.

2 Theoretical framework

This chapter gives an overview of the literature in the relevant subjects. The review starts with transport modelling in general and the network assignment phase in detail. Furthermore, literature about aggregation is reviewed.

2.1 Transport modelling

Transport modelling is a frequently used method and tool for transportation planning. Transportation planning consists of activities that change the way that transportation services are provided. Most commonly these are decisions that governments make. These activities can be distinguished from transport management and operations activities, which relate to the transportation organization itself and to the provision of transportation services in a stable state respectively (Chen & Liew, 2003). For transportation planners, it is crucial to know how the demand and infrastructure utilization change as a result of changes in the transport system. A variety of different models is used in order to provide insights in future transport situations.

In general, a model is a representation of a part of the real world, which is the system of interest. The model focuses on those elements that are important for a particular point of view. Models are therefore problem and viewpoint specific (de Dios Ortuzar & Willumsen, 2011). In transportation models, the considered part of the real world is the geographical area that is assumed to have influence on the area of research. The viewpoint is mobility-focused: the socio-economic data that are relevant for mobility and transport networks are considered. Thus, important aspects of transport networks are the geographical areas that are divided in transport analysis zones (TAZ) and its socio-economic data, and the transport network.

2.1.1 Requirements for transport models

Transport models are used to support transport planning. In transport models, different scenarios can be set up, in order to compare them for the goals that transport planners have. In these scenarios, different measures can be implemented: road and rail infrastructure, public transit services, demand management policies, traffic management policies, information strategies and land-use policies (Bliemer, Raadsen, Romph, & Smits, 2013). Goals that transport planners have can be categorized according to the earlier mentioned three pillars of sustainability: economy (for example accessibility), environment (for example emissions) and social well-being (for example equity of transport services). In order to reach these goals, some properties of transport modelling are important: realism, robustness, consistency, reliability and accountability of results, and ease of use.

2.1.2 History of transport modelling

The traditional transport model, the four-step model, was developed in the United Stated in the 1950's (Martens & Jong, 2009). In West-Europe, this model started to being developed and used in the 60's and 70's. These models had a macro detail level, which are models developed for a regional or national scale and mostly included main roads. In the 80's and 90's, the number of models increased due to increased technical possibilities. Models became less understandable for non-experts, so a black box for outsiders. In the 90's, traffic management, travel behaviour and pricing became more important. New models developed for these purposes, such as models on a micro level and dynamic models. From the year 2000, model results became more important for decision making due to introduction of new norms in laws, for example laws on air quality. Models used for different projects often became more consistent so that model outcomes could be compared. However, this consistency has lead to less flexibility: new kinds of model input data could not be used any more to keep consistency between models. Thus, flexibility and consistency are often contradicting values in transport modelling.

Two types of transport models currently exist: the traditional four-step model and the activity-based transport model. The first one is described here.

2.1.3 Four-step model

The four-step transport model is trip-based. It generally consists of four steps: trip generation, distribution, modal split and assignment. Those four steps are often applied iterative: for example, the outcomes of the network assignment can function as an input for the modal split. Furthermore, there is a wide variety of models between disaggregate and aggregate (Chen & Liew, 2003). Disaggregate models consider the transport choices (among different alternatives) that individuals or households, which differ in socio-economic situation, make in a given situation. Disaggregate models are often microscopic demand models. Aggregate models consider transport decisions of a geographic group of people based on socio-economic characteristics of the group.

Trip generation and attraction

In the first phase of transport modelling, the number of trips generated and attracted by each transport analysis zone is estimated. Disaggregate models often use household data and zone-based models use aggregate data. There are linear regression models and category analysis models to estimate the number of produced and attracted trips (Chen & Liew, 2003):

- Linear regression models assume a statistical relationship between household or zone characteristics and the number of produced and attracted trips. These are of the form $Y=\beta_0+\beta_1 X_1+....+\beta_n X_n+\epsilon$ where X_n are the independent variables of the households or socio-economic data, β_n are coefficients that weight the independent variables according to their influence, and Y is the number of produced or attracted trips. Lopes, Brondino, & Da (2014) found that models that include regression models that include spatial variables perform better than models without these variables. In the presence of spatial autocorrelation, model estimations have to consider and incorporate the spatial structure of data as well.

- Category analysis models determine a mean trip rate for different types of people and trips. These types can be based on social, economic and demographic characteristics. The trip rates can be determined with empirical data that covers different types of people and trip types.

Distribution

In the second step of transport modelling, the origins and destinations of trips are paired so for each trip it is clear which origin and destination it has. For commuting, choosing both origin and destination are the result of fairly long-term decision processes (Chen & Liew, 2003): the origin is the place of residence and the destination is the work location.

		Destination zones					
		1	2	3	...	*j*	
Origin zones	1	T_{11}	T_{12}	T_{13}	.	.	O_1
	2	T_{21}	T_{22}	T_{23}	.	.	O_2
	3	T_{31}	T_{32}	T_{33}	.	.	O_3
	...	.	.	.	.	.	.
	i	.	.	.	.	T_{ij}	O_i
		D_1	D_2	D_3		D_j	

Table 1: Principle of trip distribution

There are two types of trip distribution models: the gravity models and discrete-choice models. The gravity model is the most widely used and has certain advantages: no data collection is needed, it is also applicable in case of new zones and it can handle changes in transport networks (de Dios Ortuzar & Willumsen, 2011). Based on the gravity principles, the attraction of a city with other cities depends on the size and distance of other cities. The size increases attraction, while distance decreases interaction between cities. The doubly constrained gravity model, which is constrained by the number of total trips of both the origin and the destination, is given by $T_{ij} = A_i B_j O_i D_i f(c_{ij})$. Here, O_i is the number of trips originating at *i*, D_j is the number of trips destined for *j*, and A_i and B_j are as follows:

$$A_i = \frac{1}{\Sigma_j B_j D_j f(c_{ij})}, \quad B_i = \frac{1}{\Sigma_i A_i O_i f(c_{ij})}$$

where $f(c_{ij})$ is a function that depends on (generalized) costs, which can consist of travel time, cost, comfort among other factors. Since these functions depend on each other, an iterative process is needed to end up in an equilibrium.

Discrete choice models can also be used for distribution modelling. For every zone, a select set of other zones is chosen using destination sampling, in order to reduce the number of zones that should be considered in the choice model. Utilities for different zones define the trip distribution in the choice model. More about this model follows in the next section.

Modal split

Mode choice models are often regarded as disaggregate models: individuals make choices regarding the transport mode. The classical micro-economic theory of the costumer, that maximizes his or her utility given a particular budget, often fails because of the transitivity of preferences and inconsistent choices among different travellers (Chen & Liew, 2003). Transitivity of preferences indicates that there is no dominant choice.

Therefore, a general probabilistic model of choice was developed, where there is a systematic utility component and a random utility component, so that the total utility is $U_{n,i} = V_{n,i} + \epsilon_{n,i}$. The systematic component consists of different attributes of mode choice alternatives and their weights, so that $V = \beta_1 x_1 + ... + \beta_n X_n$ for a particular mode alternative. The random term represents arbitrariness of choice, perception errors, incomplete information and disregarded aspects.

The logit model shows that the chance of choosing a particular mode *j* for individual *i* is according to $P(i|C_n) = \frac{e^{\mu V_{n,i}}}{\Sigma_j e^{\mu_{n,j}}}$, when assuming that the residuals ϵ_i are independent and identically Weibull distributed (de Dios Ortuzar & Willumsen, 2011).

Assignment

When it is known which origin-destination relations are used, how much they are used and which modes are taken, it is possible to assign the trips to the road or transit network. Although the shortest path between two point on a plane is described by a straight line, it is impossible to travel that way. Therefore, using automobile or cycling, a path consisting of road segments should be used, while transit travellers use a path that consists of different pre-defined route segments (Chen & Liew, 2003).

More literature regarding the network assignment can be found in 2.2: Assignment.

Combining the steps

The models or steps described above can be implemented in isolation, but in transport modelling they are often combined, because they are dependent on each other. The most obvious way to treat this is to consider these steps sequentially (Figure 2a) and not having feedback loops. However, this is often not desirable, because steps require information that is estimated in later steps and these later steps can provide different information than assumed earlier (thus the model in itself is not consistent). Trip distribution, for example, depends on the travel times which are calculated in the network assignment step. Some researches suggested that an iterative approach will converge to the long-run equilibrium that is going to be realized. Others suggested that it makes sense to iterate just because the results are internally consistent (Chen & Liew, 2003).

Iterating the four steps of the transport model can be done in several ways. Current common practice is to do a simultaneous distribution and modal split (Figure 2b), so that the distribution is dependent on the travel times of different modes. Another possibility is to see the model split determination as part of the assignment: this can be applied for modelling multi-modal trips (Figure 2c).

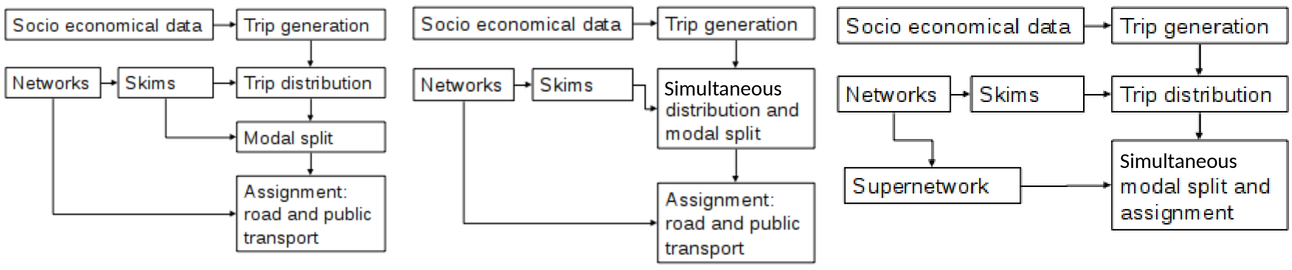


Figure 2: Four-step model iteration possibilities (Brands, 2014)

2.1.4 Model errors

One should consider the errors that may appear during building, calibrating and forecasting with models. Different types of errors are distinguished (de Dios Ortuzar & Willumsen, 2011):

- Measurement errors: these errors are due to inaccuracies when registering or measuring the data in the base year. Also estimates of variables in the future, such as fuel prices, are subject to measurement errors.
- Sampling errors: these errors come up because finite data sets are used to make estimates about the total population. Sample size must be increased to reduce this error.
- Computational errors: these errors arise when analytical solutions are not available, but are typically small compared to other errors. Congested networks could be an exception: in this case, bigger errors in assignment are possible due to more required feedback loops in the transport model. The more iterations and thus computational costs, the more accuracy.
- Specification errors: when the phenomenon being modelled is not well understood, these errors arise. These include exclusion of a relevant variable or exclusion of an irrelevant variable.
- Transfer errors: these errors come up when transferring a model set up for a particular area or time period to another context.
- Aggregation errors: these errors arise because forecasts are made for groups of people, while modelling should be done at an individual level to capture individual differences better.

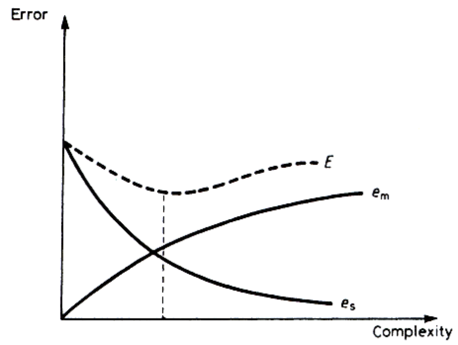


Figure 3: Error as function of model complexity

The total error is a combination of errors, which is shown in Figure 3.

Here, e_m is the measurement error and e_s is the specification error. The total error is $E = \sqrt{(e_s)^2 + (e_m)^2}$. Increasing the complexity till a certain point can improve the model quality, but increasing it too much may decrease this model quality.

2.1.5 Conclusion

This paragraph outlined the most important issues regarding transport modelling. The four-step model is the most widely used, although disadvantages exist. Of importance for the research are the inputs that the different model steps have and how zone aggregation influences the outcomes of model steps.

Attraction/production and the first iteration of distribution do not require data from the assignment step. Distribution and modal split do require data from the assignment step: generalized cost (or often travel time) is needed for the gravity model and the discrete choice model. Higher computational and aggregation errors might be acceptable as input for these steps in order to reduce computational costs. This way, the assignment results that are used as input for distribution and modal split may be less accurate, but if the effects on the eventual results are minimal, this could be appropriate in order to reduce computational costs.

Also for the final determination of network assignment results, zone and network aggregation may result in improvements in computational costs. Of course, merging zones in the assignment phase increases the aggregation error. By using sub-zones, this effect may be smaller. This can lead, also in the final assignment run, to a more favourable trade-off between accuracy and computational costs.

The difference between assignment output that is needed as input for other model steps and the final assignment output shows that the desired level of accuracy that results from aggregation techniques can differ.

2.2 Assignment

Almost all path choice models use the principle that travellers use the best path that is available for their trip. Here, the best path can be given again in terms of travel time, cost and comfort among others. An important assumption here is that travellers do have all the information they need to consider all the choices, that they are behaving independent and are identical in their changing behaviour and always make the correct route decisions (Sheffi, 1985). Infrequent travellers could, for example, take the most obvious route. Nowadays, modern techniques such as navigation systems and mobility applications could improve the information available to travellers.

The road transportation network consists of nodes, representing intersections, and links, representing road segments between intersections. Links have some properties: length (km), maximum speed (km/h) and capacity (veh/h). Having these properties, the following static transport assignment methods are possible (de Dios Ortuzar & Willumsen, 2011):

- All-or-nothing: all trips between a particular origin and destination are assigned to the shortest route and no trips to other routes;
- Multiple routing: trips are assigned proportional to routes according to attractiveness;
- User equilibrium: depending on the varying level of service of different routes, trips are assigned so that an equilibrium can settle down.

The assignment of traffic to different routes can differ, depending how to deal with link capacity:

- Capacity restraint: the link-cost function makes busy routes less attractive but traffic flows can exceed capacity. This type is further discussed here.
- Capacity constraint: the modelled traffic cannot exceed link capacity.

2.2.1 Deterministic and static user equilibrium

Transport network assignment models find solutions for the Traffic Assignment Problem by equilibrating demand (desired trips between origin and destinations) with supply (available road or transit network). This problem can be described as link costs that depend on the traffic volume on that link. The traffic volume on the other hand depends on the link cost (Dafermos & Sparrow, 1969). It is assumed that the total demand is given (fixed) for each origin-destination pair.

Wardrop stated that, in the equilibrium, the travel times on all the routes between a particular origin and destination that are actually used are equal, and less than the experienced travel time experienced by a single vehicle on any unused route (Wardrop & Whitehead, 1952). This means that no vehicle can reduce its travel time by switching to another route. It is assumed that the process, where individual commuters choose a particular path and adapt their path choice based on previous experiences, takes place so the Wardrop equilibrium will settle down in some time. This equilibrium is considered by most assignment models.

Heuristic equilibration techniques

A heuristic equilibration technique is the capacity restraint algorithm. This algorithm has the problem that it does not converge to one solution, but instead performs a flip-flop. Therefore the modified capacity restraint method performs smoothing of travel times. The algorithm is as follows:

1. Initialisation: AON assignment in unassigned network
2. Update the travel times according to the AON load
3. Perform smoothing so that the smoothed travel times are a weighted average of the travel times of the previous step and the current step
4. Perform AON to load the network again based on updated travel times
5. When the number of iterations is reached, then go to 6, otherwise n: = n + 1 and go to 2.
6. Average the loaded networks of each iteration and stop.

Another heuristic method is the incremental assignment. With this method, in every iteration a fraction of the demand flows are assigned to the network using the AON method for every fraction.

It can be expected that a user equilibrium leads to more realistic transport assignment results than an AON assignment. This is confirmed by the research by (Bovy & Jansen, 1983) who compared a network assignment model for the city of Eindhoven with traffic counts. It was found that equilibrium loads agree much better with traffic counts than an AON assignment method, even when the level of congestion is minimal. This was the case for all levels of zonal and network aggregation.

2.2.2 Dijkstra's algorithm

In every iteration, the assignment finds the shortest path between every OD pair. For this, a shortest path algorithm is used; the shortest path algorithm by Dijkstra is frequently used. Understanding the structure of this algorithm is of importance, as it forms the basis for aggregation techniques.

Dijkstra's algorithm originally was node-based, but a variant called Link-Based Single Tree Algorithm is also available (Suhng & Lee, 2013). In this research, this algorithm is used, but it still referred to as Dijkstra's algorithm.

Dijkstra's algorithm requires the following input:

- A network consisting of nodes, which are connected using weighted links. These weights represent the link (generalized) cost for the current iteration of the assignment;
- A starting point for the algorithm (the source node);
- The target nodes to where the shortest path from the source should be determined (optional, if not given, the algorithm will run till all nodes are visited).
- Whether to do an origin-based or a destination-based algorithm (default is origin-based).

The steps of the link-based algorithm are as follows:

1. Mark all nodes in the network as not visited.
2. Initialize a link queue; in this queue, links and their shortest path (generalized) cost (the cost from the source till the end of the link) are added. To start, add links to this queue which are connected with the source.
3. Consider all links in the link queue. Choose the link with the shortest path (generalized) cost; this is the *current link*. Mark the furthest node of this link, the node that was not yet visited, as visited.

4. Consider the links that are connected to the node that was just visited. Add the links that connect to a node that was not visited yet, to the link queue. Update their shortest path (generalized) cost. For those links, mark the current link as its predecessor.
5. Check whether all target nodes are visited. If not, go to step 3. Otherwise, stop the algorithm.

After performing this algorithm, one can determine the shortest path from the source to at least all target nodes. To find a path to a particular target node, the following steps are performed:
1. For the target node, check which link is on the shortest path to that node. Start generating the shortest path by adding this link to the path.
2. Get the predecessor of this link and add it to the shortest path.
3. Check whether this link is connected with the source node. If not, go to set 2.

The shortest path from the source node to a target node is now known.

2.3 Aggregation

In literature, the two main kinds of spatial aggregation are discussed that are relevant for transportation modelling. These are zone aggregation, which aggregates transport analysis zones (TAZ) and network aggregation, which has influence on the network representation in the transport model.

2.3.1 Types of aggregation

As indicated in the research problem definition (paragraph 1.2), transport analysis can be done at different scales. On one extreme, one can consider the traffic light design on an intersection, on the other extreme the introduction of pricing schemes can be considered at a national level. As the geographical scope of the problem changes, the scale of the model changes accordingly. When different models are used on a different geographical scale for the same area, questions about consistency arise (Bliemer et al., 2013; Connors & Watling, 2008).

