THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

# Foliated Transportation Networks

## Evaluating feasibility and potential

Joakim Kalantari

Department of Technology Management and Economics
Division of Logistics and Transportation
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2012

Foliated Transportation Networks
Evaluating feasibility and potential
JOAKIM KALANTARI
ISBN 978-91-7385-769-7



Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 3450
ISSN 0346-718X

Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Chalmers Reproservice
Gothenburg, Sweden 2012

”Ignorance more frequently begets confidence than does knowledge”

-Charles Darwin

**Foliated Transportation Networks**
**Evaluating Feasibility and Potential**

Joakim Kalantari
Department of Technology Management and Economics
Division of Logistics and Transportation
Chalmers University of Technology

# Abstract

Neither the critical role that transportation plays for economic activities nor its negative external effects can be denied. At the same time, the resources required to sustain and develop the transportation system to an adequate degree and at the required pace are limited. The problem that these properties result in is that the inputs of the system are finite and scarce and the external effects of it negative, but the output is too critical to do without. The most obvious path to a solution is, therefore, trying to increase output obtained from the same or a lesser amount of input. In light of this, the attempt to utilize the existing overcapacity, whatever its extent may be, and improving the operational efficiency with sustained or improved effectiveness, stands out as one of the most viable approaches.

In this thesis the concept of Foliated Transportation Networks (FTN) is evaluated with regard to feasibility of its implementation and its potential impact on the performance of the transportation network. The main objective of the concept of FTN is achieving performance improvements by foliating two different network structures, i.e., direct shipment and hub and spoke, in order to minimize the underutilized units in the network and thereby achieving performance improvements.

The studies show that FTN is feasible to implement in existing networks with limited requirements for additional investments in new technologies. Even though new technological platforms and innovations would be beneficial for the implementation of FTN, a majority of its identified potential can be accessed using existing technologies and rule-of-thumb control. The performance improvement potential that is identified and measured in number of units, traffic work and load factor is substantial. In addition, partial implementation is possible and about 80% of the potential could be realized when about 20% of the system is available for foliated control. The identified potential has proved to be robust following numerous sensitivity analyses.

*Keywords: Transportation networks, mixed model transportation, hybrid transportation systems, transportation network modeling, transportation network simulation, transportation planning and control, transportation network optimization, transportation network performance, transportation network efficiency, intelligent transportation systems.*

## List of appended papers

### Paper 1

**Research Outlook on a Mixed Model Transportation Network**

Joakim Kalantari and Henrik Sternberg

*Published in* European Transport: International Journal of Transport Economics, Engineering and Law*, 2009, Vol. 41, pp. 62-79.*

### Paper 2

**Stepwise Replacement of a Direct Shipment Network with a Hub and Spoke System**

Joakim Kalantari and Kent Lumsden

*Published in the proceedings of the Nofoma conference, June 7-8, 2007, Reykjavik, Iceland.*

### Paper 3

**Quantifying the Performance Improvement Potential of Foliated Transportation Networks**

Joakim Kalantari and Per Medbo

*Published in* European Transport: International Journal of Transport Economics, Engineering and Law*, 2011, Vol. 49, pp. 24-37.*

### Paper 4

**The Impact of Differentiated Control on the Performance of Foliated Transportation Networks**

Joakim Kalantari

*Published in* European Transport: International Journal of Transport Economics, Engineering and Law*, 2012, Vol. 52, no. 6.*

### Paper 5

**In-Transit Services and Foliated Control: The Use of Smart Goods in Transportation Networks**

Joakim Kalantari, Jan Holmström and Per-Olof Arnäs

*Under review for publication in* Journal of Supply Chain Management*. An earlier version has been double blind refereed and presented at Academy of Management, August 5-6, 2012, Boston, MA., USA.*

# Acknowledgements

When I embarked on this journey some five and a half years ago – and by journey I mean the process that had this thesis as its ultimate goal – I did not possess the foresight to realize that this would not be different from any other journey, in that, by the end of it you are so thoroughly sick of the road that just getting there seems to be the only thing that matters. I am confident though that retrospective reflection will confirm this feeling to be a fleeting one. Home-stretch impatience and frustration notwithstanding, there would not have been a journey, or at least not one that would successfully reach its destination, without the instrumental aid and backing of many people and organizations. I am indebted to all of you!

First of all, I would like to gratefully acknowledge Vinnova (The Swedish Governmental Agency for Innovation Systems) and LTS (Logistics and Transport Society) who have funded my project and without whose financial support this thesis would not have been possible. Special thanks go to Bengt Wennerberg for his enthusiastic and unwavering support.

Many friends and colleagues that I had the pleasure to get to know throughout these years have influenced me and my work in ways that are much appreciated but difficult to express. Others have been involved more directly in creating this thesis, like my co-authors, whose contributions, inspiration and cooperation have made possible most of the papers appended to this thesis. Please accept my gratitude – Professor Kent Lumsden, Dr. Henrik Sternberg, Mr. Per Medbo, Dr. Per-Olof Arnäs and Professor Jan Holmström. Special thanks to Associate Professor Daniel Hellström for his meticulous final review of this work. The final result is so much better thanks to your efforts.

I have no choice but to keep repeating some names, because of the many roles in which they have generously offered their support, without expecting anything in return. Professor Lumsden – who, in me, saw abilities that I did not and gave me this life-altering opportunity and acted as my head supervisor in this process – Dr. Per-Olof Arnäs – who was my co-supervisor and an inspiration, in that, he was the respondent at the very first PhD-defense I attended – and Dr. Violeta Roso – who has been a co-supervisor and a friend with the special talent of building me up with encouraging words and sound advice when the burden has felt almost unbearable. I owe all of you a debt that I cannot even hope to repay. I'll do my best to pay it forward to someone else. You have given selflessly to me of your time and wisdom and I genuinely appreciate that! Once again, thank you for this unique and invaluable experience.

To all my colleagues at the department; I'm grateful for having had the privilege of getting to know all of you. You have enriched my life! I would like to thank you for all the laughs we've shared. I know I will miss seeing all of you every day. Every morning, I've looked forward to coming to the office because of all the joy that you all have brought in to my life. Thank you!

Also, if this is even possible, I would like to thank Sweden. My family and I migrated here in 1992 and the way we were welcomed here and the opportunities that were made readily available to us has left me humble with gratitude. This is not just a country in which I've settled. This is a country that allowed me to make it my home. I'm a better, more fulfilled person for having spent my formative teenage years and adult life in a society built on the values of enlightenment and tolerance. Growing up here has made it possible for me to realize more of my potential than I could have ever had wished for. I hope that I can return the favor in whatever small way that I'm able to.

Kent; you are my hero! I can't think of anything that would ever change that.

Finally, to my family; I am the man I am today largely because of my loving parents and how they brought up my sister and myself. I'll do my best to thank you the only way I know how – by trying to be as good a parent to Kimia. My daughter Kimia and my wife Sanam – you are the lights of my life. I couldn't have done this without you. Even if I could, I wouldn't have had any reason to. You make me want to strive to be better. And now, even if I fail, at least we can hope to get better tables at restaurants (well, while vacationing abroad anyway).