The concept of aggregation refers to the level of detail that is used in the network and behaviour models, but can also refer to methods that are used to summarize characteristics for larger scale analysis. The following aggregation types are considered (Connors & Watling, 2008):

- Decision aggregation: the travellers that are modelled in transport models are decision makers. Decision maker aggregation deals with the distribution of attributes or characteristics across the population of decision makers.
- Traffic aggregation: this type contains an analysis traffic flow dynamics on a single link, for example the aggregation of individual car-following models.
- Zone aggregation: this type considers the method that continuous space is divided into zones so that trip demand patterns (production and attraction) are well represented.
- Network aggregation: a discrete network of congestion links is simplified or summarized in some way. The approaches can be split into four sub-types: continuum models, area speed-flow relationships, link abstraction and link extraction.

This research will deal with zone aggregation, so these methods will be further investigated.

Attempts to create a quick-scan transport model regard possible zonal and network aggregation methods. In the research of De Feijter (2012), requirements for a quick-scan model have been examined using a questionnaire. This research proposed the following aggregation methods:

- Deleting zones: small zones will be deleted and not filled;
- Hard zone simplification: merging small zones with big zones in their neighbourhood;
- Flexible zone simplification: splitting small zones up, after which each part of the zone is added to adjacent bigger zones;

- Connection link dependent zone simplification: merging a small zone with a big zone in its neighbourhood. The zone to choose can be determined using connector distance on the network;
- Least traffic flow: link extraction based on low traffic volumes in user equilibrium;
- Combining links: links can be combined based on a short distance to a parallel link.

These aggregation methods will be further investigated here.

2.3.2 Zonal aggregation

Transport analysis zones are used to reduce the number of trip origins and destinations in the transport model. The centroid of these zones are used to represent these trip origins and destinations instead of the actual trip ends, which are individual households and/or employment centres (You, Nedović-Budić, & Kim, 1998).

The MAUP

Geographical space is continuous in nature. Defining TAZ is an instance of the modifiable areal unit problem (MAUP), which indicates that there are many ways to determine boundaries and therefore areas, to split up the continuous earth surface into smaller areas (Fotheringham & Rogerson, 2009). The MAUP can be split up in two effects or problems (Páez & Scott, 2005):

- The zoning effect, which indicates where the boundaries of zones are for a fixed number of zones. The effect of zoning will be minimal when the phenomenon that is considered shows a random pattern. On the other hand, when there is a high level of spatial autocorrelation, the zoning effect will be high.
- The scale effect, which describes the inconsistency due to the use of data at a different geographical scale (aggregate or disaggregate).

Requirements for zonal structure

TAZ are assumed to accurately effect the characteristics of the units that they include. You et al. (1998) argue that defining these TAZ is crucial, as their definition will influence all the steps of transport modelling and will affect the outputs. Martínez et al. (2009) developed an TAZ definition algorithm that uses individual household data. They consider the following constraints for defining TAZ:

- Trip generation/attraction homogeneity: a TAZ should be homogeneous in population, socio-economic and land use characteristics (Chang, Khatib, & Ou, 2002). This is mainly important for the trip generation and distribution steps of the model (Baass, 1980);
- Contiguity (one piece) and convexity of zones: the units that compose a TAZ should be adjacent, otherwise the trips ends are modelled to and from a central point, while in reality these are much more scattered. It is noted that contiguity and homogeneity are sometimes mutually exclusive and thus hard to satisfy at the same time (You et al., 1998). Because contiguity affects the assignment phase of the transport model, this constraint is preferred for this research. Convexity of TAZ is desired;
- Compactness of TAZ shapes: the ideal shape of a TAZ is a circle (Baass, 1980) so that the single trip end assumption is not violated too much (same reason as for contiguity). This constraint is also of importance for the assignment step;
- Exclusiveness of zones: this indicates that there may be no islands of other zones between a zone, for the same reason;
- Equity in terms of trip generation: this means that there should be a small standard deviation of trip generation across zones;
- Adjustment of TAZ boundaries to political, administrative or statistical boundaries: this is important from a planning point of view (Baass, 1980);
- Respect of physical separators;

- Decision maker’s preferences are considered in the determination of the number of TAZ;
- Avoid main roads as zone boundaries;
- Zone size so that TAZ centroid assumption causes not too big aggregation errors;
- Minimization of intra-zonal trips: this means retaining a maximum of interaction between newly established zones, which is important for trip distribution and assignment models (Baass, 1980). Not doing so results in undesirable intra-zonal trips which are hard to model in the assignment step of the transport model;
- Maximisation of the statistical precision of the estimation of the OD matrix cell (mainly important for trip generation/attraction).

Chang et al. (2002) used some of these requirements and studied the effects of zoning structure and network detail on the outcomes of transport modelling. The V/A ratio is used, which is the ratio between the estimated traffic volume (V) and the ground count (A). They also used the root mean square error (E_{RMS}) to represent the V/A ratios in a single number (lower is better). It was found that smaller TAZ generated shorter trip lengths, higher proportions of inter-zonal trips, better V/A ratios and a lower E_{RMS}, which confirms the requirement that smaller TAZ will achieve better results. The level of network detail impacted E_{RMS} in two ways. Larger TAZ resulted in lower E_{RMS} values on a less detailed network. On the other hand, the detailed network results in lower E_{RMS} values regardless the TAZ scale. This means that matching the level of network detail with the zone aggregation level is of importance.

This is confirmed by an earlier research by Bovy & Jansen (1983), who state that refining the network and TAZ details always improved assignment outcomes. However, beyond a certain level further refinement increases accuracy only marginally. They used larger zones and link extraction to match the level of detail for TAZ and the network.

Daganzo (1980) refers to the spatial aggregation problem to describe that centroids of TAZ mis-allocate all the traffic to/from a zone to one single point. More zones and centroids could be used to solve this problem, but this increases the computational costs too much. Therefore, more centroids (the sub-centroids) to each TAZ are added so that the trip end are distributed over the entire zone. The sub-centroids in every zone are connected to the access nodes for the corresponding zone. The Frank-Wolfe algorithm is adapted in steps 0 and 2 so that the AON assignment is done in two stages. In the first stage, the shortest paths between access nodes of both the origin and the destination TAZ are computed. In the second stage, the best access nodes are found to connect any given pair of sub-centroids. This way, shared routes between the zones can be used to decrease the computational costs.

Hagen-Zanker & Jin (2015) used the knowledge that smaller zones lead to improved precision and developed the concept of adaptive zoning. They formulate the problem of large TAZ so that local traffic near the zone centroid is overestimated and that it is underestimated else in the TAZ. The error in the route is mainly near the origin and the destination: only for the extremes of the trip, locational accuracy is crucial. Therefore, an adaptive zoning method is adopted that uses small zones at one side of the route and large zones at the other side of the route. Using the one or the other half of the those routes leads to an assignment procedure that requires less computational cost and is thus able to improve accuracy.

Centroid and connector placement

The location of the centroid in a TAZ and its connector to the transport network is an issue not much considered in the literature. Often, transport modelling packages automatically generate the centroid on the location of the TAZ geometric centre. Chang et al. (2002): considered four types of centroid location: the geometric centre, the location of the biggest city in the zone, the population-weighted centre and the household-density-weighted centre. It was concluded that the effect of derived centroids instead of geometric centroid increased with the size of the TAZ. However, it was noted that additional effort in computing derived centroids is not always justified without further study.

The isolated problem of the connector location has gained attention from Sean Qian & Zhang (2012). Centroids of TAZ are connected to the roadway network by centroid connectors. The discussed method is to choose roadway nodes with intensive trip generation/attraction as the nodes to connect the connectors with. However, it is stated that the nodes with intensive trip generation/attraction may not be the actual access/egress points. Therefore in the research, a connector optimization algorithm is proposed, where the location of connectors and their travel times are chosen, in order for the maximum intensity/capacity ratio of some characteristic links to be minimized. Examples in the study showed that this algorithm produced more realistic flow patterns. However, the research does not justify that connecting with an existing node may misrepresent junctions as they are in reality.

2.3.3 Network aggregation

Unlike zonal aggregation which deals with continuous geographical space, network aggregation deals with a discrete network. Assuming that a detailed network representation is available, the following methods are available to aggregate networks (Connors & Watling, 2008):

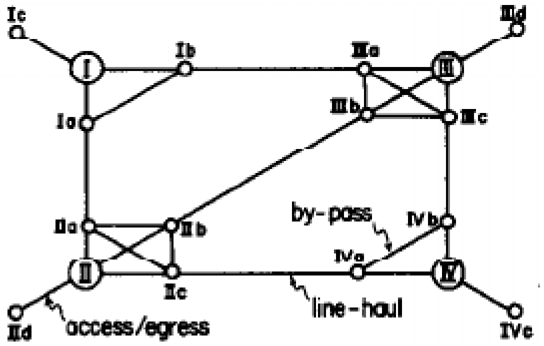


Figure 4: Link categories of network abstraction (Chan, 1976)

- Continuum models: a dense urban network is represented as a continuum, where flows are represented by a vector field;
- Area speed-flow relationships: this approach uses the macroscopic fundamental diagram;
- Link extraction: links that are deemed not important for the assignment to the network are deleted, often in an 'ad hoc' fashion. Often the aggregated network does not replicate in any consistent matter the equilibrium that is found in the detailed network, except for very low traffic demand (Chan, 1976). Shortcomings are that the reduction in network capacity is hard to estimate, the diverted traffic from the removed links leads to overloading of the aggregated network and that extraction could disconnect the networks too much;
- Link abstraction: this method replaces the network nodes and links with abstract nodes and links that are able to represent them. (Connectors connecting TAZ centroids and a network node are in fact abstract links.) Link abstraction constructs an aggregate link between two points on a map that are served by a set of detailed links. The aggregate link should have the same level of service as the detailed links that is represents (Chan, 1976).

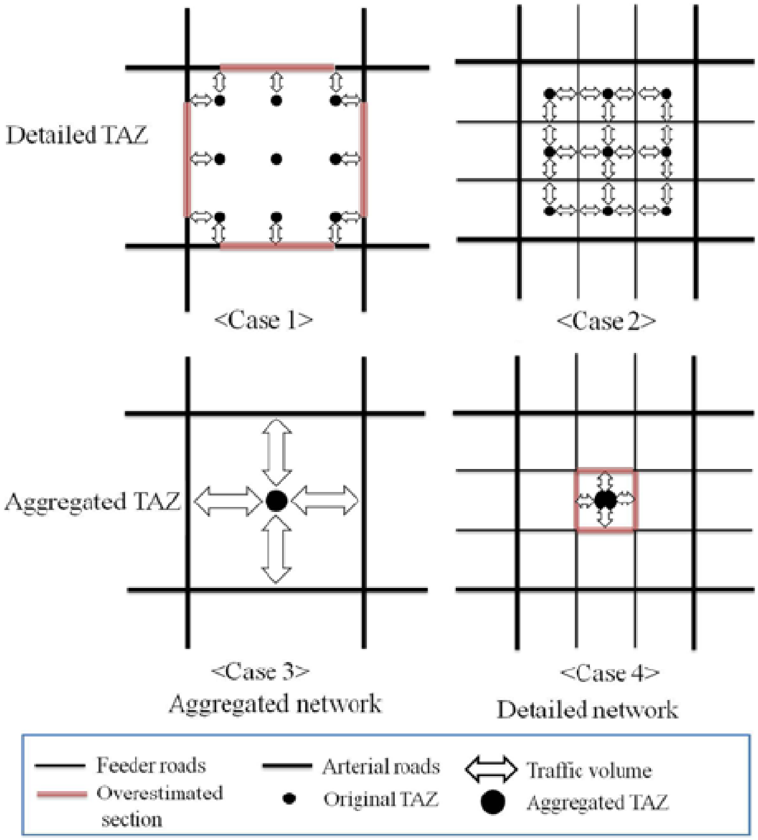


Figure 5: Consistency between network and zone detail (Jeon et al., 2012)

Chan (1976) defines three requirements that transport networks must have after aggregation, so that it is desirable for most evaluation purposes:

- The total trip time should be more or less the same in the detailed network and in the aggregated network (conservation of travel);
- The average intra-zonal travel time over the whole aggregated network (the average trip length) is the same as the average of the detailed network;
- The trip volumes between two aggregated zones are more or less the same as detailed traffic counts between the aggregated zones.

Therefore, after evaluating different kinds of abstract networks, Chan took the following approach for the network aggregation. He first aggregated TAZ in four macro-zones. Then he defined four different link types: access/egress links, bypass links, line-haul links and intra-links. The access/egress links connect the original TAZ centroid with a line-haul link. Line-haul links connect aggregated zones. There are only line-haul links there, where physical counterparts exist in the detailed network. Bypass links facilitate turning movements, so that an aggregated zone can be bypassed in order to reach another aggregated zone. Finally the intra-links serve the "internal" traffic in the aggregated zones. An advantage is that the traffic over line-haul links can be clearly separated from the traffic that stays in the aggregated zone. This approach has the advantage of being a simplified network but at the same time retaining most of the essential properties of network flow. The disadvantage of this method are that the interaction between different traffic flows on the same roads disappears.

Jeon, Kho, Park, & Kim (2012) also acknowledge that TAZ structures and network models are closely related. Much difference in detail may lead to false conclusions. When there are many TAZ in a less detailed network, the load on higher-class roads would be overestimated (case 1). On the other hand, few TAZ in a detailed network overestimates the local roads near the centroid (case 4). The paper assumes that aggregation of TAZ structures and the network detail have more influence on the assignment step than on the attraction/production, distribution and modal split steps. The authors adopt the method of the 'hole': by removing lower-class links (network extraction), holes are made between the remaining links, and the TAZ of each hole are merged to a new macro-TAZ. It was concluded that TAZ structure and the network detail have considerable effects on the results of transport modelling. As may be expected, aggregation errors are mostly caused by the flow shift from lower-class links to higher-class ones. As a last note, the authors state that the conclusions of this study are the best suited for cities that are geographically similar to Seoul city. Generalisation of the results is needed.

2.3.4 Conclusion

This literature review chapter considered literature regarding aggregation techniques. Studies that regarded zone aggregation investigated the requirements that newly formed zones should have. In the review, it is identified which TAZ determination requirements are especially of importance for the assignment step. However, these requirements have been determined for the sake of traditional zoning. This research investigates whether these requirements are still important for aggregated zones. Furthermore, the question arises whether the zone formation criteria can still be fulfilled when zones are aggregated.

Researches have found that increasing the number of TAZ, improves the accuracy. However, a lot less is said about the degree of accuracy losses due to TAZ aggregation. This research defines that degree of accuracy loss as a result of using aggregation, so that different considerations can be made depending on the requirements of output.

Some researches adapted the TAZ definition to decrease computational cost and/or improve accuracy, such as the adaptive zoning method or using more centroids in one TAZ. These researches do not consider the network definition in relation to these zones and a question arises whether that is valid. This research implements zone aggregation, and should consider whether it is necessary to use network aggregation as well. This is because it turns out that zone and network definitions are strongly related to each other: when the one is aggregated and the other not, a mismatch of assigned traffic can occur.

3 Aggregation principles

This chapter describes the principles that form the basis for understanding the impact of aggregation. In section 2.2.2 in the theoretical framework, Dijkstra’s algorithm was described, which is implemented in this research and forms the basis of all assignment methods. Section 3.1 reports on the scalability of the algorithm in general. Section 3.2 describes a first comparison between zone and network aggregation based on the scalability of Dijkstra’s algorithm.

3.1 Scalability of Dijkstra’s algorithm

The calculation time of the Dijkstra's algorithm (more specifically, the Link-Based Single Tree Algorithm) depends firstly on the number of links; every link is considered one time, requiring O(l), where l is the number of links in the network.

Furthermore, the calculation time of adding links to and extracting links from the link queue is important. These processes can be implemented most efficiently by using a “priority queue”. This means that the queue is always saved in sorted form (sorted according to shortest path cost from the source). When inserting a link in the queue (in step 4 of Dijkstra's algorithm, see section 2.2.2), the link must be inserted at the right place in the queue, so the queue remains sorted. This calculation requires at most O(log l). Extracting a link from the queue (in step 3) can then be performed independently of the length of the priority queue, taking O(1). Contrary to the node-based algorithm, the link-based version does not require changing the priority of elements in the queue.

Because every link is inserted and extracted from the priority queue at most one time, the total required time for Dijkstra's link-based algorithm with one source and multiple targets is assumed to be O(l log l). This is called one Dijkstra *sub-problem*.

To test whether the implementation of Dijkstra’s algorithm really fits this description of complexity, the algorithm is applied to networks with various sizes, which are the sub-problems. These networks are automatically generated, consisting of a square-shaped street pattern.

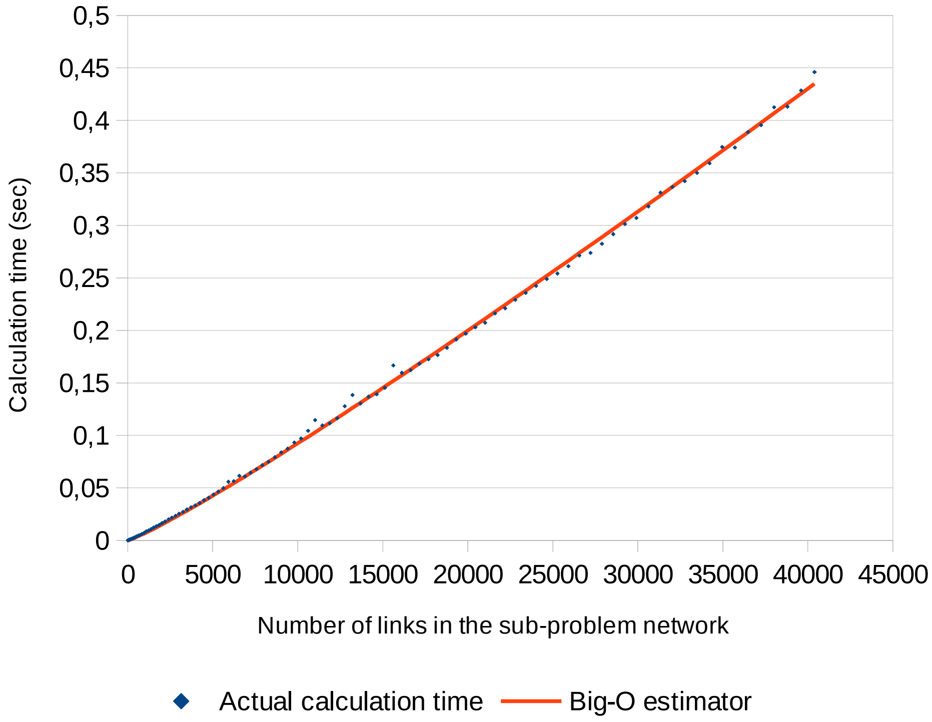


Figure 6: Calculation times of Dijkstra's algorithm for different network sizes

The Big-O estimator is able to estimate calculation times, using the formula $t=c*l\log l$, where t is the estimated calculation time, l is the number of links in the network and c is a constant, depending on the actual implementation of the algorithm and the computer speed. Dijkstra's algorithm implemented in Ruby (v2.3) reveals c ≈ 1/94000. Figure 6 shows that the Big-O estimator fits the measured computational times very well.

As we have seen, the complexity of one sub-problem of Dijkstra's algorithm is:

$$O(subproblem)=O(l\log l)$$

where l is the number of links in the network.

For a regular static assignment, the number of sub-problems equals the number of centroids, where each centroid is the source of its own sub-problem. The sum of all sub-problems in one iteration of the assignment is the *problem*; its computational cost is:

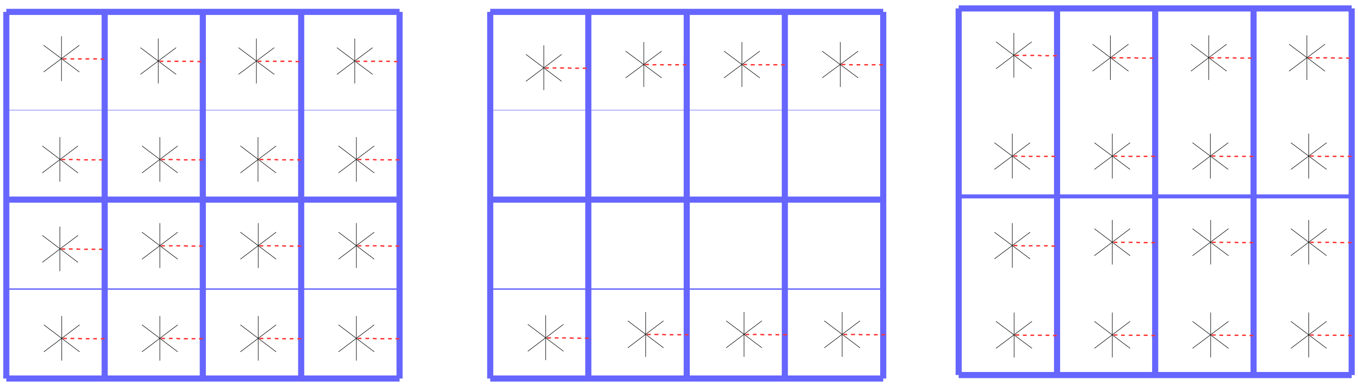
$$O(iteration)=O(j\,l\log l)$$

where j is the number of sub-problems, i.e. the number of centroids.

3.2 Basic aggregation comparison

The scalability of the Dijkstra's algorithm is constrained by the number of links in each sub-problem. In this section, the most basic forms of zone and network aggregation are tested for scalability. The basic form of zone aggregation is defined as: merging zones (with individual zone centroids) to form a bigger aggregated zone with a single zone centroid. The basic form of network aggregation is defined as: remove the links of the lowest order.

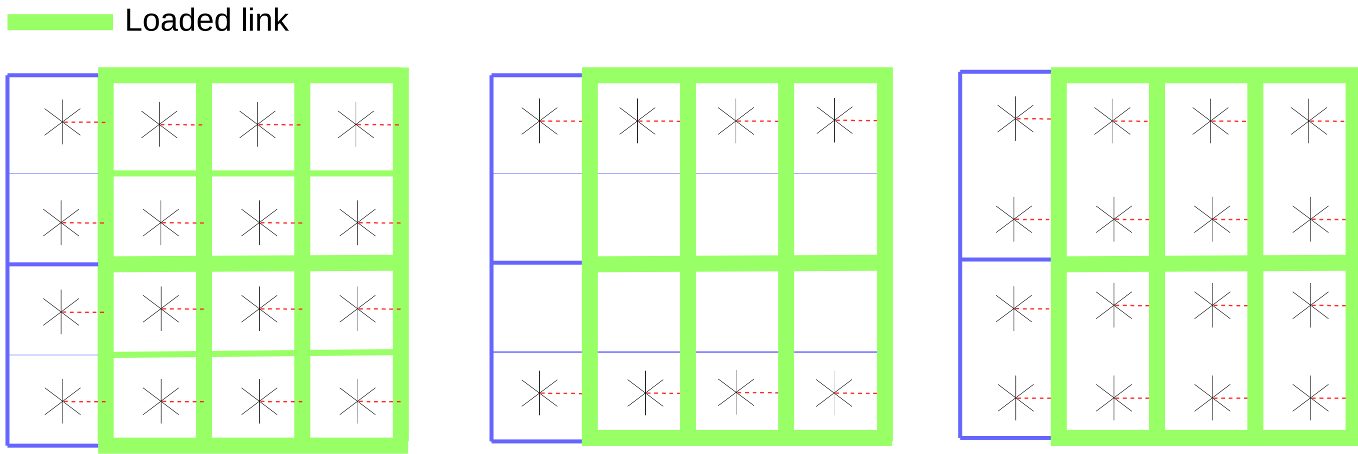
To illustrate both forms of aggregation, consider the synthetic network in Figure 7a. It consists of 16 centroids, each connected to one place in the network. The network consists of 40 links, of which 8 are of low order (local roads) and 32 of high order (main roads). (Connectors are not included.) The higher order links have significantly higher capacities and free flow speeds than the lower order links. Note that the boundaries of the zones are not shown.
Figure 7b shows the same network with zone aggregation: each two centroids (zones) are merged into one centroid, resulting in 8 centroids. Figure 7c shows the original network after network aggregation is applied, resulting in 22 links of high order.



(a) The original network (b) Zone aggregation (c)Network aggregation

Figure 7: Artificial network

This network can be loaded with a OD matrix with arbitrary traffic demand. Figure 8a shows the links with traffic in the original network; it is assumed that lower order links are used as well as they provide short routes between some centroids. Figure 8b shows the loaded network with zone aggregation; the low order links are not used (they are not part of the shortest path between any centroid combination). Figure 8c shows the loaded network with network aggregation; low order links are naturally not used as they are not present.



(a) The original network (n = 16 zones, l = 40 links)

(b) Zone aggregation (n_z = 8 zones, l_z = 40 links)

(c) Network aggregation (n_n = 16 zones, l_n = 22 links)

Figure 8: Loaded artificial network

Figure 8 shows that the link loads, resulting from these forms of zone and network aggregation, are equal. It is assumed that for real networks, these outcomes resulting from these methods are still comparable and that they are equally valuable for policy and decision makers, while being less valuable than the original loaded network.

In general, the number of sub-problems and the number of links in a square network like Figure 7 depends on the number of original centroids. After performing the basic form of zone aggregation (merging two zones into one) or network aggregation (removing low level links), the number of sub-problems and the number of links is as follows:

	Size of the problem (number of sub-problems)	Size of each sub-problem (number of links)
Regular assignment	$n=n_0$	$l=2n_0+2\sqrt{n_0}$
Zone aggregation	$n_z=\frac{1}{2}n_0$	$l_z=2n_0+2\sqrt{n_0}$
Network aggregation	$n_n=n_0$	$l_n=n_0+1\frac{1}{2}\sqrt{n_0}$

Table 2: Number of sub-problems and links for basic aggregation variants

where n is the number of sub-problems, l is the sub-problem size (number of links) and n_0 is the number of original centroids.

When applying the basic form of zone aggregation, the number of sub-problems is half the original number of sub-problems. The size of each sub-problem is equal to the size of each sub-problem in the regular assignment; the network has not changed.

When applying the basic form of network aggregation, the number of sub-problems is equal to the number of sub-problems without aggregation. However, the size of each sub-problem is less than in the regular assignment; for large networks ($n_0 \rightarrow \infty$) the size of each sub-problem is approximately half the size of sub-problems without aggregation.

Thus, when the number of zones is half the original number of zones, as applied in this example of zone aggregation, computational complexity halves. When the number of links in the network is approximately

half (which is true for large networks), as applied in the example of network aggregation, computational cost decrease according to *j l log l*, which results in almost halving the computational cost.
The effect on computational costs of both aggregation methods can be considered equal. Of course, these aggregation methods are not the most advanced methods and consequently lead to a significant decrease of accuracy. However, they decrease the computational costs also significantly, by around half.

3.3 Conclusion

A sub-problem is defined as one Dijkstra’s algorithm with one source and multiple targets. The size and computational cost of this sub-problem depends on the number of links that is being considered before the algorithm is finished. Dijkstra’s algorithm can stop when all targets have been found. Note that the source does not necessary has to be the origin, it can also be a destination-based sub-problem, where the source is the destination and the targets the origins.

The number of sub-problems as well as their (possible varying) sizes determine the computational cost of one iteration of the assignment step of the four-step model. The computational costs for the full assignment step depends on the number of iterations and their individual computational costs. In theory, the cost of every iteration is equal; in practice it can differ.

The example in this chapter, which is a simple and regular network, showed that the most basic forms of zone and network aggregation lead to similar reductions of computational costs and assignment results. Literature shows that the network aggregation was mainly used for getting an impression on the main link flows; that makes it harder to adapt to the goals and boundary conditions of this research: to estimate traffic flows on all links that are in the detailed network and use all original zones. The zone aggregation methods described in the literature, such as shared routes and adaptive zoning, which can use all original zones and estimate flows on all links in the network, better suit the research goals. Therefore, this research further develops and evaluates these methods.

The next chapter describes these zone aggregation methods.

4 Aggregation alternatives

Aggregation can be applied and performed in various ways, which are called the aggregation alternatives. This chapter describes the alternatives that are possible when applying zone aggregation, which are based on the aggregation principles described in chapter 3. These alternatives consist of “building blocks”; the ones that are different in the alternatives are described in section 4.1. The most important combinations of these building blocks, the main alternatives, are consequently described section 4.2.

The literature showed that, in order to overcome the paradox between computational costs and accuracy, different new methodologies have emerged, such as adaptive zoning and shared routes. These are the methods that fit well within the concept of zone aggregation, which is chosen as the main aggregation method in this research. These methods only use aggregated zones in the assignment step and are able to estimate loads on all links in the non-aggregated network. This chapter further elaborates on these methods and how they are adapted to fulfil the research objective and meet the boundary conditions.

4.1 Important building blocks

This section describes the most important building blocks that explain the difference between the chosen aggregation alternatives. Other building blocks that can be used in aggregation but do not explain the difference between the described alternatives in this chapter, are described in Appendix B.

The two main building blocks are “aggregated zone hierarchy” and “route reconstruction”, where each building block is available in two options.

Building block	Options	
Aggregated zone hierarchy	Fixed aggregated zone definition	Adaptive zoning: every original zone interacts with different aggregated zones
Route reconstruction	Reconstruct total route with first/last mile routes	Reconstruct total route with two-sided routes

Table 3: Main building blocks used in aggregation alternatives

Both building blocks are used in the chosen aggregation alternatives. Aggregated zone hierarchy is about the structure of aggregated zones, which form the basis for aggregated assignment methods. Route reconstruction is about the way of constructing a route for OD pairs.

4.1.1 Aggregated zone hierarchy

Using aggregated zones is the key component for the zone and network aggregation methods. For this building block, the options are to use a fixed aggregated zone layer or adaptive zoning.

Fixed aggregated zone definition
One fixed higher level of aggregated zones means that every original centroid is included in exactly one aggregated zone. Every aggregated zone consists of one or more original centroids. The aggregated zones are complementary to each other and can be at, for example, neighbourhood, district or municipality level (Figure 9).

Adaptive zoning
Adaptive zoning means that every original zone is included in multiple aggregated zones. Adaptive zoning acknowledges the fact that origin destination relations with greater distance are less frequently used. The idea of adaptive zoning is that every original zone (centroid) interacts with a different set of aggregated zones: small original or aggregated zones nearby and big aggregated zones at distance. The idea is based on Hagen-Zanker & Jin (2015) but adapted in order to incorporate existing zone structures.

The aggregated zones are layered and are at neighbourhood (zone level), district and at municipality level. This set of (aggregated) zones is called the *neighbourhood*[1] of that original zone.

In the highest level of aggregation (level 0) this *neighbourhood* consists of the original zones within the neighbourhood level of that original zone, plus the neighbourhood zones within the district level of that original zone, plus the district zones within the municipality of that original zone, plus all the remaining municipalities in the transport model. When higher zone levels are available, those levels can be used as well.

Figure 9: Original centroids within different levels of aggregated zones (case study: The Hague)

In level 1 of adaptive zoning (one level less aggregation), the *neighbourhood* shifts one level and increases in size: it consists of all original zones within the district level of that original zone, plus the remaining district zones in the transport model.

In general, the neighbourhood of an original zone is:

$$neighbourhood(l) = Z_0(l) \cap Z_1(l+1) \cap \ldots \cap Z_L$$

where $Z_0(l)$ consists of all original zones in the aggregated zone at level l in which the original zone is included, and Z_L consists of all remaining aggregated zones on the highest used level.

When the *neighbourhood* on a particular level is defined, the *inverse neighbourhood* can be defined. The *inverse neighbourhood* of an aggregated zone (on any level) is defined by all original zones which have that

1. The term neighbourhood can either refer to (i) a zone at the neighbourhood level, or (ii) the set of zones that an original zone interacts with in adaptive zoning (in the latter case, *neighbourhood* is in italic typeface).

aggregated zone in their *neighbourhood*. The efficiency gains of adaptive zoning comes from the use of the *inverse neighbourhoods* of aggregated zones as targets of Dijkstra's algorithm: there are few large aggregated zones with a large sub-problem and many small aggregated zones with a small sub-problem.
Appendix A clarifies the definitions of *neighbourhood* and *inverse neighbourhood* using a small network example.

4.1.2 Route reconstruction

This building block is used to reconstruct a route for every OD pair. This means that the original demand OD matrix is not aggregated. Instead, finding a route for every OD pair is done in a more efficient way than in a regular assignment, using either first/last mile and shared routes, or using two-sided routes. Thus, this building block consists of these two options with which the total routes can be found:

First/last mile and shared routes
The first option in this building block consists of using first and last mile routes within aggregated zones in addition to a shared route between two aggregated zones. It is based on Raadsen, Schilpzand, & Mein (2009) and Benezech (2011), who both use some kind of shared route between aggregated zones, in addition to access and egress routes to connect original zones with the shared route.

This methods starts by generating shared routes between aggregated zone centroids. The location of these centroids is thus of importance for determination of the shared route. The shared route is cut within the aggregated zones; more precisely, from the aggregated zone centroid until the "boundary node". This boundary node is defined as the first node that is encountered after entering the aggregated zone.
After that, the first and last mile routes can be determined, which stretch between original zones and boundary nodes within an aggregated zone. The total route between any origin and destination in different aggregated zones, is the sequence of first mile, shared route and last mile links and is illustrated in Figure 10.

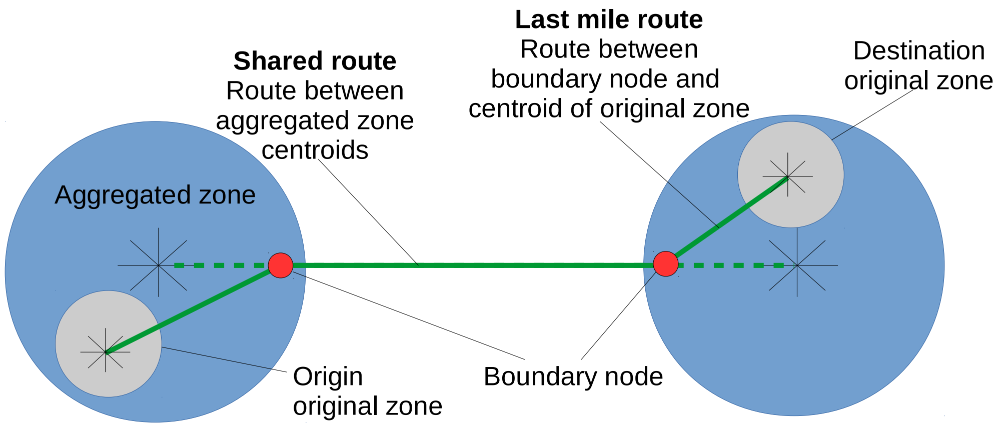


Figure 10: Route reconstruction using first/last mile and shared routes

Because this method generates the full route from origin to destination out of different routes, it can happen that a certain link is passed twice (in both directions). Therefore, the links that occur twice and in opposite direction in the full route are removed.

Two-sided routes
Two-sided route finding means that two routes are generated and reconstructed into one route. One path goes from the origin original centroid to the destination aggregated zone (path *a*, detailed near the origin). The other route goes from the origin aggregated zone to the destination original centroid (path *b*, detailed near the destination).

Two options are available to reconstruct the total route between origin and destination.

The first option is to find a common node in both routes. This occurs when both paths have at least one node in common. The total path is constructed so that the first part of path *a*, from the origin till the common node, is used and the second part of path *b*, from the common node to the destination (Figure 11).

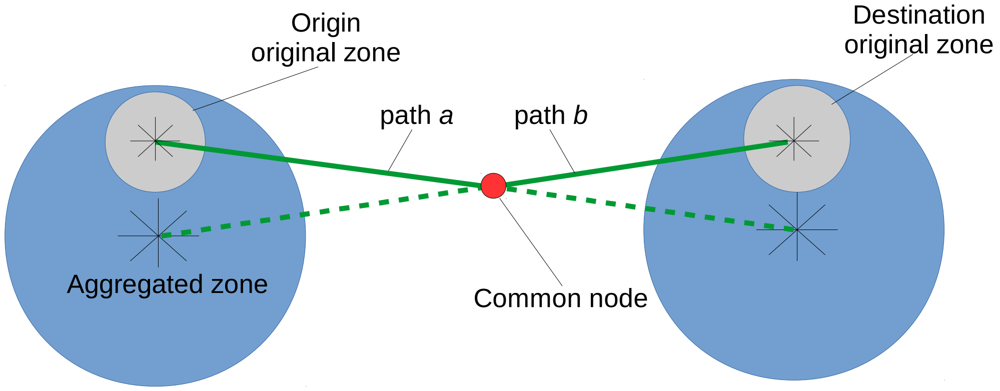


Figure 11: Route reconstruction using two-sided routes: common node

The second option is to use a first mile or last mile path, which is connected with the aggregated zone side of path *a* or *b* to form a total path between origin and destination. This method works as follows. The distances between the origin original centroid and origin aggregated zone centroid (d_1) and between the destination original centroid and destination aggregated zone centroid (d_2) are calculated using the Euclidean distance.