Göteborg
29/10 -2012

Joakim Kalantari

# Table of contents

# List of figures

# List of tables

# List of definitions

| | |
|---|---|
| *Bin packing problem* | Bin Packing Problem is a special case of a multiple subset sum problem where the capacity is constant (Martello and Toth, 1990). Solving the bin packing problem is to give the most efficient way to use a limited resource. |
| *Business model* | Refers to models for planning and controlling the transportation network with regard to business considerations as opposed to purely technical ones, which is the focus of the thesis. |
| *Change making problem* | A particular bounded case of the general knapsack problem is the so-called Change Making Problem where the profit is constant and set to 1 and the capacity is constraint (Martello and Toth, 1990). |
| *Direct shipment (DS)* | A transportation network setup where all the nodes are directly connected and the goods travel the shortest way to their destination without additional coordination with other nodes. |
| *Effectiveness* | A measure of quality of the output of a process. It is a ratio of the actual output and the nominal output of a system, e.g., shipments on time/total shipments. |
| *Efficiency* | A ratio of inputs and outputs of a system. Performance consists of an efficiency and an effectiveness component. Efficiency is a compound of utilization and productivity. |
| *Fill rate (load factor)* | The ratio of amount of goods and the capacity employed. |
| *Foliated Transportation Networks* | Conceptual transportation network model where different transportation network setups are simultaneously used in the same system, i.e., where networks are foliated. |
| *Hub and spoke (HS)* | A transportation network setup where all the nodes are only connected to the hub and the goods always travel via the hub. |

| | |
|---|---|
| *Key performance indicator* | Measurable indicators that are used for assessing the performance of a system. |
| *Knapsack problem* | Knapsack Problems are a mathematical formulation of a general group of maximization problems related to capacity constraint (Martello and Toth, 1990). |
| *Overcapacity* | Denotes operational underutilization of resources. Overcapacity is an indication that efficiency improvements are possible. |
| *Performance* | A construct consisting of effectiveness and efficiency. |
| *Productivity* | A construct that measures transformational efficiency. It is a ratio of the actual output and the actual input of a system, e.g., transport work/traffic work. This is one of the two components of the construct of efficiency. |
| *Spare capacity* | Spare capacity; to be distinguished from overcapacity in that it does not denote underutilization in an operational sense. Spare capacity is present and idle in order to manage fluctuation in demand or un-/expected disturbances. |
| *Standby shipment* | Shipments that are accepted for delivery on the condition of available capacity, i.e., the shipments are accepted and delivered only when overcapacity arises. |
| *Subset sum problem* | A special bounded case of the general 0/1 knapsack problem where the weight and profit are equal (Martello and Toth, 1990). |
| *Traffic work* | Truck capacity multiplied by the distance traveled. |
| *Transport work* | Total amount of goods shipped multiplied by the distance traveled. |
| *Transportation network* | The nodes in a transportation system, which represent geographical positions, are interconnected by links to a network (Lumsden, 1995). |
| *Transportation network layer* | In an FTN, each foliated sub network is considered a network layer in the overall system. |

| | |
|---|---|
| *Utilization* | A construct that measures “input usage.” It is a ratio of the actual input and nominal input of a system, e.g., utilized capacity/available capacity. This is one of the two components of the construct of efficiency. |

# 1 Introduction

*In this section the object of study will be introduced along with a description of the problem area. Also, the purpose and research questions are presented in this chapter.*

It is not without some doubt and deliberation that one includes a term in the title of one's dissertation that few have heard of and possibly even fewer know what it is to mean in this context. It was, however, unavoidable in this particular case as the concept of Foliated Transportation Networks (FTN) is the central object of study in this thesis. The Merriam-Webster dictionary online[1] defines "foliated" as: "composed of or separable into layers." This is supposedly the most accurate translation of the Swedish term "överlagrad," which was coined by Professor Kent Lumsden to illustrate the core tenet of the concept. The concept of FTN is based on the idea of layering more than one system design in the same physical transportation network.

The specific FTN design that is the object of study in this thesis is one where a hub and spoke (HS) network is foliated on a direct shipment (DS) network (Figure 1-1). Note that the physical network needs no altering when an HS layer is foliated on an existing DS structure.

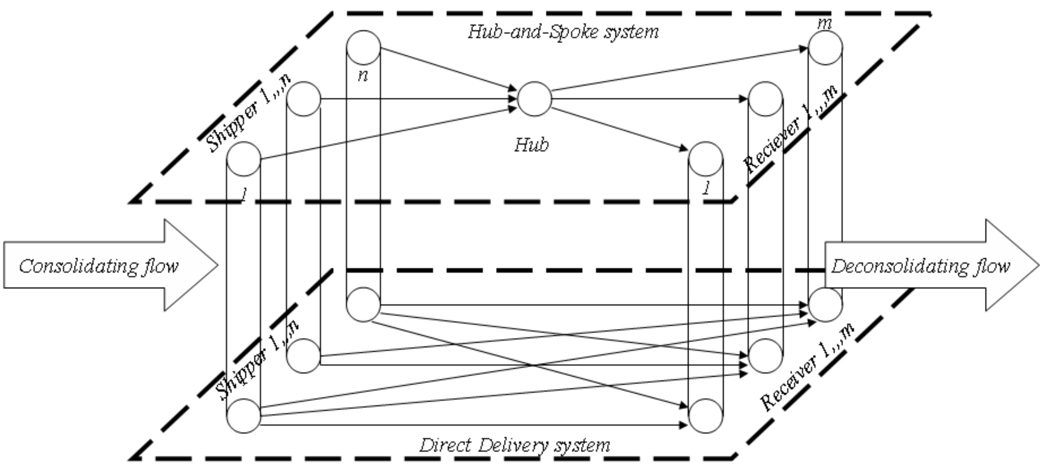


*Figure 1-1 Conceptual model of FTN (Persson and Lumsden, 2006)*

The basic idea of FTN here is to send only full units directly (DS layer) in each relation. Any amount of goods in any relation that is not enough to fill

[1] http://www.merriam-webster.com/dictionary/foliated (2012-09-13)

an entire unit is to be consolidated in the HS layer of the network. Doing so leads to significantly reducing the number of links where underutilized units may be sent. At the same time, efficiency benefits of consolidation in the HS layer will also be realized. These properties ought to result in FTN being able to achieve performance levels that are not possible to achieve if only one of the constituting layers were utilized in isolation, given everything else equal.

The purpose of this thesis is to evaluate the concept of Foliated Transportation Networks.

## 1.1 Problem area

The role of an effective and efficient[2] transportation system as a critical enabler of trade and economic growth is well established in the literature (e. g. McKinnon, 2006; Rodrigue et al., 2006; Cowie, 2010). Equally, the negative external effects and risks of our current transportation system are well documented as well (Stern, 2007; Taylor, 2007). Transportation affects the environment negatively in terms of emissions, e.g., $NO_X$, HC, $CO_2$, particles and air and water pollution but also in terms of noise pollution, congestion and traffic hazards. The negative external effects of transportation coupled with its significance for a vibrant and growing economy, makes the performance of the transportation system relevant for society as a whole (Vinnova, 2004; Regeringen, 2008).

The performance of the transportation system has lately garnered a lot of attention from researchers and policy makers alike (European Commission, 2006). Even though the critical role of an effective transportation system is not lost on policy makers, the attention of researchers and policy makers appears to be focused primarily on transportation efficiency and the mitigation of the sectors' negative external effects (Miljödepartementet, 2008; Regeringen, 2008). In addition, changing cost structures due to the development of energy prices and policy-driven incentives aimed at dampening the transport sector's negative environmental impacts also contribute to the growing interest in addressing the transportation system from an efficiency perspective.

Looking at the road-bound freight transportation industry; the data reveals a system-wide overcapacity and low utilization with regard to a number of key performance indicators (McKinnon, 2010). The reason for the origin of this overcapacity is not as clearly indicated. The conclusion can be drawn that the demand and cost structure of the road-bound transportation

[2] Performance = [Utilization, Productivity, Effectiveness] = [Efficiency, Effectiveness]. Efficiency = [Utilization, Productivity]. See section 2.2.

system promotes overcapacity (Lumsden, 1995; Sternberg, 2011). This observation is not uncommon when viewing any service-producing industry. The term "overcapacity" should be distinguished from spare capacity. Overcapacity is signified by underutilization and is therefore detrimental to operational efficiency. Spare capacity regards passive capacity or additional capacity, e.g., extra trucks, personnel etc. that can be activated when the need arises, e.g., to handle fluctuation in demand. On these bases, a conclusion could be drawn that there exist real problems regarding the efficient use of transportation resources. This interpretation is widespread among researchers and policy makers.

The invocations of inefficiencies of the transportation systems are, however, in part unsubstantiated. This is the result of our demonstrated inability to concisely define, operationalize and assess the efficiency of transportation systems with a single, or a set of, comprehensive and universally accepted units of analysis. This has in turn resulted in a lack of reliable and valid data (McKinnon, 2009; McKinnon, 2010). Existing estimates of the efficiency of various transportation systems rely on key performance indicators (KPI) that individually cannot provide a complete picture and more often than not end up presenting measures of constructs that are not always meaningful as a representation of transportation efficiency (McKinnon, 2009). Aside from the inadequacies of constructs aimed at measuring efficiency, the data supporting the statistics are often less than fully applicable for the purpose as well. Often 50%-max systems[3] and structural and technical flow imbalances are not distinguished when collecting and aggregating data for official statistics (Trafikverket, 2011). Physical and economical KPI do not overlap to a sufficient degree so as to not produce divergent results, and KPI are not measured uniformly across modes, load units and organizations. These properties render recorded statistics incompatible on a higher system level than where the data was collected (Trafikanalys, 2011). These factors contribute to undermine the validity of the claims of systemic inefficiencies in the transportation system; partially due to the poor understanding of what is meant by efficiency but mainly because of an inability to reliably operationalize a comprehensive and meaningful set of measures of efficiency when it comes to transportation systems.

Nevertheless, the potentially wanting validity of the claims of an inefficient system does not preclude the necessity of efficiency improvements. Scale economy and low margins are two distinctive characteristics of the trans-

[3] Systems that can at best achieve 50% capacity utilization as measured on the round trip e.g. the distribution systems of sectors such as mining, forestry, crude oil, agriculture etc.

portation industry (Hultén, 1997). Here, the physical and economical characteristics act as mutually reinforcing conditions favoring even small increments in the improvement of efficiency. This means that overtly poor utilization and wasteful consumption of resources in the transportation system is not a precondition for motivating research seeking radical or incremental efficiency improvements. It is logical to stipulate that efficiency improvements hold intrinsic value as long as the cost for realizing them can be motivated by the potential gains.

The major constraint when attempting to increase transportation efficiency is that efficiency improvements cannot come at the cost of the effectiveness of the system. Existing research shows that even though the deciding factor regarding the choice of transportation providers is almost always the price, the qualifying conditions such as lead time, flexibility, service quality and so on are often non-negotiable and both conflicting and difficult to efficiently obtain (e.g. Saxin et al., 2005; Lammgård, 2007). In fact, deviation from the qualifying conditions would in many cases automatically disqualify a transport service provider regardless of the price offered. This also supports the reasoning that deliberate introduction and maintaining of overcapacity is a strategy for coping with the qualifying demands of the service buyers (Lumsden, 1995).

Regardless of whether the starting point is a system riddled with overcapacity or one with limited potential for efficiency improvement, the magnitude of the problem, i.e., the negative external effects of transportation, is increasing in the foreseeable future. It is not only the rapid past growth or even the sizeable projected growth (European Commission, 2006) of the transportation demand that amplifies the significance of this issue but the fact that historically, the growth has not been absorbed by identified existing overcapacity (Rodrigue, 1999; Flodén, 2007). This observation further indicates that transportation systems require or are benefited by maintaining overcapacity. This development is alarming because what the overcapacity aims to satisfy is being disabled due to the projected near saturation of the infrastructural capacity, e.g., the traffic system (Rodrigue, 1999; Crainic et al., 2004).

Taking measures to limit the growth of the transportation sector is not likely to be the preferred option due to the vital role that transportation plays in a functioning, healthy and growing economy. Transport activity has historically been considered as an economic indicator. A rapid expansion of the infrastructural capacity allows for only a temporary relief and would also have limited impact in the short term (Brandt et al., 2007). Besides, this approach is not likely to be preferable as it does not address the negative external effects of transportation. Also the cost and feasibility of these types of solutions are not undisputed. Technological advancement, though promising, cannot be counted on to provide short-term relief. Hence, increasing the performance of the existing transportation systems is a goal shared by industry and society alike.

The gist of the problem is that the inputs of the system are finite and scarce and its external effects predominantly negative, but at the same time, the output is too critical to do without. The most obvious path to a solution is, therefore, trying to increase output obtained from the same or a lesser amount of input. In light of this, the attempt to utilize the existing overcapacity, whatever its extent may be, and improving the operational efficiency with sustained or improved effectiveness, stands out as one of the most viable approaches. The main objective of the concept of FTN is achieving performance improvements by doing just that.

## 1.2 Purpose and research questions

For reasons cited above there is a need for making more efficient use of underutilized resources in the transportation networks. However, this cannot lead to any deterioration in service quality for the shippers and the services offered to them. The conceptual model of FTN is designed to provide such an improvement. It is further assumed that the rapid development of information, communication and identification technologies is a key factor in creating a technological basis for concepts such as FTN to be designed and developed to a state of operational feasibility (Persson, 2006a; Persson and Waidringer, 2006).

Fragments of the idea of FTN are already in use in the transportation networks of today as part of a strategy to make use of the existing overcapacity in the networks. However, the current state of development of the design of FTN is not close to what would be considered an operational model. To determine whether FTN could be considered a relevant solution to the problem area discussed above, it is necessary to evaluate the feasibility of an operational design of FTN as well as its potential impact on the transportation network performance. As it stands now, the operational feasibility of FTN has only been rather loosely theorized and hypothesized (Persson, 2006a; Persson and Lumsden 2006; Persson and Waidringer, 2006), as opposed to substantially evaluated. This means that the need for research on the feasibility of introducing FTN in existing real-world systems is urgently required. Furthermore, the expected improvement potential of FTN for network performance is largely hypothesized from the existing theory and logical inferences. These hypotheses need to be tested based on empirical evidence. In short, FTN needs to be evaluated based on the criterion of its operational feasibility and potential impact on network performance.

Hence the purpose of this research is:

*To evaluate the concept of Foliated Transportation Networks.*

When pursuing this purpose, the two above identified key issues need to be addressed primarily; namely those of operational feasibility and potential impact on the network performance. This is accomplished by devising two research questions (RQ) where each one is aimed to address one of these issues respectively. The first RQ pertains to evaluating the operational feasibility of introducing a Foliated Transportation Network model. The second RQ aims to evaluate the potential impact from the implementation of such a foliated transportation model on network performance as compared to existing network designs.