When d_1 is smaller than d_2, a path between the origin original centroid and origin aggregated zone centroid is calculated, which is the first mile path. This is done using a destination-based Dijkstra's algorithm with the aggregated zone centroid as source and all original centroids in that aggregated zone as targets. Together with path *b*, the total path between origin and destination is completed (path *a* is not used).

When d_2 is smaller than d_1, a path between the destination original centroid and destination aggregated zone centroid is calculated, the last mile path. This is done using an origin-based Dijkstra's algorithm with the aggregated zone centroid as source and all original centroids in that aggregated zone as targets. In this case, the total path is completed using path *a* (path *b* is not used). Figure 12 illustrates this case.

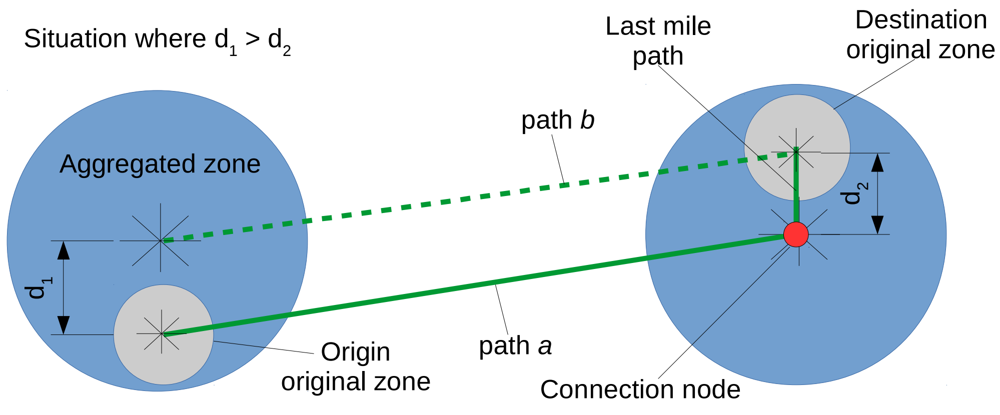


Figure 12: Route reconstruction using two-sided routes: first or last mile path

The first option, to find a common node, is preferred as it has the most direct paths and computational costs are lower (no new Dijkstra's algorithms are needed). Therefore this method is performed first. However, it can happen that no common node can be found in both paths. This is possible when path *a* and *b* do not intersect or when they intersect at a grade separation. When this is the case, the second method is performed.

As mentioned, the Dijkstra's algorithms performed in the second method have all original zone centroids as target. In principle it would be sufficient to use only the original zone centroid for that particular first or last mile path. However, by using all zone centroids in the aggregated zone, all shortest paths from the

aggregated zone centroid are already known and saved in memory. This saves computational costs in case a first or last mile path in the same aggregated zone must again be determined.
The concept of two-sided routes is inspired by Hagen-Zanker & Jin (2015), who assign an OD pair load to the most accurate halves of both paths, but those paths do not necessarily form one single path.

4.2 Alternative descriptions

This section describes the aggregated assignment alternatives that result from choosing various options from the described building blocks.

4.2.1 Alternative 1: fixed zones with first/last mile routes (FZ-FL)

This aggregation method uses fixed aggregated zones and constructs routes for every OD pair using first/last mile paths. That means, the options from the following building blocks are used:

Aggregated zone hierarchy	Fixed aggregated zone definition
Route reconstruction	Reconstruct total route with first/last mile routes

For this alternative, the level of aggregation is determined by the level of aggregated zones that is used (for example, neighbourhood or district level). Only a single layer of aggregated zones is used.

4.2.2 Alternative 2: fixed zones with two-sided routes (FZ-TS)

This aggregation method uses fixed aggregated zones too, but generates two-sided routes to reconstruct the total path for every OD pair. That means, the options from the following building blocks are used:

Aggregated zone hierarchy	Fixed aggregated zone definition
Route reconstruction	Reconstruct total route with two-sided routes

The level of aggregation is again determined by the level of aggregated zones and only a single layer is used.

4.2.3 Alternative 3: adaptive zoning with two-sided routes (AZ-TS)

This aggregation method uses adaptive zoning, which means that every original zone centroid interacts with a different set of aggregated zones (the *neighbourhood* of every original zone). Furthermore, two-sided routing is used to reconstruct a path for every OD pair. That means, the options from the following building blocks are used:

Aggregated zone hierarchy	Adaptive zoning: every original zone interacts with different aggregated zones
Route reconstruction	Reconstruct total route with two-sided routes

The level of aggregation for this alternative is determined by the size of the *neighbourhood* of every original zone centroid (either level 0 or 1).

5 Estimations

In this chapter, we estimate computational costs and accuracy for different aggregation methods from the alternative study in chapter 4 and compare them with the regular assignment. It is similar to chapter 3 in calculating computational costs based on Big-O notation, but the difference to that approach it that it is not assumed that the accuracy for the alternatives is equal.

To a certain degree, it is possible to estimate evaluation criteria outcomes for each alternative before really implementing it:

- It is hard to measure the impact on accuracy before implementing the alternative. However, educated guesses are possible. The actual accuracy for these advanced aggregation methods can only be determined in a case study, which is performed in chapter 6.
- It is possible to estimate the complexity of the used algorithms before actually implementing the alternative, using the big-O notation for Dijkstra's algorithm. However, the actual running times depend, apart from the complexity, also on the implementation itself and on the transport network.

One iteration of a transport assignment performed using any (aggregated) method requires

$$O(iteration)=O(j\,l\log l)$$

where *j* is the number of sub-problems and *l* is the size of the sub-problem.

In this chapter, *n* is the number of original zone centroids in the whole transport model study area. The number of original zone centroids per aggregated zone is indicated with *h*.

5.1 Regular assignment

Consider a square network of *n* centroids (Figure 13). The cost of one iteration of the regular static assignment is $O(regular)=O(n\,l\log l)$, as discussed in chapter 3. Recall that this means that there are *n* sub-problems (Dijkstra's algorithm, which has one source and many targets) and the size of one such sub-problem is *l* links, which is the full area of the transport model.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Area of 1 original zone: *n*

Figure 13: Square network with n*=16 original zones*

Assumed is that the number of links in a particular area is proportional to the number of zone centroids in that area, so $n \sim l$. This assumption is true when the zone centroids are evenly spread over the area in the transport model (see appendix D for a justification of this assumption). It makes it easier to express the size of an area as function of the number of (original) centroids in that area. Then the complexity of the regular assignment is $O(regular)=O(n^2\log n)$.

5.2 Fixed aggregated zones with first/last mile paths

The alternative FZ-FL uses two kinds of sub-problems: the first/last mile routes and the shared routes.

For the first/last mile routing (including the inner traffic routes), a Dijkstra's algorithm is performed for every original centroid twice: an origin-based algorithm and a destination-based algorithm. Those algorithms search paths within the area of the aggregated zone: to the boundary nodes and to the other original centroids. These sub-problems have as source the original zone centroid and as targets all other original zone centroids and boundary nodes within the aggregated zone. Therefore the size of these sub-problems is proportional to the number of original centroids per aggregated zone, *h*.

For route searching between aggregated zones, the number of required algorithms is the number of aggregated zones (n/h). The source of this sub-problem is the centroid of the aggregated zone and the targets are the centroids of all other aggregated zones. The size of these sub-problems covers the whole transport model study area, n.

The table below summarizes these findings.

	Number of sub-problems	**Size of sub-problems**
First/last mile routes (including inner traffic)	$2n$	h
Shared routes (routes between aggregated zones)	$\frac{n}{h}$	n

Table 4: Sub-problems for fixed aggregated zones with first/last mile paths

The sub-problems that calculate the shared routes have a size of n, which is the same size as all sub-problems in the regular assignment. However, the number of these big sub-problems is a factor h smaller than the regular assignment. For example, when the number of original zones per aggregated zone is 4, the number of big sub-problems (with have size n) is 4 times smaller.

Especially for very small transport network sizes, the computational costs of calculating the first/last mile routes is significant. These should not be higher than the gains that were generated by using shared routes. For very large transport networks (n is big) and a small number of original zones per aggregated zone (h is very small in comparison to n), the size of the sub-problems that calculate the first/last mile routes can be neglected. In this case, the aggregated assignment is h times more efficient than the regular assignment.

The accuracy of the aggregated assignment also depends on the ratio between n and h. Two factors are important for the accuracy: the number of routes that are found directly, and the size of the aggregated zones for routes that are not found directly (instead using shared, first/last mile paths).
A low number of h leads to more aggregated zones and a high accuracy of routes between aggregated zones. A high number of h leads to bigger but a lower number of aggregated zones. The routes for traffic within these aggregated zones are found directly, but routes between original zones in different aggregated zones are found with less accuracy. This is especially the case for routes between original zones that are close to each other but in different aggregated zones.

5.3 Fixed aggregated zones with two-sided routes

The alternative FZ-TS uses two kinds of sub-problems: the routes for inner traffic within aggregated zones, and the two-sided routes for building routes between original zones in different aggregated zones.

For the traffic within the aggregated zones, the paths are calculated using n origin-based Dijkstra's algorithms. Here, the source of the sub-problem is the original zone centroid and the targets are the other original zones within the aggregated zone. The area size of these sub-problems is h, the number of original centroids in an aggregated zone.

For route searching between aggregated zones, the number of sub-problems is twice the number of aggregated zones (n/h), because every source forms both an origin-based algorithm and a destination-based algorithm. The source of the sub-problem is the centroid of the aggregated zone and the targets are all original zone centroids in other aggregated zones. The size of these sub-problems covers the whole transport model study area (n).

The following table summarizes these findings.

	Number of sub-problems	**Size of sub-problems**
Inner traffic routes	n	h
Two-sided routes	$2\frac{n}{h}$	n

Table 5: Sub-problems for fixed aggregated zones with two-sided routes

The sub-problems for two-sided routes have the same size of all sub-problems in the regular assignment, which is also equal to the size of sub-problems in the method FZ-FL. The number of these big sub-problems is a factor ½h smaller than the regular assignment. That is two times the number of big sub-problems in the FZ-FL method.

Contrary to the FZ-FL method, FZ-TS does not calculate first/last mile routes. Therefore, only the inner traffic routes are calculated and n sub-problems are sufficient instead of $2n$. The size of these sub-problems is h. Thus, the same reasoning applies: for very big networks with a small number of original zones per aggregated zone, the size of the sub-problems that calculate the inner traffic routes can be neglected. Then, the aggregated assignment is ½h times more efficient than the regular assignment.

As with the FZ-FL method, the ratio between n and h is important for the accuracy of the aggregated assignment method. When h increases, the number of routes that is found directly increases, but the routes between aggregated zones are constructed less accurate. When the same network and aggregated zones are used for both the FZ-FL and FZ-TS methods, the only difference in accuracy is the method of reconstructing routes.

5.4 Adaptive zoning with two-sided routes

The alternative AZ-TS uses two kinds of sub-problems: the routes for inner traffic between the lowest level of aggregated zones, and two-sided routes for building routes between original zones that are not in the same lowest level aggregated zone.

Within the lowest level of aggregated zones, original zone centroids interact directly with each other. That means that for every original zone centroid, one Dijkstra’s algorithm is performed. The targets of the sub-problem are the other original zones within this aggregated zone. Thus the total number of sub-problems on this level is equal to the number of original zones (n). The size of this lowest level of aggregated zones is equal to h.

For the higher levels of aggregated zones, the number of sub-problems and their sizes depends on the zone level. The number of sub-problems on a particular zone level is twice the number of aggregated zones on that level: one for the origin-based and one for the destination-based sub-problem. This is equal to $2n/h^l$, where l is that particular level of aggregated zones and h is both the number of original zones in the lowest level of aggregated zones and the number of aggregated zones in higher levels of aggregated zones. Both sub-problems have as source the aggregated zone centroid and as target all original zones in the *inverse neighbourhood.* These sub-problems find routes in the area of one level aggregated zone higher, and the size of this area is h^{l+1}.

These findings are summarized in the following table.

Level	Number of sub-problems	Size of sub-problems
0 (lowest level)	n	h
1	$2\frac{n}{h}$	h^2
l	$2\frac{n}{h^l}$	h^{l+1}

Table 6: Sub-problems for adaptive zoning with two-sided routes

The sub-problems on the highest level of aggregated zones L are of the same size of all sub-problems in the regular assignment, which is the whole transport network area, n. This gives $h^{L+1}=n$. The number of levels above the lowest aggregated zone level can then be calculated with $L={}^{h}\log n-1$. The number of sub-problems of the highest level is $2\frac{n}{h^L}$.

Consider the example network in Figure 14. It consists of eight original zones (l = 0), four neighbourhood zones (l = 1) and two district zones (l = L = 2). The number of sub-zones per aggregated zone (h) is two.

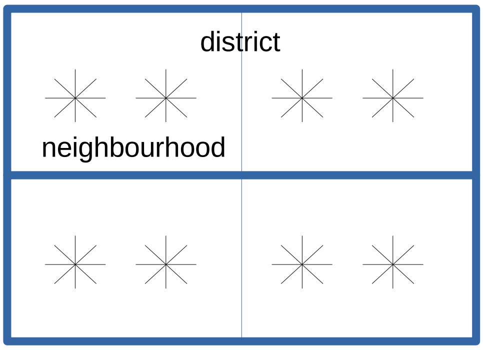


Figure 14: Example network

In this case, where there are two layers of aggregated zones (L = 2), the number of sub-problems on the highest level is $2\frac{n}{h^L}=2\frac{8}{2^2}=4$.

In order to compare this method AZ-TS with the two fixed zone methods, we have to make an assumption on the level of aggregated zones for these methods. Here we assume that this level is the district level (l = 2). For FZ-FL, the number of sub-problems for the aggregated zone level would be 2. For FZ-TS, the number of sub-problems on this highest level would be 4 as with AZ-TS. Thus, the computational costs for FZ-TS and AZ-TS on the highest level are the same.

The difference between AZ-TS and FZ-TS is in the way that routes between zones in the same district zone but different neighbourhoods zones is found. FZ-TS finds these routes directly, because level 1 (the neighbourhood zone level) does not exist for FZ-TS. The method AZ-TS however, reconstructs these routes using two-sided routes on zone level 1. The FZ-TS method's lowest level sub-problems search routes for inner traffic within the district zone, while the AZ-TS method's lowest level sub-problems search routes for inner traffic within the neighbourhood zone. Thus, whether the computational costs of AZ-TS or FZ-TS are lower, depends on which option has lower computational costs:

- four sub-problems with district size (FZ-TS), or
- two sub-problems with district size and four sub-problems with neighbourhood size (AZ-TS).

In this example, the accuracy of FZ-TS will be higher than AZ-TS, because for the first method, routes between all original zones in a district are found directly, while with the latter this is done with two-sided routing. Routes between original zones in different district zones are found in the same way in both methods.

5.5 Conclusion

When comparing the fixed zoning alternatives, we see the following. The alternative FZ-FL is h times more efficient than the regular assignment, when applied a sufficient large network. In comparison, the alternative FZ-TS is ½h times more efficient than the regular assignment. The difference is directly a result of one route versus two routes between each aggregated zone. Without further knowledge, the accuracy of both methods cannot be estimated apart from the aggregation level.

Comparing the alternatives with two-sided routes shows that AZ-TS and FZ-TS have comparable computational costs with two levels of aggregated zones. In this case, whether AZ-TS is more or less efficient than FZ-TS depends on the zone definitions in the network that is used. When more levels of aggregated zones are used, the efficiency of AZ-TS is expected to be higher. With regards to the assignment results, FZ-TS is more accurate than AZ-TS, because the routes of more OD pairs is calculated directly.

6 Results

In previous chapters, the different aggregation alternatives were described, together with the expected gain in computational costs. Although these gains can be estimated on theoretical networks, these results can vary in real-life networks. Furthermore, it was stated that the effect on the assignment outcomes cannot be estimated at all, apart from some educated guesses. Therefore, this chapter describes the results of both computational costs and assignment accuracy for a real-life transportation network. Section 6.1 starts with the evaluation framework, with which these results can be measured. Section 6.2 presents a case study where this evaluation framework is applied to the described aggregation alternatives.

6.1 Evaluation framework

In this section, the evaluation framework is described, with which the different aggregation alternatives can be compared. It gives the definitions of accuracy and computational cost, which are the two main factors for evaluation. Hereby the non-adapted, regular assignment model with maximum available zone and network detail (the assumed “ground truth”) is compared with the adapted assignment model. Both methods are performed without calibration in order to compare them purely on model characteristics.

6.1.1 Accuracy

The accuracy of the adapted model can be assessed in different ways, depending on the purpose of the assignment outcomes. The eventual transport model outcomes will, of course, be used by transportation planners. They are interested in both traffic flows and travel times and costs. Traffic flows can be used to calculate for example congestion, emissions and noise, while the generalized cost (or absolute travel time or cost) can be used to calculate accessibility.

Secondly, the outcomes of the assignment model are, in some transport models, used as input for another iteration of trip distribution, modal split and assignment. In the first iteration, these steps are based on an empty network and its travel times. However, for the long-term equilibrium, it is assumed that trip distribution and modal split adapt to the traffic and transport situation. Therefore, the second iteration uses the generalized costs that were calculated after the assignment of the first iteration. Thus, for the purpose of iterating, the re-calculated skim matrices with their generalized cost are of importance.

The following two parameters give a first sight on the results of the aggregated assignment models.

Relative difference in sum of vehicle kilometres
This indicator shows the increase or decrease of the total number of kilometres travelled by all vehicles. The sum of vehicle kilometres is calculated as follows:

$$VKT=\sum_l x_l * s_l$$

where x_l is the traffic load on a particular link l and s_l is the length of that link.