In regard to the first RQ it is helpful to consider that no real-life implementation of FTN exists today. In fact, this is the very reason why it is interesting to explore and evaluate the operational feasibility of implementing FTN. Any implementation is impeded by the fact that the present model of FTN is on a conceptual level, i.e., the level of abstraction at which the model is expressed is not readily transferable to an operational design. In light of that, any evaluation of the operational feasibility of FTN should entail the identification of the prerequisites for an operational FTN design. More plainly put, a first step is to identify the design gaps to be bridged between the existing theories and systems and the conceptual model of FTN. Such gaps may both regard the need for new knowledge as well as new applications for existing theories and technologies.

Furthermore, given a set of identified challenges for designing an operationally feasible model of FTN, the need for and the approach necessary for overcoming these challenges would not likely be readily apparent. Therefore, it is of interest to focus the evaluation on what is operationally feasible to achieve as compared with the conceptual model. Hence, the first research question is as follows:

RQ1: *What are the challenges for designing an operationally feasible Foliated Transportation Network, and how can these challenges be overcome?*

The second RQ is posed in order to address the issue of evaluating the potential impact of FTN on the transportation network performance. The answer to the first RQ provides the feasibility parameters regarding the design of an operational model of FTN. The identification of what can be done in the existing systems and what remaining design effort is necessary helps decide the parameters of a feasible operational design of an FTN. This will provide the modeling underpinning that is necessary for creating a model at the appropriate level of abstraction that would allow evaluation of the impact of FTN on network performance.

However, before attempting to create a model on which to base the evaluation of the potential impact, it is necessary to be able to do the same regarding the individual layers of the network. In this thesis, the two predominant network structures of direct shipment (DS) and hub and spoke (HS) are studied. It is required to model and measure network performance in the process of pursuing the second RQ.

The individual network layers, in this context, are made up of the same physical network and cannot physically be distinguished from one another. The network layers exist only as different ways of controlling the goods and the network resources as the goods are transshipped from an origin to a destination. In any case, the individual layers of the foliated network must be modeled as well as performance defined regarding measurable constructs vis-à-vis network performance. The ability to do so is a key factor in fulfilling the purpose of this thesis. Based on this, the final research question is devised as follows:

RQ2: *How would foliating a hub and spoke network over a direct shipment network affect the network performance?*

These RQs are meant to capture the key issues mentioned above as necessary to fulfill the purpose stated.  RQ1 aims to provide an evaluation of the operational feasibility of the purposed Foliated Transportation Network design, whereas the second RQ is concerned with the evaluation of the potential impact of an FTN on the network performance as compared to the original design. The purpose and research questions are presented in Figure 1-2.

In answering the first research question, it will be clarified what remains to be done regarding the design effort for developing the concept of FTN into an operational model. However, the evaluation of the operational feasibility of the concept with regard to identified challenges is very much dependent on to what degree the identified obstacle influences the identified potential impact on network performance. This property makes the first and second RQ closely intertwined. The potential impact on network performance needs to be evaluated based on an operationally feasible design at the same time as different challenges posed generate different effects on the potential of the network.

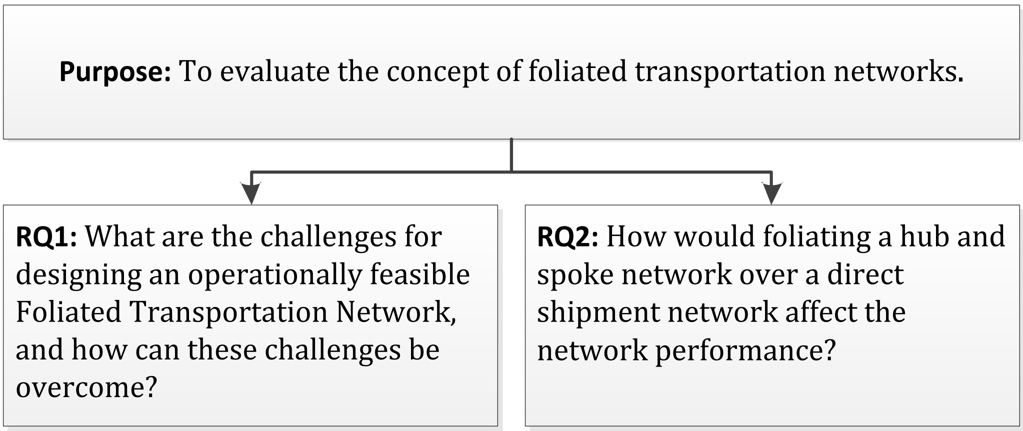


*Figure 1-2 The purpose and research questions*

## 1.3 Scope and delimitations

This research takes the perspective of the transport network operator and limits its empirical sphere to that of Swedish domestic general cargo freight transportation. Only the long haul portion of the transportation network (i.e. between terminals) is considered and the pick-up/delivery operations to and from the terminals in the network are delimited. From this empirical sample, generally valid conclusions are to be drawn about FTN.

The research focuses on network performance with regard to physical units of analysis and deliberately delimits monetary units of analysis. For one, costs are difficult to unambiguously assign to different activities. Secondly, the primary driver of this research aims to improve the performance of transportation networks with regard to the utilization of physical resources but not necessarily the economic ones. Of course economics is an important factor from a business perspective, but the research is driven from a technical perspective. Monetary cost can be aggregated indiscriminately where superior performance in one area can compensate sub-par performance of another. Therefore, in an analysis from a technical perspective, using monetary units of analysis can lead to sub-optimum solutions. Finally, the transferability of technical parameters is likely to improve the prospect of arriving at results that are insensitive to external economic factors such as changes in the price structure (e.g., volatility in the price of energy, labor, capital, etc.), political decisions (e.g., taxes and regulation) or local differences and price variations.

It is worth noting that the phenomenon of foliation is a broader construct than the special case of foliating DS and HS networks under study here. In this thesis, the FTN studied will consist of a DS and an HS layer. Any comparison is done between the new design, i.e., FTN, and DS. The logic behind this choice is that, examining previous research and the theoretical underpinnings of the design of FTN, it becomes clear that foliating an HS layer on a DS network is the rationally viable combination, and that the converse does not apply.

For one, the improvement in efficiency is hypothesized to stem from reducing the number of links in which underutilized units may be sent. Such reduction is possible only when foliating an HS over a DS network and not the other way around. Secondly, if the volumes of goods being transported through the network are limited enough to warrant the application of a pure HS network, other concepts, e.g., shortcuts by Lumsden et al. (1999) or similar modifications by Woxenius (2007), would be the more logical choice.

Depending on the disciplinary perspective of the beholder, different parts of the system are essential to consider (Sjöstedt, 2005). The disciplinary focus of this thesis is on transportation, which is signified by the entity focus on vehicles. This means that the underlying conditions or decisions that give rise to the transportation demand is not of interest here, only the manner in which the demand is satisfied. Evaluation is based on the degree to which customer demands are met and the amount of physical resources consumed in achieving demand satisfaction.

## 1.4 Structure of the thesis

The remainder of this text is disposed as follows:

*Chapter 2* includes a review of the existing theories on transportation networks in general and direct shipment, hub and spoke and foliated transportation networks in particular. Also, a discussion of transportation network performance is included in this chapter.

*Chapter 3* discusses the methodological aspects of this thesis. The methodology employed to carry out the research is presented in this section. Also, research quality and the match between method choices with regard to the posed research questions and the compatibility of the different methods used are discussed.

*Chapter 4* introduces the models developed. This includes a mathematical model of HS and DS and a simulation model of FTN.

*Chapter 5* contains the presentation of the appended papers. This thesis is based on the five appended papers that will be briefly introduced in this section. The relationship between the papers, the research questions and purpose is also presented here.

*Chapter 6* presents the results of the studies and a discussion of the results with regard to the purpose and research questions.

*Chapter 7* takes up the results and conclusions of this thesis. In this section, the research questions posed are addressed based on the combined results of the studies. Also, the practical, theoretical implications of the results are discussed. Moreover, the need and direction for future research will be discussed.

# 2 Frame of reference

*In this section, existing theories on transportation networks in general and direct shipment, hub and spoke and foliated transportation networks in particular is reviewed. Also, a discussion of transportation network performance is included in this chapter.*

Transportation is normally associated with the movement of goods from one node in a distribution network to another. In transportation, attempts are made to solve this problem by ensuring that goods are moved as quickly, efficiently and consistently as possible from the point of origin to the point of consumption (Ross, 1996).

Any movement is prompted by a demand that has arisen as a result of economic activity. The top level of the conceptual model of Wandel et al. (1992) denotes the supply chain where the demand originates (Figure 2-1). The properties of demand in terms of frequency, lead-time, shipment size, delivery precision and flexibility are also results of activities and decisions on this level. On the second level, where the physical movement of goods is performed, lies the focus of this thesis.

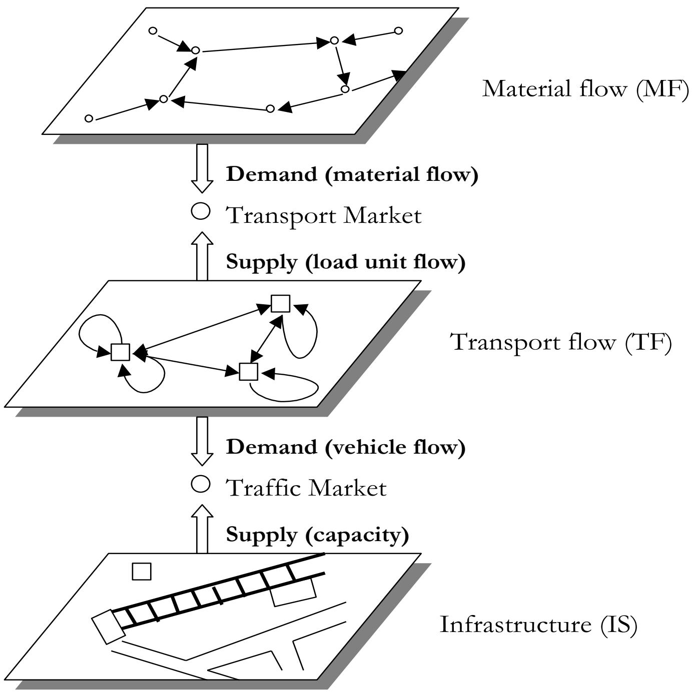


*Figure 2-1 Three layer model of freight transportation systems (Wandel et al., 1992)*

There are many individuals, groups and organizations whose decisions interact to affect the transportation system and thus the pattern of flows. The user of the transportation, i.e., a shipper of goods, makes a decision about when, where and whether the goods should be transported and how often. Here, regardless of the logic behind the decision for a movement or the utility sought by the shipper, the aim is to satisfy that demand as effectively and efficiently as possible. The third layer, concerned with land use, also falls outside the scope of this thesis.

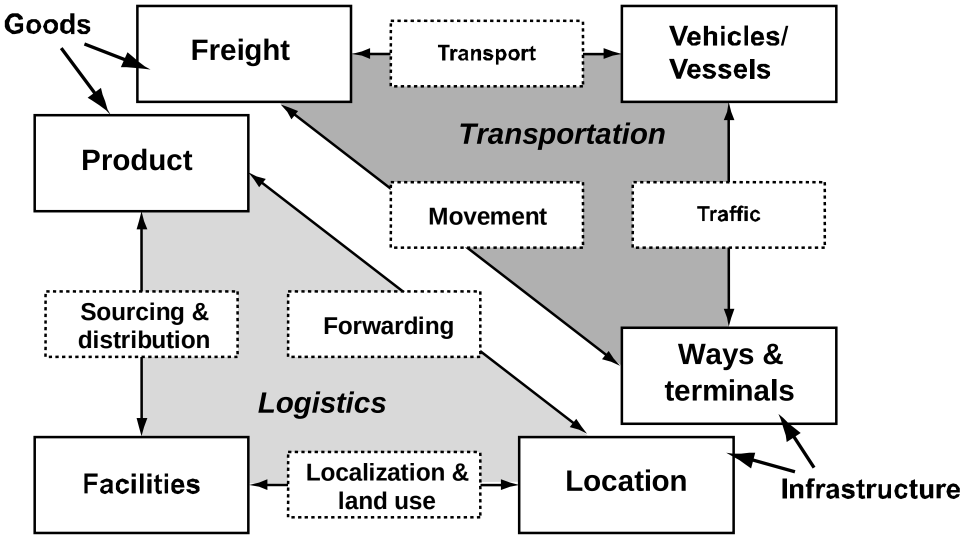


*Figure 2-2 The relationship between logistics and transportation (Sjöstedt, 2005)*

In the conceptual model of Sjöstdt (2005), the relationship between the logistics and transportation systems is viewed neither as a hierarchy nor as one where one is a subsection of the other (Figure 2-2). Logistics and transportation are viewed as parts of a whole where each part contains its own independent systemic logic.

In the remainder of this chapter, transportation network theory is reviewed with special focus on direct shipment (DS) and hub and spoke (HS) networks as well the existing literature and the theoretical underpinning of foliated transportation networks (FTN). Furthermore, it is attempted here to present meaningful operationalization of transportation network performance and to present relevant constructs to that effect. The chapter is concluded with the presentation of some relevant constructs from the field of combinatorial mathematics. Though this is not a theoretical domain that is applied, it is necessary to be included in order for the reader to be familiarized with the formal mathematical representation of problems that are theoretically relevant.

## 2.1 Transportation networks

Transportation networks can be described in terms of nodes and links (see Figure 2-3). The nodes and the links make up the physical network. Many different combinations of nodes and links may be configured in order to complete a relation. A relation is a direction-dependent connection between two nodes in a network, meaning, A to B is a different relation than B to A. A relation is to be considered as an abstraction, i.e., disconnected from the physical network (Lumsden, 2006).

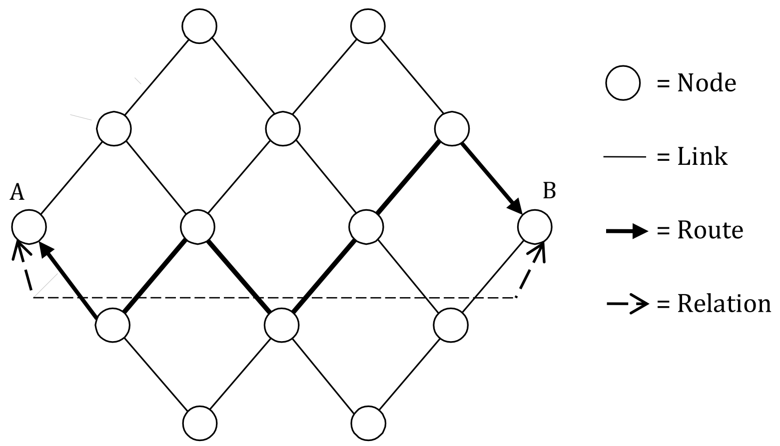


*Figure 2-3 Model of a transportation network (Lumsden, 2006)*

A relation may be completed via any number of nodes and links configurations, referred to as a route. Nodes in a transportation network could act as sources and sinks, where a source node is where the transport is initiated and the destination node is referred to as a sink (Lumsden, 1995).