Relative difference in sum of vehicle hours
This indicator shows the increase or decrease of the total number of vehicle hours that all vehicles have spent on the roads. The sum of vehicle hours is calculated as follows:

$$VHT=\sum_l x_l * t_l=\sum_l x_l * \frac{s_l}{v_l}$$

where t_l is the time that one vehicle needs to traverse link l and v_l is the (average) speed on link l.

The following **link volume** indicators are used:

Link load ratios
A link load ratio can be calculated for every direction of every link in the network. They are calculated as the

relative load on the adapted model outcomes with respect to the load resulting from the regular assignment.
Two ways are used to assess the link load ratios: using a map of the network which displays the load difference per link and using a graph which shows relative link load differences.
Furthermore, to gain more insight, the link load ratios are also indicated per road type. For each road type, a maximum tolerable error is determined based on advice of a transport model specialist. The following road types are used:

- High order city roads. These roads include motorways and the main city roads, consisting of at least two lanes per direction. The maximum speed of these roads varies between 50 and 100 km/h. A maximum link load error of 5% is used for this road type.
- Second-order city roads. These roads include the other city roads, consisting of 1 lane per direction and with a maximum speed of 50 km/h. A maximum link load error of 10% is used for this road type.
- Local roads: These roads include small roads in neighbourhoods with a maximum speed of 30 km/h. For this road type, a maximum link load error of 20% is used.

The maximum link load error for all links together is 10%.

Link load RMSE
The link load root mean squared error (RMSE) is an indicator for the average error caused by aggregation and is calculated as follows:

$$RMSE=\sqrt{\sum_{l}\left(\frac{x_a}{x_o}-1\right)^2}$$

where x_a is the load on link *l* in the adapted model and x_o is the load on that link in the regular model.
The link load RMSE is calculated for all roads together, and for the different road types (high order city roads, second-order city roads and local roads) for more insight of the link load errors per type of road.

As stated before, these indicators are only used when the adapted assignment model is assessed for its *eventual outcomes* (the outcomes that transportation planners desire).

The following **travel time** indicators are used:

Travel time ratios
A travel time ratio can be calculated for every origin-destination (OD) pair in the demand OD matrix. It is calculated as the relative travel time resulting from the adapted assignment with respect to the travel time from the regular assignment. The travel time ratios are displayed in a graph which shows the distribution of travel time ratios.

Travel time RMSE
The travel time RMSE is an indicator for the average error in travel times caused by aggregation and is calculated similarly to the link load RMSE.

As stated before, the travel time indicators are used in order to determine to what extent the adapted assignment model is able to *generate input for trip distribution and modal split* and for its *eventual outcomes.*

6.1.2 Computational cost

To quantify the computational cost of the adapted assignment model with respect to the regular assignment model, the following indicators are used:

Actual computational time
The actual computational time is the number of seconds that the assignment model requires.

Number of considered links

The actual computational time is a good indicator for the computational costs, but also heavily depends on the implementation method and the used machine. The total number of actual considered links in all Dijkstra's algorithms, used during the assignment, is an indicator that is independent of the implementation method.

6.2 Case study: the Hague

In this section, the evaluation framework is applied to a real-life transportation network, for which we use the road network of The Hague. The road network stretches from the North Sea in the north-west to the A4 motorway in the south-east, and from the N14 ring road in the north-east to the N211 ring road in the south-west. This area contains not only the municipality of the Hague, but also parts of Leidschendam. The model is "cut" at the boundaries of the area of interest, which means that traffic from and to external areas is represented by zone centroids at the edge of the model (the "outer centroids"). The study area is shown in Figure 15.

Figure 15: The Hague study area

This network is chosen because the density of original zone centroids is fairly high, thus allowing for the different aggregation methods. Furthermore, the network is not too big so that multiple aggregation methods can be tested within reasonable computational costs.

The road network consists of around 1200 original zone centroids, of which 81 (7%) are outer centroids. These outer centroids represent the external traffic and each is connected to one road at the boundary of the study area. Furthermore, the outer centroids typically generate high levels of traffic. Aggregation of outer centroids often leads to high intensities assigned on the wrong roads. Therefore, no zone aggregation is applied on them. This means that all traffic from and to the outer centroids is assigned in the regular way. Compared to aggregating these centroids as well, computational costs increase but also the accuracy increases.

The OD matrix that is assigned represents the morning rush hour. Therefore, there is some congestion in the network at the main roads in the city.

6.3 Comparison of model characteristics

To compare the three different aggregation methods purely on their characteristics, this section presents assignment results as the result of one iteration (an all-or-nothing assignment). With one iteration, the route choices generated by the models are not influenced by link capacities. This way, the aggregation model's ability to produce the right routes can be assessed in a more isolated manner.

Because the model characteristics are evaluated in this section, per aggregation method a single level of aggregation is used:

- FZ-FL: neighbourhood level
- FZ-TS: neighbourhood level
- AZ-TS: neighbourhood+district level (highest aggregation: level 0)

First, the results of the evaluation framework are presented. It consists of the general indicators, link load indicators, travel time indicators and computational costs as discussed before. Furthermore, for the methods with two-sided routes, it shows how often a common node is found and how often a first/last mile path was used. The indicators of the evaluation framework for all three methods are shown in Table 7.

One iteration (AON)		Regular assignment	Fixed aggregated zones with first/last mile routes (FZ-FL)	Fixed aggregated zones with two-sided routes (FZ-TS)	Adaptive zoning with two-sided routes (AZ-TS)	Color legend	
General indicators							
Sum of vehicle kilometres		778951	4,5%	0,4%	1,9%		
Sum of vehicle hours		41407	14,7%	5,8%	4,4%		
Link load indicators							
Link load ratios	All roads (within 10%)		59%	89%	77%	100%	50%
	High order city roads (within 5%)		53%	99%	90%		
	Second-order city roads (within 10%)		45%	92%	75%		
	Local roads (within 20%)		64%	89%	83%		
Link load RMSE	All roads		53	5	16	0	50
	High order city roads		127	13	27	0	150
	Second-order city roads		69	7	25	0	100
	Local roads		17	2	10	0	20
Travel time indicators							
Travel time RMSE			1,08	0,29	2,1	0	2
Weighted travel time RMSE			12,26	3,48	10,44	0	10
Route reconstruction							
Common node				87,8%	83,1%		
First/last mile path				12,2%	16,9%		
Computational cost							
Dijkstra number of considered links		16.095.423	-66%	-62%	-70%		
Actual computational time (sec, %)		880	490	530	450	400	900
			-44%	-40%	-49%		

Table 7: Evaluation framework for one iteration

A first sight on the results shows that FZ-TS performs best in terms of accuracy and can reduce computational costs with around 40%, which is the least of the three methods. AZ-TS performs best in terms of computational cost with a computational cost almost half of the regular assignment method, and gives an accuracy that falls between the other two methods. The FZ-FL method shows the least accuracy of all methods and falls between the other methods in terms of computational cost.

6.3.1 Computational cost

The evaluation framework shows that adaptive zoning with two-sided routes performs best, followed by the fixed aggregated zoning with first/last mile routes and fixed zoning with two-sided routes.

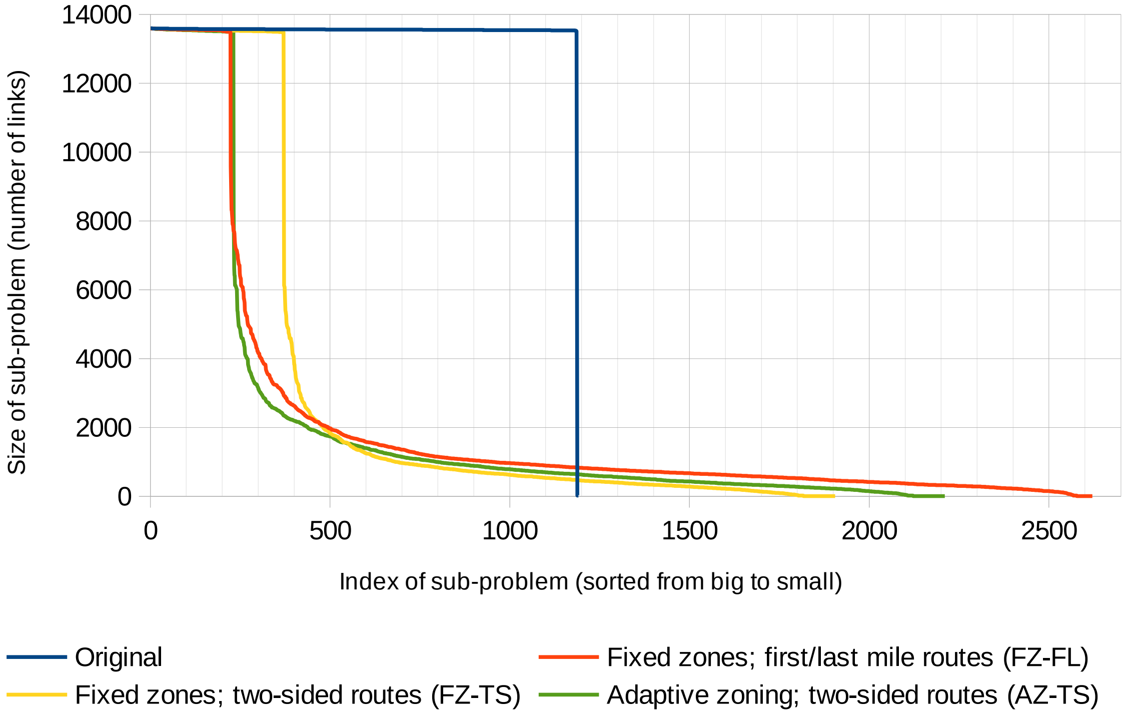


Figure 16: Sub-problems and their sizes for different aggregation methods (one iteration)

Figure 16 shows how the sizes of various Dijkstra's algorithms sub-problems are distributed for the different aggregation methods. It shows that the regular assignment method performs 1200 sub-problems (origin-based Dijkstra's algorithms), which is equal to the number of original zone centroids. Furthermore, the sizes of these sub-problems are constant (around 13.600 links), which means that every sub-problems considers the total road network.

For the method FZ-FL, the biggest sub-problems are of equal size to the sub-problems in the regular assignment. These sub-problems, around 200 in total, are origin-based algorithms from the aggregated zone centroids and outer centroids. The remaining sub-problems, 2400 in total, are getting smaller very fast and are represented by searching the first/last mile routes. For every original centroid (1200 in total), both an origin-based and a destination-based algorithm is performed, with targets being the other original centroids and the boundary nodes in that aggregated zone.

The method FZ-TS shows a higher number of sub-problems with an equal size to the sub-problems in the regular assignment: almost 400 sub-problems with the highest sub-problem size. This is almost twice the number of big sub-problems with the first/last mile routes. The same zone level is used, but for every zone two sub-problems are performed: one origin-based and one destination-based algorithm. The remaining sub-problems, around 1500, is represented by the route searching for the route reconstruction when no common node can be found.

The method AZ-TS shows that more than 200 sub-problems of the highest size are needed, which is represented by around 100 zones of the highest used level (district level) for which both an origin- and destination-based algorithm is performed (two-sided routing). The number of remaining sub-problems is around 2000 and consists of the lower level aggregated zones (neighbourhood level) performing two-sided routing and route searching for the route reconstruction when no common node is found.

This results in the following ranking for the computational cost: AZ-TS performs best with almost halving the calculation time, followed by FZ-FL with 44% lower calculation times, and concluded with FZ-TS with 40% lower calculation times. In chapter 5 we have seen that in theory for fixed aggregated zones, the first/last mile method could be up to twice as fast as the two-sided routes method; however in this network of limited size the difference is not that large.

6.3.2 Link load indicators

As the evaluation framework shows, the method FZ-TS gives the best link load accuracy, followed by AZ-TS and FZ-FL. This is true for all link load indicators.

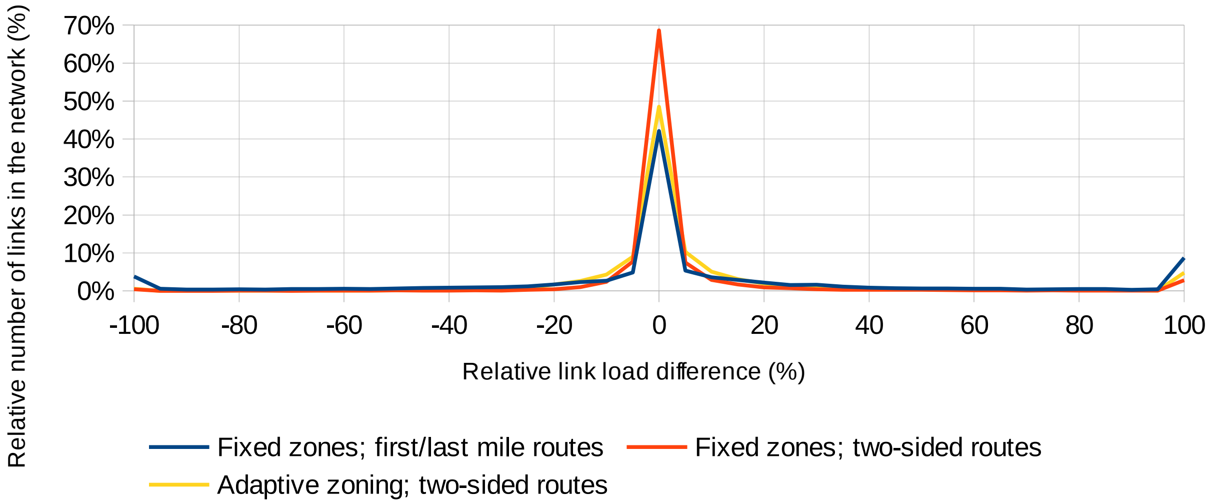


Figure 17: Link load comparison (one iteration)

Figure 17 shows that all methods produce link loads that are close to the regular assignment. The figure confirms the finding that fixed zones with two-sided routes produces the most accurate link loads. For one iteration however, still a significant number of links sees its load doubled or disappeared with all aggregation methods; for the method first/last mile routes this is around 13%.

Furthermore, for all methods is true that on the average, the link loads increase. This is explained by the route searching methods: with all aggregation methods, the routes that are found between OD pairs are at least as long as the shortest routes found in the regular assignment. Consider the first/last mile routes: especially on OD pairs close to each other (but in different aggregated zones), the found route depends on the shared route which is often out of the way of the direct route. The shared route often uses higher level roads and indeed it is found that the first/last mile routes method overestimates link loads especially on high level roads.

The evaluation framework shows that the method FZ-FL produces the worst routes between OD pairs; RMSEs are highest for this method. On the other hand, two-sided routes produce better results resulting in very low RMSEs, especially in fixed aggregated zones. The method fixed aggregated zones with two-sided routes could use higher numbers of common nodes than the adaptive zoning; which is expected as the two routes are closer to each other with on average smaller zones. This results in the most accurate link loads, with 90% of links within the accepted error.

6.3.3 Travel time indicators

Because routes are generally longer using aggregated methods, the number of vehicle kilometres and hours is higher than in the regular assignment. This finding is reflected in the travel times: on average, the travel time between any OD pair in an aggregated assignment is higher (Figure 18). Still, for all three methods the most common travel time difference is 0%. Fixed zones with first/last mile routes shows the highest increase in vehicle loads and thus in OD travel times. Adaptive zoning performs second-best and fixed zones with two-sided routes produces most accurate travel times.

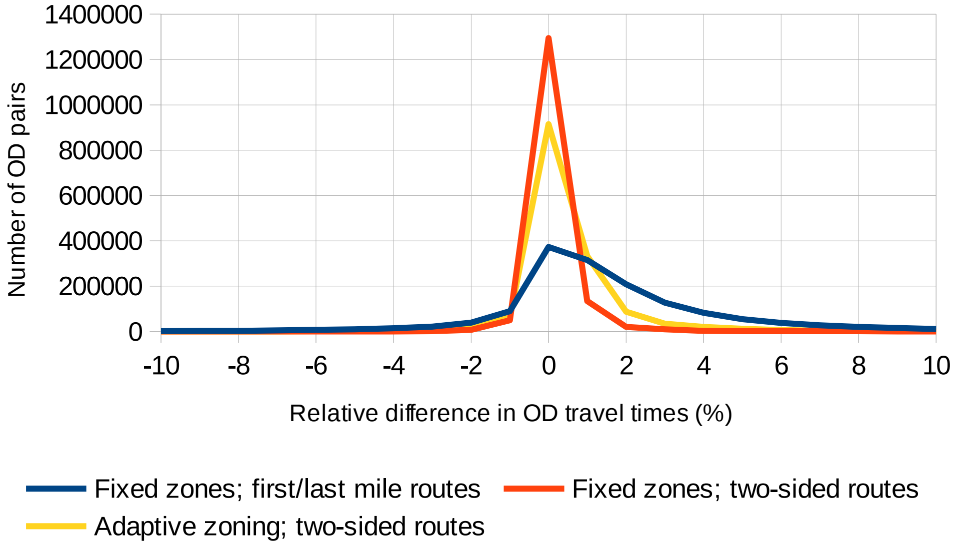


Figure 18: Travel time comparison (one iteration)

6.3.4 Conclusion

Based on this case study with one iterations, the first conclusions can be drawn. Although computational cost savings differ not much between the aggregation methods, travel times and link loads do. The lowest computational cost is reached with adaptive zoning for reasonable results, while most accurate results are produced with fixed zones and two-sided routes for still less calculation time than the regular model.

Fixed aggregated zones with first/last mile routes generate far less accurate link loads and travel times than fixed aggregated zones with two-sided routes, for only slightly better computational time. Therefore, this method is not not longer considered in the model application study with multiple iterations. This does not mean that it should not be used in any case: when the lowest calculation times are desired and only a single layer of aggregated zones is available, this method will still give reasonable results and can be used.

6.4 Model application

In this section, the aggregation alternatives are tested in a situation like in a real model application: namely with multiple iterations until an equilibrium in the transport network is reached. It is assumed that this equilibrium is reached when the *relative gap* is below 0.01% (Boyce, Ralevic-Dekic, & Bar-Gera, 2004).

The following alternatives and aggregation levels are evaluated:

Aggregation alternative	**Low aggregation**	**High aggregation**
Fixed zones; two-sided routes (FZ-TS)	Neighbourhood level (FZ-TS N)	District level (FZ-TS Q)
Adaptive zoning; two-sided routes (AZ-TS)	Level 1 (AZ-TS L1)	Level 0 (AZ-TS L0)

The indicators of the evaluation framework for the four combinations of methods and aggregation levels are shown in Table 5.