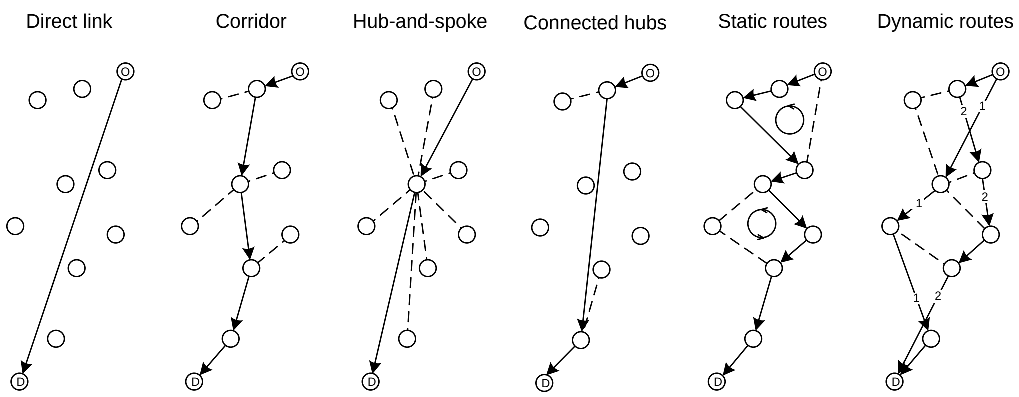


*Figure 2-4 Options for transport from origin (O) to destination (D) in a network (Woxenius, 2007)*

The transportation network's most common designs can be divided into two principle categories: direct shipment or hub and spoke (Crainic, 2003; Lumsden, 2006). In practice, one seldom finds any pure systems (Crainic, 2002; Woxenius, 2007). Figure 2-4 is an illustration of some of the occurring variations according to Woxenius (2007).

The links in a transportation network correspond to highways, rail tracks, seaways or urban streets, and the nodes express the connectivity relations of links in the network, e.g., warehouses, distribution centers, freight terminals and ports (Manheim, 1979). Nodes are collecting and consigning points in the network where the goods are collected, transshipped or sorted for transport (Rodrigue et al., 2006). Hultén (1997) describes the function of a node beyond that of connectivity alone, to include bridging the gap of frequency, capacity and time between transportation demand and supply (see Figure 2-5). Buskhe (1993) also includes the load-bearing resources or vehicles in the network model from a capacity utilization perspective.

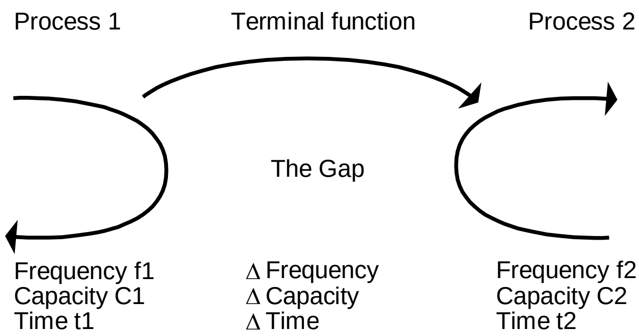


*Figure 2-5 Model of the functionality of a terminal (Hultén, 1997)*

Options or decision variables are those aspects of the transportation system that can be directly changed by the decisions of one or several individuals or organizations. Manheim (1979) outlines possible options when it comes to transportation system operating policies. This set of options includes the full spectrum of decisions about how the transportation system is operated. The networks of nodes and links, and vehicles establish an envelope of possibilities; within that envelope a large variety of detailed operating decisions must be made. These options include vehicle routes and schedules, types of services to be offered, including services auxiliary to transportation (diversion and re-consignment privileges for freight) and regulatory decisions. A transportation setup is a set of decisions based on options available for a certain flow of goods.

A key enabler of effective transport operations management is having necessary information available (Sternberg, 2008). Every setup requires exchange of information among all the involved participants in order to avoid execution hurdles (Stefansson and Sternberg, 2011).

### 2.1.1 Direct shipment networks

In a direct shipment network (see Figure 2-6) all nodes are interconnected with direct relations (Lumsden et al., 1999). Direct relation means that the only nodes involved are the origin (O) and destination (D), the goods are not consolidated along the way and the transportation is independent of other origin/destination (O/D) pairs (Woxenius, 2007), or in other words, that the transport is dedicated (Lumsden et al., 1999; Crainic, 2003). A DS network is best utilized when the number of nodes in the network is limited, the demand for transportation in every O/D pair is sufficient and the primary optimization parameter is time and flexibility (Hultkrantz, 1999).

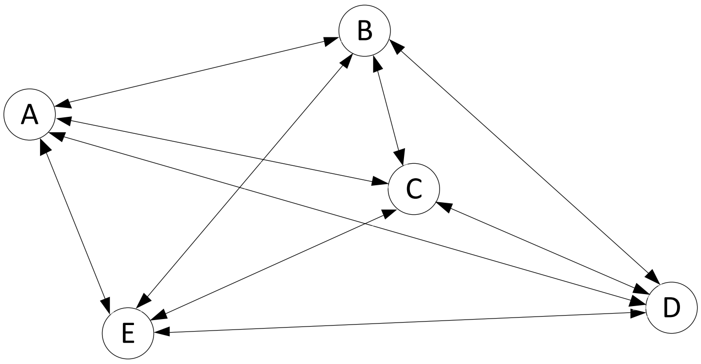


*Figure 2-6 A model of a direct shipment network (Lumsden, 2006)*

A DS setup by default leads to the shortest time in transit as the goods always travel directly, the shortest way and without any additional stops, consolidation operations or handling. A DS network is easily managed, due to the simple governing rules and the fact that transports are independent of one another. In return, the DS setup requires a greater number of resources, e.g., trucks in the system, leads to a lower transportation frequency, and its performance is dependent on sufficient volumes, i.e., the demand in each relation must match the capacity reasonably well in order to achieve acceptable levels of resource utilization (Lumsden et al., 1999).

The number of links in a DS network, assuming perfect connectivity, is equal to Equation 2-1 (Lumsden, 2006).

$$DS_{Links} = \sum_{k=1}^{n-1} K = \frac{n(n-1)}{2} \qquad \textit{Equation 2-1}$$

In Equation 2-1, (n) is the number of nodes in the network. The properties of high need for resources, low frequency of transports and the need for high threshold demand for adding a new node to the network are all derived from this relation. The number of links analogously drives the number of resources required in the network, which in the case of DS is equal to at least (n-1) for any additional node included in the network. The high threshold for adding new nodes to the network refers to the fact that for every new node to be added to the network, there have to exist enough transportation demand to/from that node to every other node in the network that would motivate the commitment of new resources. This amount is roughly equal to at least the amount required to fill the (n-1) load-bearing unit to/from the additional node.

In an ideal typical network of terminals, each terminal serves a specific geographical area. Terminals typically function both as sources and sinks. In a DS network, shipments are subjected to only one consolidation/ deconsolidation step, i.e., consolidation at the origin terminal, transported directly to the destination terminal for deconsolidation and distribution to the individual consignees (Crainic, 2002).

### 2.1.2 Hub and spoke networks

In an HS setup, all the nodes are only interconnected with a/the hub (see Figure 2-7), and in cases where more than one hub exists, all the hubs are also interconnected (Lumsden et al., 1999; Crainic, 2002). Hence, in a single hub network the number of links is equal to Equation 2-2 (Lumsden, 2006).

$$HS_{Links} = n - 1 \qquad \text{Equation 2-2}$$

In Equation 2-2, (n) represents the number of nodes. As compared to a DS network, the reduction in the number of links leads to either fewer re-courses necessary or higher frequency of service (Lumsden et al., 1999). Furthermore, unlike the DS setup, the threshold of demand for including an additional node in the network is the amount needed to fill only one additional load-bearing unit, e.g., truck. A hub network creates a larger spatial coverage and high transport frequency for the network as the volume that flows between the O/D pairs does not need to be very large to be included (Bryan and O'Kelly, 1999).

Other advantages of the HS setup are high resource utilization rate regard-ing load capacity, a lower number of resources in the network and more leveled flows (Lumsden, 2006). An HS setup is preferred in a network with a vast number of nodes, where aggregation of demand is necessary to attain adequate flow and the primary parameter of optimization is resource utilization and coverage (Bryan and O'Kelly, 1999). On the other hand, deliveries in an HS network almost never run the shortest way, which means the transport work will be greater in an HS setup compared to a DS one. This also means that the mean time between nodes of the network increases, i.e., it takes longer to transport goods in the network (Crainic, 2002).

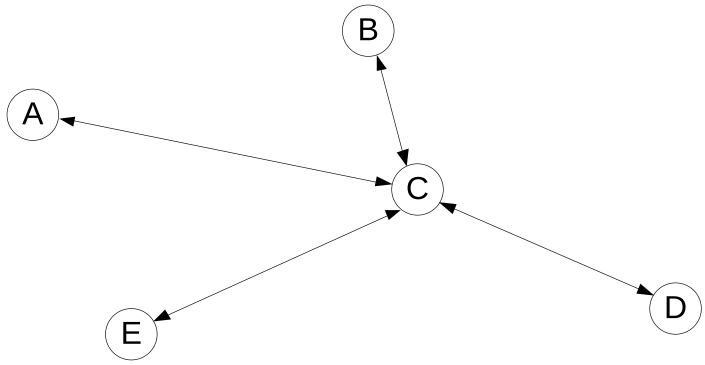


*Figure 2-7 A model of a hub and spoke network (Lumsden et al., 1999)*

The additional time required to transport a shipment through the network is only partially dependent on the longer distance traveled. Terminal handling regarding the extra deconsolidation/consolidation step at the hub is also a prime contributing factor. Time spent in the terminal is not exclusively dependent on the actual terminal operations. Some of the idle time at the terminals will be due to the coordination constraints that hub transshipment entails. Plainly put, the shipment from every node must have arrived at the hub before the deconsolidation/consolidation operations can be completed and the goods can be sent on their way to the destination terminals. This results in "quicker" goods, i.e., goods from origins that are geographically closer to the hub or goods that require less handling, etc., need to stand idle waiting for all the goods from every node to arrive and be handled at the hub. Finally, the longer transit time naturally means that the time window to accept goods for transport for the network operator will in effect be reduced when compared with the DS setup.

The hub and spoke system is at least a two-level system (O'Kelly, 1998). This refers to the fact that at least two consolidation/deconsolidation operations are necessary in a hub and spoke system: one in the origin/destination terminal and at least one in the hub. In cases of more than one hub, additional consolidation/deconsolidation operations are required in the inter-hub transports. The extra handling puts additional strain on transshipment terminals and also increases the risk for lost and damaged goods (Lumsden, 2006). In the hub and spoke setup, the coordination of flows between O/D pairs are dependent on all other O/D pairs and require complementary handling, which all leads to more time in transit (Taylor et al., 1995). The required coordination increases the complexity in the system and witch makes the system more difficult to manage (Arnäs, 2007).

Analysis of hub and spoke networks can be divided into two major directions: spatial organization (O'Kelly and Bryan, 1998) and scheduling and routing (Dobson and Lederer, 1993). The common denominator of these approaches is that of network optimization, though from different perspectives. It is recognized that generally few pure hub and spoke systems can be observed (Taha and Taylor, 1994; O'Kelly, 1998). Identified variations on the different principle designs or routing procedures are all based on exceptions and/or modifications of the original design (see for example, Woxenius (2007), Lumsden et al. (1999), Aykin (1995), Liu et al. (2003), Roy and Crainic (1992) and Zäpfel and Wasner (2002)).

### *2.1.3 Foliated transportation networks*

Bjeljac and Lakobrija (2004), Persson and Lumsden (2006) and Persson and Waidringer (2006) present Foliated Transportation Networks (FTN) as a conceptual model that is designed to improve the efficiency of a transportation network by foliating the two predominant network structures, i.e., direct shipment (DS) and hub and spoke (HS) in the same network. It should be noted that this intervention does not alter the physical network in any real sense, as compared with a pure DS network. The distinction between the different network layers come from the route that the goods travel from origin to destination through the network. For instance, goods routed via the DS layer would be shipped directly between origin and destination, whereas goods routed via the HS layer would be shipped via a hub, i.e., an additional consolidation step. Both routes are possible given a network where it is physically possible to connect any node, to any other, directly, e.g., a DS network.

Bjeljac and Lakobrija (2004) describe FTN[4] as a DS network where only "full" units, with regard to the carrier's loading capacity, are sent directly and all units that are not full are consolidated in an HS sub-layer of the system (see Figure 1-1). The authors examine the feasibility of such an implementation within an existing system with regard to time constraints. The findings suggest that implementation would be possible with negligible impact on service quality of the service provider. The largest modifications needed would be the introduction of a hub terminal and the consolidation operations needed to route quantities of goods that do not fill a whole unit through the hub.

[4] Bjeljac and Lakobrija do not actually use the term FTN. They used the Swedish term "överlagrad," which was originally coined by Professor Kent Lumsden.

In the model presented by Bjelajc and Lakobrija (2004), the hub volumes are identified after the departure of all DS trucks, i.e., the decision to send a sub-set of the total volume in an O/D pair via the hub is made after it is operationally apparent that the remaining volume is not sufficiently large to fill a whole unit. The major drawback of this model design is the fact that the goods that are routed through the hub, i.e., the portion of the shipment in the network that will require the longest time in transit, will depart last. Though feasible, this approach will not be optimal regarding the mean time between the nodes of the network in an FTN setup.

Persson and Lumsden (2006), in their principle design of the model, suggest a setup where the hub volumes are identified in advance and are sent first in order to improve the system's performance regarding mean time between the nodes of the network. The argument is that because the required transit time through an HS network is inherently longer compared to a DS setup, it would, on a system level, be beneficial to afford that portion of the goods the longest time window by shipping it first. This approach would likely require a very high level of accuracy regarding operations planning and control when it comes to identifying the hub volumes in advance (Persson, 2006b; Persson and Waidringer, 2006).