Iterative till equilibrium			Fixed aggregated zones with two-sided routes (FZ-TS)		Adaptive zoning with two-sided routes (AZ-TS)			
		Regular assignment	Neighbourhood level (FZ-TS N)	District level (FZ-TS Q)	Level 0 (AZ-TS L0)	Level 1 (AZ-TS L1)	Color legend	
General indicators								
Sum of vehicle kilometres		784522,2	0,5%	1,7%	2,0%	0,4%		
Sum of vehicle hours		1,79E+04	0,7%	2,3%	2,7%	0,7%		
Link load indicators								
Link load ratios	All roads (within 10%)		85%	77%	74%	86%	100%	50%
	High order city roads (within 5%)		97%	88%	84%	97%		
	Second-order city roads (within 10%)		89%	77%	77%	90%		
	Local roads (within 20%)		87%	83%	80%	88%		
Link load RMSE	All roads		7	13	14	6	0	50
	High order city roads		16	25	27	17	0	150
	Second-order city roads		8	18	18	8	0	100
	Local roads		3	10	11	2	0	20
Travel time indicators								
Travel time RMSE			0,07	0,11	0,12	0,07	0	2
Weighted travel time RMSE			0,23	0,39	0,43	0,23	0	10
Route reconstruction								
Common node			90,3%	82,8%	82,6%	91,4%		
First/last mile path			9,7%	17,2%	17,4%	8,6%		
Computational cost								
Actual computational time (sec, %)		7857	5551	4299	4260	6916	4000	8000
			-29%	-45%	-46%	-12%		
Number of iterations		10	8	10	11	8		

Table 8: Evaluation framework for equilibrium

The indicators in the evaluation framework show that the highest aggregation levels have the lowest computational cost but also the highest inaccuracies.

6.4.1 Computational cost

The method with the lowest computational cost is AZ-TS L0; it has the lowest calculation time per iteration but needs one more iteration than FZ-TS Q, which makes the two methods almost equal fast. The method FZ-TS N performs in the middle range and needs only 8 iterations. The method AZ-TS L1 is the slowest of all: calculation time per iteration is even higher than with the regular assignment. However, it needs less iterations, so is still slightly faster. The reason for high computational times for this method is the high number of sub-problems (every neighbourhood has two sub-problems) in combination with the big size of the sub-problems (the sub-problem size of the original zones covers whole its own district).

6.4.2 Link load indicators

The evaluation framework shows that FZ-TS N and AZ-TS L1 have the highest link load accuracy.

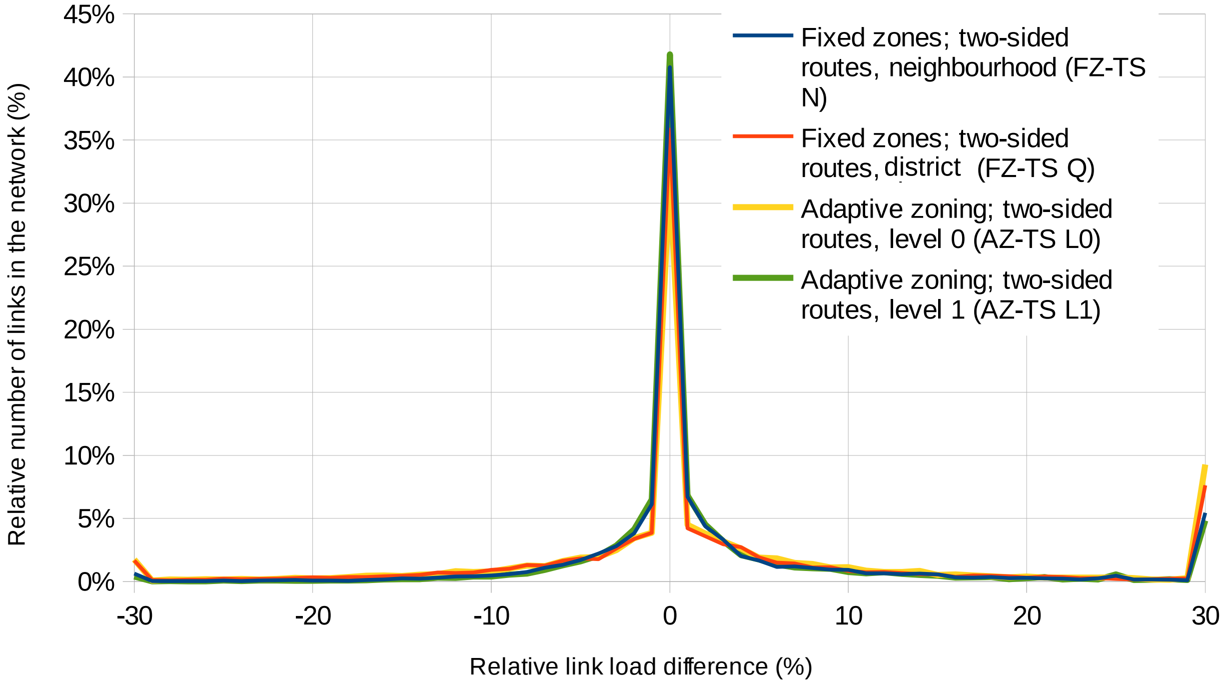


Figure 19: Link load comparison (equilibrium)

This is confirmed with Figure 19, which shows that both have the highest peak in link accuracy and the lowest spread in the frequencies. Both lead to only 6% of links that have a relative link load difference of 30% or more and more than 40% of links with exactly the right load. The links that have a relative difference of more than 30% are almost all very low level links with a small load in the original assignment; an over- or underestimation of loads results in high relative differences. More than 85% of links fall within the acceptable limits for the different levels of roads. Also RMSEs are lowest for these methods.
The difference between these methods is that AZ-TS L1 finds routes between zones in the same district but different neighbourhood directly, while FZ-TS N does this with two-sided routing. This difference leads to slightly better link loads for FZ-TS L1.

The other two methods, FZ-TS Q and AZ-TS L0, perform also reasonable well but have a higher spread in the frequency figure. They lead to higher RMSEs and lower numbers of links that fall within the reasonable limits (more than 74%).
The difference between these methods is that FZ-TS Q finds the route between zones in the same district but different neighbourhood directly, while AZ-TS L0 uses two-sided routing for these zones. This is the reason FZ-TS Q leads to slightly more accurate link loads.

6.4.3 Travel time indicators

An assignment till equilibrium shows an increase in vehicle kilometres and hours for all aggregation methods, although lower than for one iteration. This leads to higher OD travel times on average, which is illustrated by Figure 20. AZ-TS L1 produces the most accurate travel times shortly followed by FZ-TS N. FZ-TS Q and AZ-TS L0 also produce quite accurate travel times, although slightly less. Still, 93% of all OD pairs have a travel time error of less than 2%.

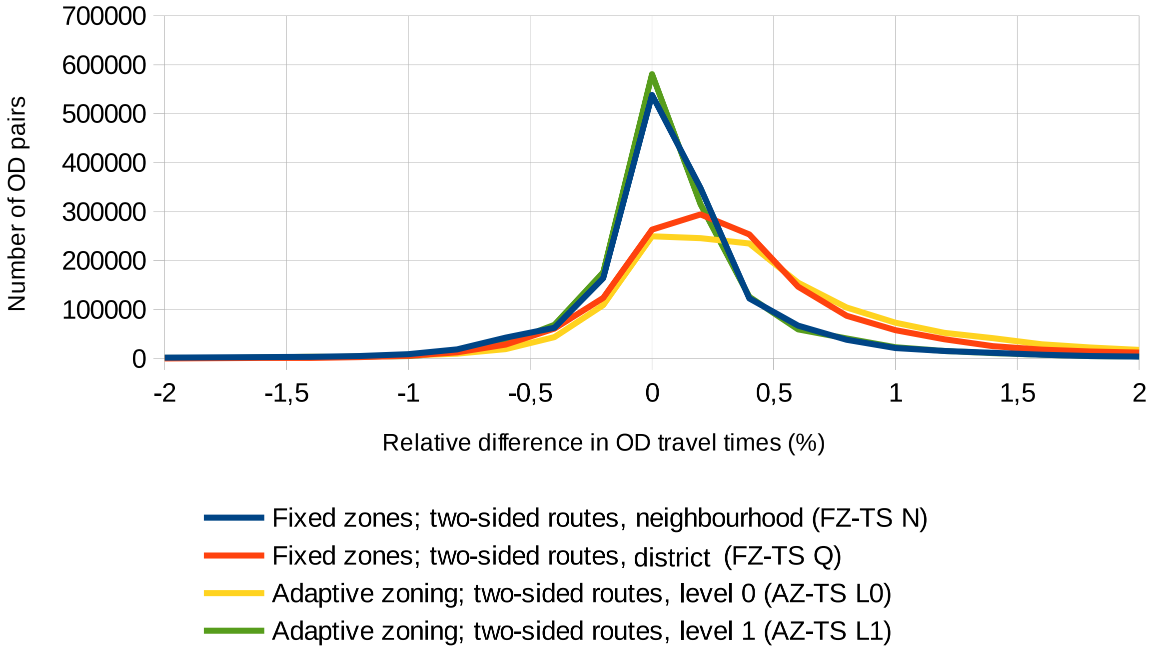


Figure 20: Travel time comparison (equilibrium)

6.4.4 Conclusion

This comparison of methods and aggregation levels shows that there is a clear trade-off between computational cost and accuracy. For this network, the difference in computational cost and accuracy differs much between level of aggregation but less between aggregation method. There are some differences in accuracy between the aggregation methods for the same aggregation level, but those are minimal.

For the low level of aggregation (FZ-TS N and AZ-TS L1) the difference is how routes between zones in the same district (but different neighbourhood) are determined. FZ-TS N does this more efficient and has significantly lower calculation times. Thus, when higher accuracies are desired, FZ-TS Q can be considered, because it gives almost as good accuracy as AZ-TS L1 but at much lower calculation time.

For the higher levels of aggregation (FZ-TS Q and AZ-TS L0), the same difference holds, but in this case the adaptive zoning does two-sided routing in more cases than fixed zoning, which is more efficient in computational costs. Thus for the lowest computational costs, AZ-TS L0 can be considered; it performs faster than FZ-TS Q, especially per iteration.

When more zone levels would have been available, it is possible to use even higher levels of aggregation. It is expected that adaptive zoning generally reduces the computational cost for slightly lower accuracy than the regular model. As we have seen, only for the lowest level of aggregation, adaptive zoning increases complexity and is generally not the optimal choice. For all other levels of aggregation, it is expected to generate almost as accurate results as fixed zoning for less computational cost.

The boundary conditions have been met with all described methods:

- The existing low-level zone structures and OD matrices can be used, which makes switching from the regular model to the aggregated model quite easy.
- Total routes between OD pairs are reconstructed, which means that the flow conservation criterion has been met.
- The number of required additional data is limited to the aggregated zones that should be imported in the model.
- The aggregated model is flexible: different levels of aggregation can be applied, depending on the input of the aggregated zones.

The following recommendations are done for specific applications of a transport model.

Recommendation for a quick view of the traffic situation

When a quick view on the total traffic situation in a larger area is needed, adaptive zoning with two-sided routes is often the preferred way of aggregation. There are some differences in traffic flows, but these are small, especially if the network as a whole is considered. Computational costs can decrease with 45% in this case.

Recommendation for performing variant studies

For performing variant studies where differences in traffic flows can be marginal, more accuracy is needed. This can be achieved by using fixed zoning with two-sided routes on neighbourhood level (low level of aggregation). A computational cost gain of 30% can be reached and link load errors are small, as shown in appendix C. This appendix shows link load differences on the total network, both including and excluding shared loads.

Recommendation for estimating travel times

For estimating travel times and skim matrices, all investigated methods perform well: the differences in skim matrices produced with aggregation methods and the regular assignment are very small. To profit the most from aggregation and computational costs savings, it is recommended to use a high level of aggregation in this case. The revealed maximum of 2% error in travel time of a particular OD pair is often marginal compared to the travel times of other OD pairs in the network (as used in trip distribution) or compared to other modalities for the same OD pair (as used in modal split). Therefore, it is expected that the effects on trip distribution and modal split are also marginal.

7 Conclusion and discussion

This chapter concludes the report by answering the research questions. Furthermore, it discusses assumptions behind the research and makes recommendations for future research.

7.1 Conclusion

The objective of this research is to develop a static transport assignment model by using smart forms of aggregation to reduce computational costs and maintain accuracy as much as possible. This research explored and tested different aggregation methods. This section answers the sub research questions which in turn lead to the answer of the main research question.

What types of aggregation are known in transport modelling literature?
Transport models always use some sort of aggregation. Traffic demand data is aggregated in small geographical zones and often not all roads are included in the network. However, due to limitations in the computational capacity or model questions on a higher geographical scale, further aggregation is often applied. In transport modelling, the two main aggregation principles are zone and network aggregation. The most basic form of zone aggregation is merging some original zones into one aggregated zone and aggregating the OD matrix accordingly. More sophisticated aggregation alternatives use aggregated zones in which a shared route is calculated in order to reduce computational costs, or adaptive zoning where original zones interact all with a different set of aggregated zones. Network aggregation is typically performed in two ways: link extraction and abstraction. In the first case, links that are not used are removed from the network. In the second case, links are replaced by abstract links based on the function of the link. In any case, the detail of zones and links in the network should match in order to prevent links with largely under- or overestimated link flows.

Which principles explain the computational cost efficiency of aggregation?
Dijkstra's shortest path algorithm is the most used algorithm to find shortest routes in transport networks, and an understanding of the computational properties of this algorithm is important for implementing aggregation. It calculates the shortest routes from a single source to all possible targets in a network, which is one sub-problem. The larger the network (in terms of number of links or nodes), the longer the algorithm takes. Furthermore, the more sub-problems are performed, the longer one iteration in the assignment step takes. This is represented by the expression of computational complexity: one iteration of a transport assignment performed using any method requires $O(iteration) = O(j\, l \log l)$ where j is the number of sub-problems and l is the size of the sub-problem. The basic form of zone aggregation leads to a reduction of sub-problems, while network aggregation reduces the size of sub-problems.

Which main aggregation alternatives are feasible in the context of traffic assignment?
The chosen aggregation alternatives in this research should be able to fulfil the objectives of the research: to model traffic flows on all roads and using the original OD matrix. Aggregation alternatives are split up in building blocks, so the effect of different aggregation elements can be measured the best. The combination of these building blocks leads to three different zone aggregation alternatives: fixed aggregated zoning with first/last mile routes, fixed aggregated zoning with two-sided routes and adaptive zoning with two-sided routes.

To which extent does the level of aggregation influence the accuracy and computational costs?
The methods with two-sided routes were tested with different levels of aggregation: on neighbourhood and district level. On district level, the average link load and the number of vehicle kilometres increased in comparison with the neighbourhood level. This can be explained by the longer routes between OD pairs that are found on average with higher aggregation levels. It is noted that, for different aggregation levels, the differences in link loads are higher than the differences in travel times between OD pairs. This is caused by

link load differences on a route between two zones that balance out. Thus, a higher aggregation level mainly influences the link loads: the RMSE doubles from 7 to 14 vehicles when going from neighbourhood to district level. More aggregation influences with smaller degree the travel times RMSE: from around 0.07 to 0.12 when aggregating. Furthermore, higher aggregation levels reduce the computational cost significantly.

Which aggregation methods are best suited for generating skim matrices?
Based on the analysis performed on the network in this research, the differences in skim matrices produced with aggregation methods and the regular assignment are very small. Of course, a higher degree of aggregation increases differences in skim matrices, but they stay very small. Therefore, it is assumed that these slightly different OD matrices are still very well able to produce input for new model steps: distribution and modal split. To have maximum computational gains from the aggregation methods, it is recommended to use a high level of aggregation when skim matrices are to be produced. For this highest level, adaptive zoning is recommended as it has the lowest computational cost. The case study showed that adaptive zoning with two-sided routes on the highest aggregation level results in 97% of all OD pairs with a relative travel time error of at most 2%.

Which aggregation methods are best suited for accurately estimating traffic flows?
As has been noted, on higher levels of aggregation, adaptive zoning decreases the computational costs significantly in comparison with fixed zoning, for just slightly lower accuracy. On by far the most links in the network, no difference in load is visible with the eye compared with the original assignment, even with high aggregation. Therefore, adaptive zoning with two-sided routes is often the preferred way of aggregation, especially when the network is big and considered in its totality. For adaptive zoning on the high level of aggregation, 74% of all links have a relative link load difference of at most 10% for a reduction of 45% in computational costs.
In cases when variant studies are done, fixed zoning with two-sided routes should be used. Additional advantages are that this method is probably easier to understand. This means that it could be easier accepted by policy makers. Furthermore, it is the best option if just one level of aggregated zones is available. Additionally, on the lowest level of aggregation, fixed zoning is even faster than adaptive zoning and in that case this is the preferred option. Thus, when small, detailed parts of the network are considered as in variant studies, fixed zoning with two-sided routes on the lower level of aggregation is often the preferred way. Fixed zoning with two-sided routes on the low level of aggregation results in 85% of all links within the acceptable link load difference of 10%, for a computational cost gain of almost 30%.

How and to what extent can the balance between accuracy and computational costs of the transport modelling network assignment be improved, by using aggregation?
This research has shown that the accuracy of assignment results and computational costs are not static for a given transport model. Instead, when lower computational costs are desired, generating assignment results and skim matrices is still possible with good results. Three aggregation alternatives have been developed, which all satisfy the boundary conditions of integration with current zone systems and having consistent assignment results. All aggregation alternatives can also be used with different levels of aggregation, making the balance between accuracy and computational cost really flexible.
Of course, it is the modeller's decision whether to accept slightly lower accuracy in return for much lower computational costs. This will depend on the modeller's value of accuracy in comparison with computational time. Thus, the extent to which the balance is improved with these aggregation methods, varies between situations. However, it is expected that in many cases, this ratio is in favour for lower computational costs.
As noted, the preferred level of aggregation depends on the application and requirements. For higher levels of aggregation, adaptive zoning is often favourable, while for the lowest level of aggregation, fixed zoning is regarded more optimal.

7.2 Discussion

During the research, choices and assumptions have been made, which may have impact on the results or conclusions of the research. This section describes them and estimates the possible impact.

This research focussed mainly on zone aggregation; network aggregation was not part of any aggregation alternative. Analysis of the literature showed that the network level should match the level of zones used in the network. However, in this research the original zone structure remains untouched and, as in the regular assignment, traffic from and to the original zones is still modelled. This means that traffic remains modelled on all link levels. Therefore it is suggested that for these more advanced methods of zone aggregation, network aggregation is not strictly needed.

The case study The Hague in which the different aggregation methods are tested, is “cut” at the boundaries of the area of interest. In all of the aggregation alternatives, the zone centroids that represent external traffic are not aggregated. Because the external traffic makes up a significant share of the total traffic in the transport network, this can be seen as an “unfair” way to improve accuracy of the aggregated methods. On the other hand, the aggregated models still lead to lower computational costs.

Another characteristic of the case study is that the aggregation difference between neighbourhood and district zones is much smaller than the difference between districts and the total study area. The consequence is that the accuracy of different aggregation methods for the same aggregation level is similar, while different aggregation levels really show different results. When more zone levels were available and used in the research, the differences between methods could have been clearer.

Dijkstra’s algorithm, which is the shortest path algorithm used in this research, normally continues until the shortest paths to all nodes in the network are known. However, this research adapted the algorithm to stop after shortest paths to all targets have been determined. This is an additional computational step that must be performed, which may slow down the algorithm.

Lastly it is noted that the indicated calculation time savings should mainly serve as an approximation of the savings that can be done on other transport networks, software environments or computing devices. Although effort has been done to re-use the same algorithm scripts for different alternatives when possible, calculation times also depend on the method of implementation. Nonetheless, the combination between calculation time, visited links in the shortest path algorithm and complexity notation gives a good indication of computational costs.

7.3 Future research

This section presents recommendations for future research on the topic of aggregation in transport networks. It starts with suggestions for further development of the discussed aggregation methods, which are expected to further increase accuracy or deduce computational costs. Furthermore, it suggests how the developed aggregation methods can be further assessed using different case studies.

7.3.1 Aggregation methods

Because the two-sided routes approach performs best compared to the other route reconstruction method (first/last and shared routes), it is reasonable to focus on further development of that method. In the case study, it was shown that alternatives with lower accuracy (on the same aggregation level) had a relative high number of OD pairs where no common node was found. Apparently, reconstructing routes with a first or last mile in two-sided routing is generally less accurate. Therefore, future research can focus on improving that method, for example by developing an improved intermediate route that connects both routes. It can also improve the first or last mile route in two-sided routing, for example by using the boundary nodes like in the first/last mile and shared routes method.

A similar direction of research is the idea of to skip generating a full route between origin and destination when the traffic demand between that OD pair is very low, for example lower than 1 vehicle. This can be especially effective when a common node cannot be found; this way computational costs can be reduced further.

To improve the computational costs of adaptive zoning, the number of aggregated zone layers can be further increased. This can be done with an algorithm that splits up existing aggregated zones, which forms a lower level of aggregated zones. This reduces the scale difference between consecutive layers of aggregated zones. This makes adaptive zoning more effective and flexible, and the definitions of aggregated zones are still equal to or based on existing zone structures.

7.3.2 Application

The application of the developed aggregated assignment models in other case studies gives more insight in their performance. First of all, it would be interesting to apply the aggregated model to a full transport model that is not cut at the area of interest boundaries. It is suggested that zones outside the area of interest are still aggregated when their density is high. When the density of zones outside the study area is low, the effects of aggregation on computational costs are limited. In the case study, this would mean that the zones in the province of South Holland are still aggregated like in the study area, but outside this province where the density of zones is low, aggregation is not applied.

Finally, in order to better estimate the desired balance between accuracy and computational costs, it would be interesting to compare both the regular and the aggregated models to actual traffic counts. Only in this way, the relative additional error as a result of aggregation can be assessed. With this information, the modeller can better decide whether to accept the additional error for the shown reduction in computational costs. In order to measure the impact of the aggregation on trip distribution and modal split, those steps should also be performed and the results assessed.

Appendix A: Example of adaptive zoning

To clarify the way adaptive zoning works with two-sided route finding, the example in this appendix is been set up. The idea of adaptive zoning comes from Hagen-Zanker & Jin (2015). This research extends that approach by enabling existing zone structures, and being able to reconstruct a complete route between origin and destination (not merely two paths that not necessarily form one path).

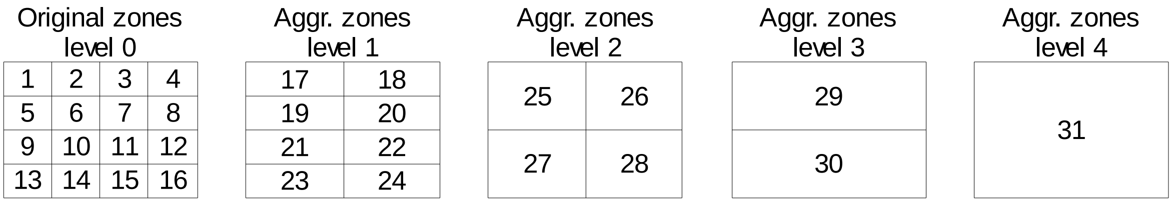


Figure 21: Original and aggregated zones

The example consists of a zone structure of 4x4, which is a total of 16 original zone centroids. In the first level of aggregation, two zones are merged and form a new one. The resulting aggregating zones form level 1, and are eight zones. In level 2, there are four aggregated zones; in level 3, two aggregated zones and level 4, one single aggregated zone (this last level is not used).

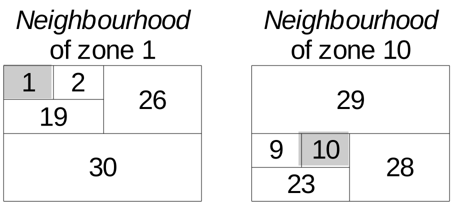


Figure 22: Example of neighbourhood maps

Then, for all original zones, the *neighbourhood* is defined. Figure 22 shows two examples. For all original zones, the *neighbourhood* is shown in Table 9. The idea of the neighbourhood is that close by zones are small and zones far away are big.

Finally, the *inverse neighbourhood* is created based on the neighbourhood table. This is done by, for every (aggregated) zone, considering which original zones interact with it. It can be automatically generated from the neighbourhood table.

Neighbourhood

1	2	19	26	30
2	1	19	26	30
3	4	20	25	30
4	3	20	25	30
5	6	17	26	30
6	5	17	26	30
7	8	18	25	30
8	7	18	25	30
9	10	23	28	29
10	9	23	28	29
11	12	24	27	29
12	11	24	27	29
13	14	21	28	29
14	13	21	28	29
15	16	22	27	29
16	15	22	27	29

Inverse neighbourhood

1	2							
2	1							
3	4							
4	3							
5	6							
6	5							
7	8							
8	7							
9	10							
10	9							
11	12							
12	11							
13	14							
14	13							
15	16							
16	17							
17	5	6						
18	7	8						
19	1	2						
20	3	4						
21	13	14						
22	15	16						
23	9	10						
24	11	12						
25	3	4	7	8				
26	1	2	5	6				
27	11	12	15	16				
28	9	10	13	14				
29	9	10	11	12	13	14	15	16
30	1	2	3	4	5	6	7	8

Table 9: Neighbourhood and inverse neighbourhood

The computational cost savings from adaptive zoning comes from the usage of the *inverse neighbourhoods* as sub-problems. Thus the number of sub-problems is three times the number of sub-problems in the regular assignment, but their sizes are on average smaller.

The following example clarifies the two-sided route finding. The total route between original zones 1 and 13 is to be found. Aggregated zone 29 has zone 13 in its inverse neighbourhood and has found routes to and from it. Aggregated zone 30 has zone 1 in its inverse neighbourhood and also found routes to and from it. To reconstruct the full path, the route from zone 1 to 30 is used on the detailed side (on the upper side), and the route from 29 to 13 is also used on the detailed side (at the lower side). How this works is described in paragraph 4.1.2 on page 32.

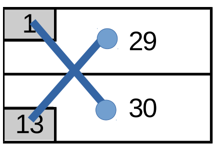


Figure 23: Two-sided routes in the context of adaptive zoning

Appendix B: Building block descriptions

This appendix describes the building blocks that were considered in the research. However, they were not part of any aggregation alternative in order to reduce the number of possible combinations to evaluate.

The zone aggregation procedure and assignment can consist of different building blocks. In case one wants to utilize zone aggregation, the fixed building blocks are necessary to use. Conditional required building blocks are required when some option in another building block is chosen.

Fixed building blocks

Building block	Options	
Aggregated zone hierarchy	Fixed aggregated zone definition	Adaptive zoning: every original zone interacts with different aggregated zones
Centre representation of aggregated zone	New node (centroid)	Use existing node
Assignment procedure	Reconstruct route for every OD-pair; assign original OD-matrix	Create routes between aggregated zones; assign aggregated OD-matrix

Conditional required building blocks

Requirement if	Building block	Options		
New node (centroid)	Aggregated zone centroid placement	Centre of sub-centroids	Weighted centre of sub-centroids	
Use existing node	Aggregated zone centre node determination	Existing node closest to centre of sub-centroids	Existing node closest to weighted centre of sub-centroids	
New node (centroid)	Number of aggregated zone centroid connectors	One	Multiple	
Multiple	Aggregate zone centroid connector node	Closest node(s)	All boundary nodes	All sub-centroid connector nodes
Multiple	Aggregated zone connector (generalized) cost	Zero	Ratio to length	
Reconstruct route for every OD-pair; assign original OD-matrix	Route reconstruction	Reconstruct total route with first/last mile routes	Reconstruct total route with two-sided routes	
Create routes between aggregated zones; assign aggregated OD-matrix	Actions per iteration	Only macro-assignment (first/last mile assignment after all iterations, as post-processing)	Both macro-assignment and first/last mile assignment	

Figure 24: All aggregation building blocks

Centre representation of aggregated zone

Aggregated zones, like the original smallest zones, require a centre representation where trips to/from that zone start and end. This centre can be either:

- An existing node;
- A new node/centroid.

An existing node is easy because it does not have to be connected to the network. Its behaviour is similar to a new node that is connected using one connector to the closest node. The other option is to create a new node/centroid, which still has to be connected to the network. In this case, a few more building blocks are required (Figure 25).

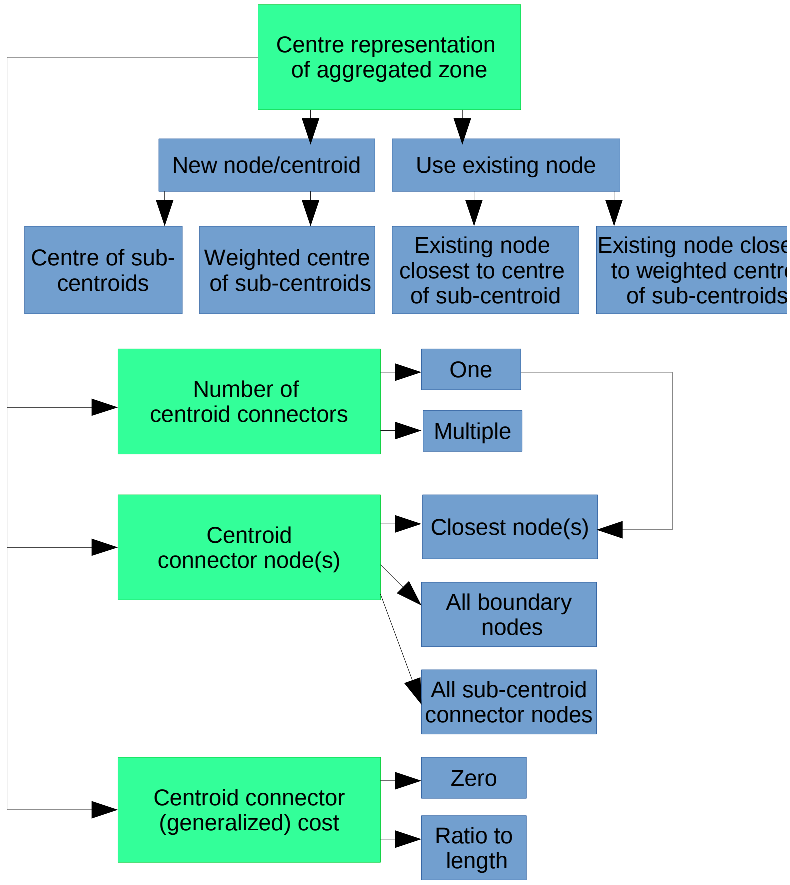


Figure 25: Options for choosing the centre representation

In the research, an existing node is chosen for each aggregated zone centroid.

Assignment procedure

This building block defines the way of assignment. Two options are possible:

- The original OD matrix is aggregated according to the aggregated zone definition and this matrix is assigned
- A route between every sub-centroid pair is reconstructed (in a smart, fast way) so that the original OD matrix can be assigned.

In the first case, the original OD-matrix for car traffic is compressed so that the result is an OD-matrix for the aggregated zones. The following formula is is used for every cell in the aggregated OD-matrix to calculate the traffic between aggregated zone centroids (the two aggregated zone centroids are indicated as m_a and m_b):

$$T_{m_a \rightarrow m_b} = \sum_i \sum_j T_{q_i \rightarrow q_j} \forall i \in m_a, j \in m_b$$

Consider for example the following original OD-matrix for a fictive network with four zones:

i \ j	1	2	3	4	
1					$\sum T_{1\to j}$
2					$\sum T_{2\to j}$
3					$\sum T_{3\to j}$
4					$\sum T_{4\to j}$
	$\sum T_{i\to 1}$	$\sum T_{i\to 2}$	$\sum T_{i\to 3}$	$\sum T_{i\to 4}$	

Table 10: Example original OD-matrix for car traffic

The resulting aggregated OD-matrix looks like this:

i \ j	5	6	
5			$\sum T_{5\to j}$
6			$\sum T_{6\to j}$
	$\sum T_{i\to 5}$	$\sum T_{i\to 6}$	

Table 11: Example aggregated OD-matrix

With the aggregated OD matrix, it is optional to detail the assignment again using the block “First/last mile assignment” (however, not using this block results in a regular assignment on a higher level, less detailed).

In the second case, the original OD matrix is not changed; rather the original matrix is assigned. Therefore, it is required that a route for every OD pair is found using the block “Route reconstruction”. In the research, only the last option is used, because it was shown to give more accurate results: as the original OD matrix can be used; it is possible to reconstruct the routes quite accurately, depending on the details.

Aggregated zone centroid placement

An aggregated zone centroid represents the different original centroids in a particular aggregated zone. The goal is to place the aggregated zone centroid in such location, that the distribution of routes between aggregated zones is equal as in the detailed “regular” assignment.

- Simply the centre of the corresponding sub-centroids: The location of the aggregated zone centroid could be defined as the geometric centre of all sub-centroids in the corresponding macro-zone *m*.
- The weighted centre of the sub-centroids.

In the first option, the formula for the x-value of the sub-centroid is as follows:

$$x_m = \frac{1}{n_q} * \sum_q x_q \quad q \in m$$

where n_q is the amount of sub-centroids within the macro-zone and x_q is the x-coordinate of the sub-centroid in the macro-zone.

In the second option, the location of the macro-centroid is the geometric centre of all sub-centroids in the corresponding macro-zone *m*, weighted according to the amount of traffic going in and out of every sub-centroid:

$$x_m = \frac{1}{\sum_q \left(\sum_i T_{q\to i} + \sum_j T_{i\to q} \right)} * \sum_q \left(x_q * \left(\sum_i T_{q\to i} + \sum_j T_{i\to q} \right) \right) \quad q \in m$$

The latter gives slightly more accurate results for the same computational cost, as the centre point moves towards the highest traffic amounts. Therefore, that approach is used in this research.

Aggregated zone centroid node determination

This block is required if an existing node is used as aggregated zone centroid. In this block, the same centre is calculated as in “Aggregated zone centroid placement”. Then the existing node that is closest to this centre, will be appointed as the zone centroid. Requirements for this node are:

- The node is located within the boundaries of the aggregated zone. This requirement is used to increase the chance of generating logical paths from the aggregated zone to the original zone centroids;
- The node is not connected to any high order link, for example motorways. This is required because nodes on motorways are difficult to logically connect to the original zone centroids.

Number of aggregated zone centroid connectors/node/(generalized) cost

These building blocks are required if it is chosen to put a new centroid in the aggregated zones and optional when an existing node is used as aggregated zone centroid. The combination of three factors, namely the number of connectors for the aggregated zone centroid, their connector node and their (generalized) cost, together determine the location where the traffic is loaded onto the network, and thus also which routes are taken for the (macro-)assignment. Choosing one (closest) node to connect to, results in traffic being assigned to a single point and tends to lead to a limited set of routes (in-links) to that aggregated zone.

In this research, existing nodes are used as aggregated zone centroids and they are not manually extra connected to the network.

Actions per iteration

This building block is required if it is chosen to use the optional block “First/last mile assignment”. In every iteration, it can be chosen to do:

- Only the macro-assignment. In this case, the first/last mile assignment is performed once after all the iterations.
- Both macro-assignment and the first/last mile assignment.

The first gives lower computational costs but also assumes that the macro-traffic is not influenced by the first/last mile traffic. With big aggregated zones (and thus many internal and first/last mile traffic), this assumption may not hold. For bigger networks, the computational costs of first/last mile assignment could be neglected, and that argues in favour of doing both macro-assignment and first/last mile assignment per iteration.

Chosen in the research is not to use the the separate first/last mile assignment; therefore, this block is not used.

Appendix C: Network plots

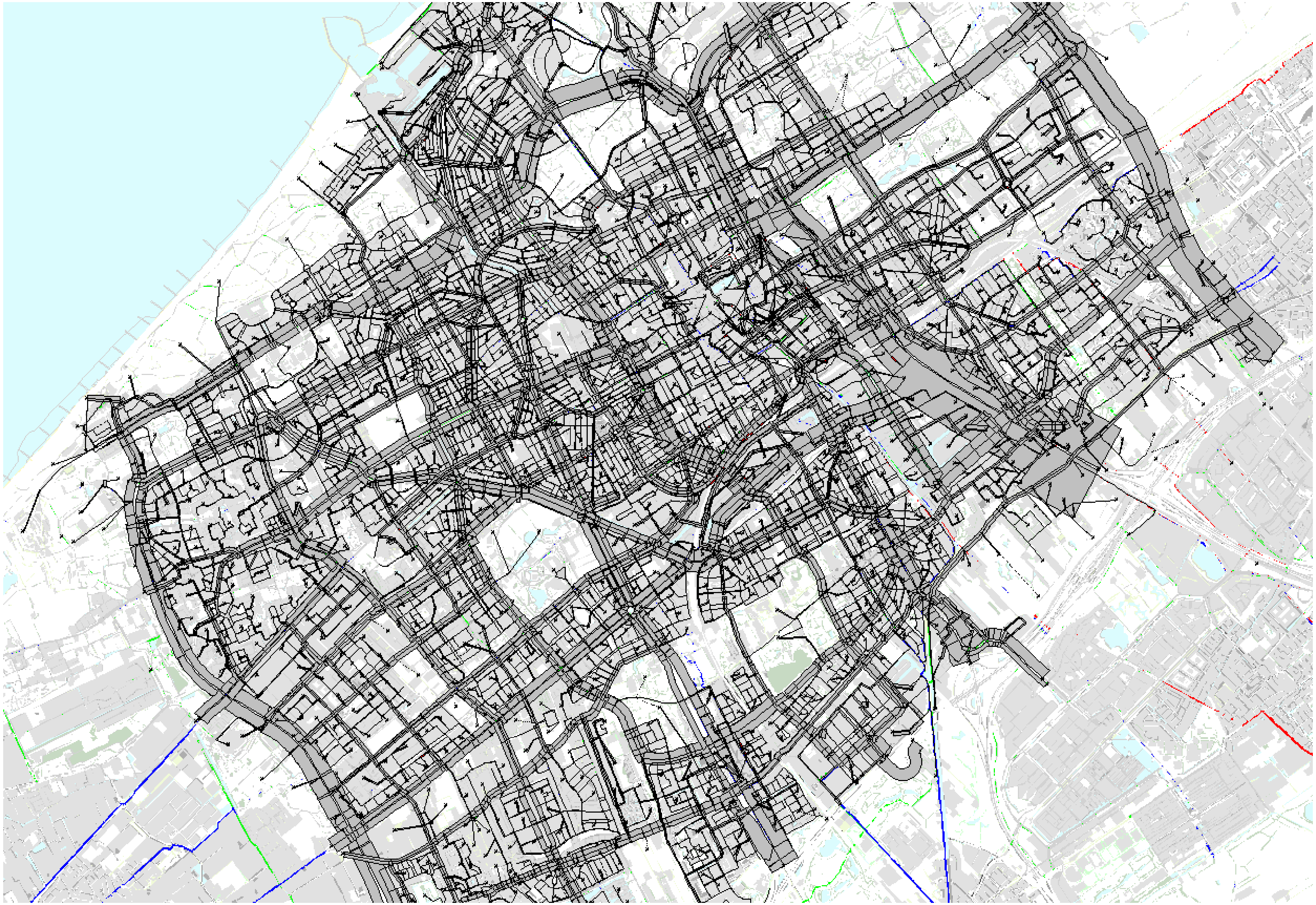


Figure 26: Link load differences for FZ-TS N in the case study (gray = no difference; red = link load overestimation; green = link load underestimation)

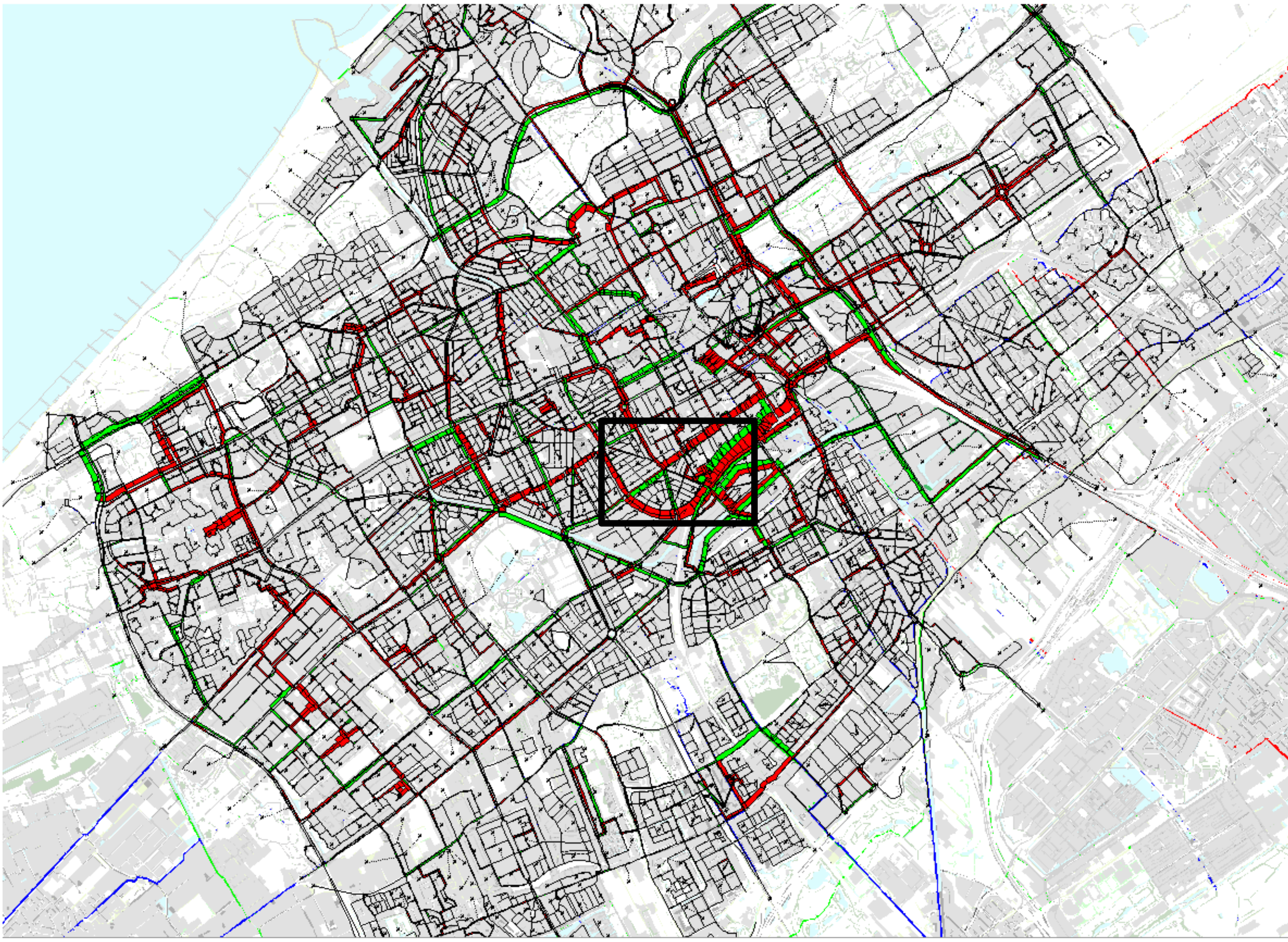


Figure 27: Link load differences excluding shared loads for FZ-TS N

(red = link load overestimation; green = link load underestimation)

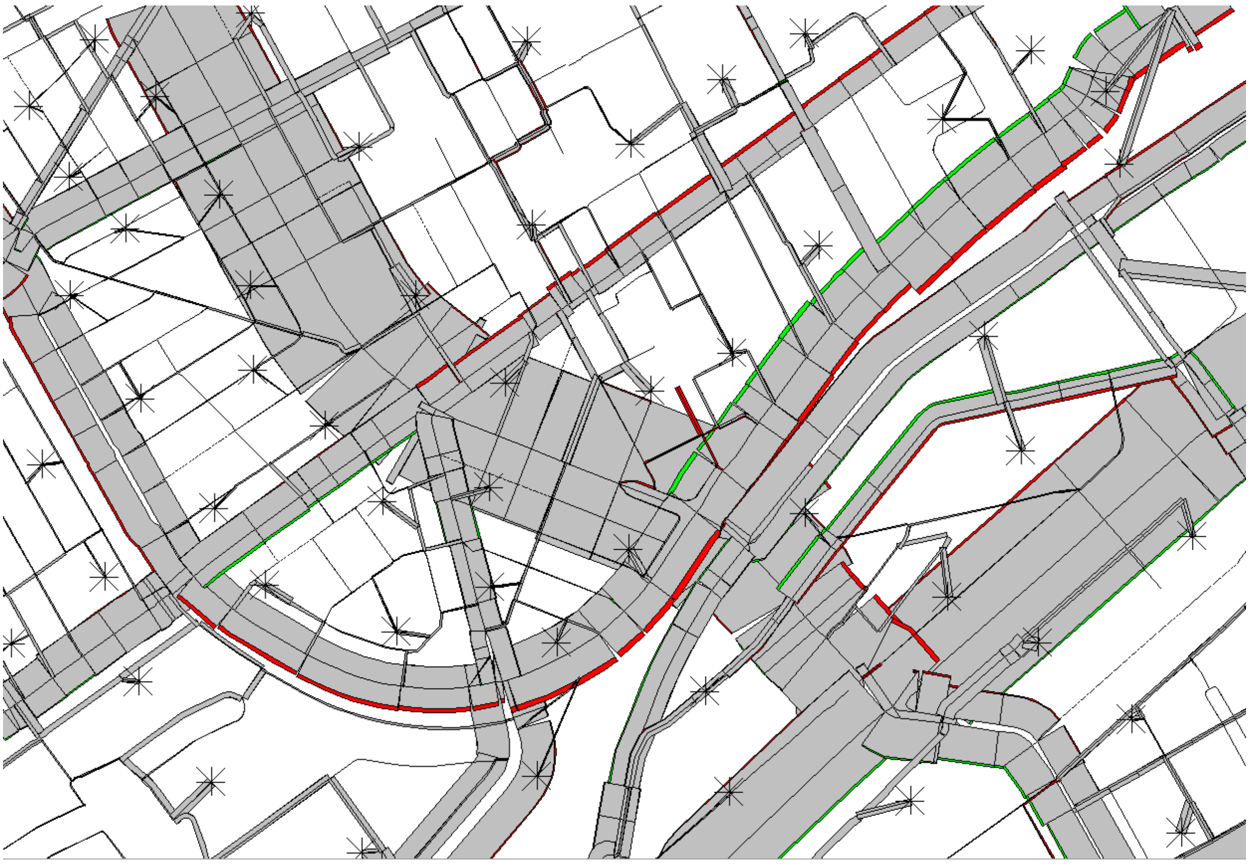


Figure 28: Map detail that shows a fraction of traffic taking another route

Appendix D: Relation number of centroids and links

The number of centroids and links in a particular transport network area is correlated. The following table and figure show the number of measured centroids and links for a number of aggregated zones in the The Hague transport model, used in the research.

Zone	Centroids	Links	Links per centroid
#1	4	98	25
#2	7	244	35
#3	7	120	17
#4	9	274	30
#5	9	210	23
#6	9	308	34
#7	10	290	29
#8	11	264	24
#9	13	396	30
#10	13	430	33
#11	14	334	24
#12	15	342	23
#13	16	344	22
#14	17	500	29

Table 12: Number of centroids and links per aggregated zone in The Hague

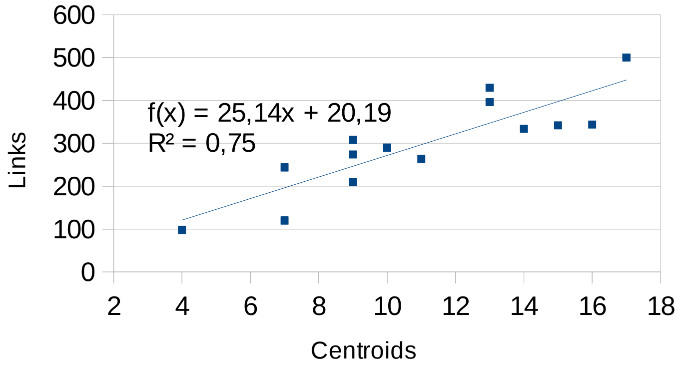


Figure 29: Number of centroids and links per aggregated zone in The Hague

For these zones, the average number of links per centroid is 27, with a minimum of 17 and a maximum of 34. A correlation is determined between the number of centroids and links in an aggregated zone: 75% of the variance in number of links is explained by the number of centroids.

References

Annema, J. A., & Jong, M. de. (2012). De Discussie over Transportmodellen, 1(48).

Baass, K. G. (1980). Design of zonal systems for aggregate transportation planning models. Transportation Research Record, 807, 1–6.

Benezech, V. (2011). A new model for disaggregate traffic assignment making explicit the spatial distribution of trip extremities. Presented at the European Transport Conference 2011. Retrieved from https://trid.trb.org/view.aspx?id=1237890

Bliemer, M., Raadsen, M., Romph, E. de, & Smits, E. (2013). Requirements for traffic assignment models for strategic transport planning: a critical assessment. Presented at the Paper presented at: Proceedings of the 36th Australasian Transport Research Forum 2013, ATRF, Brisbane, Australia, 2-4 October, 2013, Australasian Transport Research Forum.

Bovy, P. H. L., & Jansen, G. R. M. (1983). Network Aggregation Effects upon Equilibrium Assignment Outcomes: An Empirical Investigation. Transportation Science, 17(3), 240–262. https://doi.org/10.1287/trsc.17.3.240

Boyce, D., Ralevic-Dekic, B., & Bar-Gera, H. (2004). Convergence of traffic assignments: how much is enough? Journal of Transportation Engineering, 130(1). Retrieved from https://trid.trb.org/view.aspx?id=686675

Brands, T. (2014, September). Multimodal modelling. Presented at the Course Public Transport.

Chan, Y. (1976). A method to simplify network representation in transportation planning. Transportation Research, 10(3), 179–191. https://doi.org/10.1016/0041-1647(76)90073-3

Chang, K., Khatib, Z., & Ou, Y. (2002). Effects of zoning structure and network detail on traffic demand modeling. Environment and Planning B: Planning and Design, 29(1), 37–52. https://doi.org/10.1068/b2742

Chen, W.-F., & Liew, J. Y. R. (2003). The civil engineering handbook. Boca Raton [FL]: CRC Press. Retrieved from http://www.crcnetbase.com/isbn/9780849309588

Connors, R. D., & Watling, D. P. (2008). Aggregation of transport networks using sensitivity analysis. EUROPEAN TRANSPORT CONFERENCE 2008; PROCEEDINGS. Retrieved from http://trid.trb.org/view.aspx?id=935125

Dafermos, S., & Sparrow, F. (1969). The traffic assignment problem for a general network. Journal of Research of the National Bureau of Standards B, 73(2), 91–118.

Daganzo, C. F. (1980). An equilibrium algorithm for the spatial aggregation problem of traffic assignment. Transportation Research Part B: Methodological, 14(3), 221–228. https://doi.org/10.1016/0191-2615(80)90001-6

de Dios Ortuzar, J., & Willumsen, L. G. (2011). Modelling transport. John Wiley & Sons.

De Feijter, E. D. (2012, April 5). Towards a suitable quick scan transport model. Retrieved from http://resolver.tudelft.nl/uuid:76d73195-364b-4729-b247-559be0a240fb

Fotheringham, A. S., & Rogerson, P. (Eds.). (2009). The SAGE handbook of spatial analysis. Los Angeles ; London: SAGE Publications.

Guo, J., & Bhat, C. (2004). Modifiable areal units: Problem or perception in modeling of residential location choice? Transportation Research Record: Journal of the Transportation Research Board, (1898), 138–147.

Hagen-Zanker, A., & Jin, Y. (2015). Adaptive Zoning for Efficient Transport Modelling in Urban Models. In O. Gervasi, B. Murgante, S. Misra, M. L. Gavrilova, A. M. A. C. Rocha, C. Torre, … B. O. Apduhan (Eds.), Computational Science and Its Applications -- ICCSA 2015 (pp. 673–687). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-21470-2_49

Jeon, J.-H., Kho, S.-Y., Park, J. J., & Kim, D.-K. (2012). Effects of spatial aggregation level on an urban transportation planning model. KSCE Journal of Civil Engineering, 16(5), 835–844. https://doi.org/10.1007/s12205-012-1400-4

Lopes, S. B., Brondino, N. C. M., & Da, S. (2014). GIS-based analytical tools for transport planning: Spatial regression models for transportation demand forecast. ISPRS International Journal of Geo-Information, 3(2), 565–583. https://doi.org/10.3390/ijgi3020565

Martens, M., & Jong, M. de. (2009). Governance Verkeersmodellen: Uitdagingen voor de toekomst. In Colloquium Vervoersplanologisch Speurwerk, Antwerpen. Retrieved from http://www.cvs-congres.nl/cvspdfdocs/cvs09_213.pdf

Martínez, L. M., Viegas, J. M., & Silva, E. A. (2009). A traffic analysis zone definition: a new methodology and algorithm. Transportation, 36(5), 581–599. https://doi.org/10.1007/s11116-009-9214-z

Mihyeon Jeon, C., & Amekudzi, A. (2005). Addressing Sustainability in Transportation Systems: Definitions, Indicators, and Metrics. Journal of Infrastructure Systems, 11(1), 31–50. https://doi.org/10.1061/(ASCE)1076-0342(2005)11:1(31)

Páez, A., & Scott, D. M. (2005). Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal, 61(1), 53–67. https://doi.org/10.1007/s10708-005-0877-5

Raadsen, M. P. H., Schilpzand, M. P., & Mein, E. (2009). Applying Inter-Regional Shared Routes in Detailed Multiregional Dynamic Traffic Models. In European Transport Conference, 2009. Retrieved from http://trid.trb.org/view.aspx?id=1107483

Sean Qian, Z., & Zhang, H. (2012). On centroid connectors in static traffic assignment: Their effects on flow patterns and how to optimize their selections. Transportation Research Part B: Methodological, 46(10), 1489–1503. https://doi.org/10.1016/j.trb.2012.07.006

Sheffi, Y. (1985). Urban transportation network. Equilibrium Analysis with Mathematical Programming Methods, Prentice Hall. Retrieved from http://bin.t.u-tokyo.ac.jp/startup14/file/1-1.pdf

Suhng, B. M., & Lee, W. (2013, July). A New Link-Based Single Tree Building Algorithm for Shortest Path Searching in an Urban Road Transportation Network. Retrieved 7 September 2016, from http://www.dbpia.co.kr

Wardrop, J. G., & Whitehead, J. I. (1952). Correspondence. some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineers, 1(5), 767–768. https://doi.org/10.1680/ipeds.1952.11362

You, J., Nedović-Budić, Z., & Kim, T. J. (1998). A GIS-based traffic analysis zone design: technique. Transportation Planning and Technology, 21(1–2), 45–68. https://doi.org/10.1080/03081069708717601