The previously cited studies stipulate that by foliating the two structures, i.e., DS and HS, and dynamically planning, controlling and optimizing the distribution of goods between the two network layers, strengths of the individual setups will be amplified at the same time as their weaknesses will be diminished, resulting in a superior network performance. The FTN is argued to enable a higher performance than what is possible to achieve with any of its constituting layers alone. In other words, the resource utilization and productivity, i.e., efficiency, will be increased without the deterioration of the service quality, i.e., effectiveness, by effectively utilizing a portion of the extant overcapacity in the transportation network.

The idea of tapping into the overcapacity that exists in transportation networks through the implementation of mixed model transportation networks is not a thoroughly novel one. Persson and Waidringer (2006) identify *inter-city freight transportation* (Roy and Crainic, 1992) , *non-strict hubbing* (Aykin, 1995), *mixed truck delivery system* (Liu et al., 2003), *hybrid/extended transportation systems* (Zäpfel and Wasner, 2002), transportation networks with the presence of *inter-hubs* (O'Kelly and Bryan, 1998) and *shortcuts* (Lumsden et al., 1999) as related concepts. Persson and Waidringer (2006) distinguish FTN from the other concepts available in the literature on the basis of the assumption that FTN would only be possible to implement given the recent and future advances and use of information, communication and identification technologies.

What further separates the idea of FTN from other related concepts in the literature, or solutions now in practice, is the aim of foliating two different network structures that are to be managed through dynamic and systematic planning and control (Persson, 2006a; Persson, 2006b; Persson and Lumsden 2006). This objective is also a demarcating criterion between FTN and the related concepts presented above in that FTN alone aims to foliate two different network structures and not merely alter or modify one network structure with exceptions or additional rules. Hence, the phenomenon of foliation is a broader construct than the special case of foliating DS and HS networks.

Persson and Waidringer (2006), the originators[5] of the term "foliated" transportation networks, discuss the linguistic rationale behind this term in the following manner:

> "*From a linguistic point of view, the term foliated has been chosen to illustrate the characteristics of the two systems when they are combined and refers to the way they interact, i.e., the two systems overlap and foliate each other.*"

This systematic combination of two network structures in one network is what makes FTN a unique concept, separating it from other related mixed model transportation network concepts. The concept of FTN is still in the early stages of its design, which means that an empirical implementation of FTN is not to be found and studied. The model presented in previous works cited above is at an abstraction level that is far from an operational model. The fact that the existing research on FTN is based on theoretical reasoning rather than empirical grounding adds to the likelihood of the accuracy of this observation.

There are some empirical solutions to be found where elements or fragments of the same basic idea as FTN are present. Though none of them aim to foliate different networks, all seek to utilize existing overcapacity by differentiating the control of the goods flows. What further differentiates these solutions from the concept of foliation is that they also lack some other characteristics described above. They are managed ad hoc rather than systematically or at the system level. Also, the effectiveness of the

---

[5] The Swedish term "överlagrad" was first used by Bjeljac and Lakobrija (2004) in their thesis. The English translation of "foliated" is however coined collaboratively by Dr. Jonas Waidringer and Mr. Pehr-Ola Pahlén (formerly known as Persson).

system output is not uniform, meaning that depending on the control principle, the quality of the service produced differs. Some examples are (1) stand-by shipments, (2) priority classes of goods and (3) price differentiation. Analogies can also be drawn from the field of air transportation, e.g., air passenger and/or belly hold cargo transportation.

(1) Stand-by shipments refer to goods that are accepted and stored at the terminal and will be delivered only when overcapacity arises, i.e., the buyer accepts a flexible lead time. In practice, those types of goods will be loaded on a carrier any time overcapacity in the required relation occurs; for instance, cars at Scandia Harbor[6] or tires at Schenker's Bäckebol Terminal.[7]

These examples elucidate both the ad hoc nature of the solution and the lack of system overview (i.e., only the selected long-distance legs are regarded on stochastic bases).

(2) Different priority goods are based on the same basic principle as the stand-by shipments, where for an incentive the customer allows the goods a longer transit time; for example, the free delivery option of books from Adlibris[8], an online bookstore. This access to less time-sensitive goods allows the transport network operator to distribute the low priority goods so as to make use of the overcapacity (regarding both loading and sorting capacity).

Here the entire system can be affected. However, the approach still lacks systematic and dynamic planning and control.

(3) The third example is more a strategy for coming to terms with flow imbalances between regions and origin-destination (O/D) pairs. The basic idea here is to generate new flows from old destinations, e.g., Lidel[9] transporting dry waste in regular inbound distribution trucks.

Another common difference between FTN and the examples presented is the fact that the excess capacity is undervalued in the examples, whereas in the FTN this is not meant for it to inherently be the case. From a shippers'

---

[6] Volvo Cars utilizes the overcapacity of the Ro-Ro vessels for deliveries to markets overseas. The cars are temporarily stored adjacent to the port terminal in question.

[7] Michelin Tires are temporarily stored at the Bäckebol terminal and are sent across Sweden using the overcapacity of Schneker's long-haul trucks.

[8] When ordering at the online bookstore Adlibris, one gets to select a normal delivery (1-2 days) that will require a fee or to allow the shipment to arrive within 2-5 days for no charge. Naturally it is the service of the transportation provider that is interesting here, and not the pricing strategy of the online bookstore.

[9] The German grocery chain experimented with this approach before public outcry, following publicity about the scheme in the press, forced them to abandon the program.

point of view, regardless of which network layer is utilized for the production of the service, the service quality is the same. The same cannot be said about the examples above.

Looking at the air transportation networks, similarities between FTN and passenger/freight transportation can be identified. For passengers, the basic idea of FTN is implemented in all large networks with one crucial difference. In a network of air transportation for passengers, the detailed capacity planning is not a fundamental part of the equation, as passengers opt for different routes according to availability and/or price parameters at the time of booking/purchase (Dobson and Lederer, 1993). This means that the matching operation of capacity and demand on an operational level is done by the passenger. The lack of this property in a freight transport network requires a different kind of capacity allocation control from the network operator than in the case of a user-attracting system such as a passenger network (O'Kelly, 1998). Also, the cargo in the belly hold of an airliner is accepted on grounds similar to that of a stand-by/priority differentiation with almost exactly the same effect (Acharajee, 2000).

Besides the difference in the level of planning (ad hoc vs. systematic and dynamic), the level of attention (system-wide vs. specific relations/ units/origins/destinations) and pricing and quality of the service (discounted vs. regular and standard vs. prolonged lead time) the FTN is distinguished from these related concepts in that it aims to foliate two different network structures in the same network.

## 2.2 Transportation network performance

There are many ways to measure the performance of a transportation network (Chow et al., 1994; Chan, 2003; Shepherd and Günter, 2005; Shaw et al., 2010). In essence, performance measurement is an analysis of both efficiency and effectiveness in accomplishing a given task (Mentzer and Konrad, 1991; Caplice and Sheffi, 1994; Ploos van Amstel and D'hert, 1996). Caplice and Sheffi (1994) categorize the dimensions of logistics performance into: utilization, productivity and effectiveness. All three dimensions are operationalizable and measurable as ratios of input and/or outputs from given processes (Figure 2-8).

Utilization captures input usage and is operationalized as a ratio of actual and nominal input, e.g., a ratio of utilized and available capacity. Productivity is measure of transformational efficiency and is measured as a ratio of actual output and input, e.g., a ratio of transport and traffic work or transport work and fuel usage, etc. Finally, effectiveness is measure of the quality of the service produced and is measured as a ratio of actual and nominal output, e.g., a ratio of on-time and total shipments (Table 2-1). The efficiency of a network is a compound construct that regards both utilization and productivity.

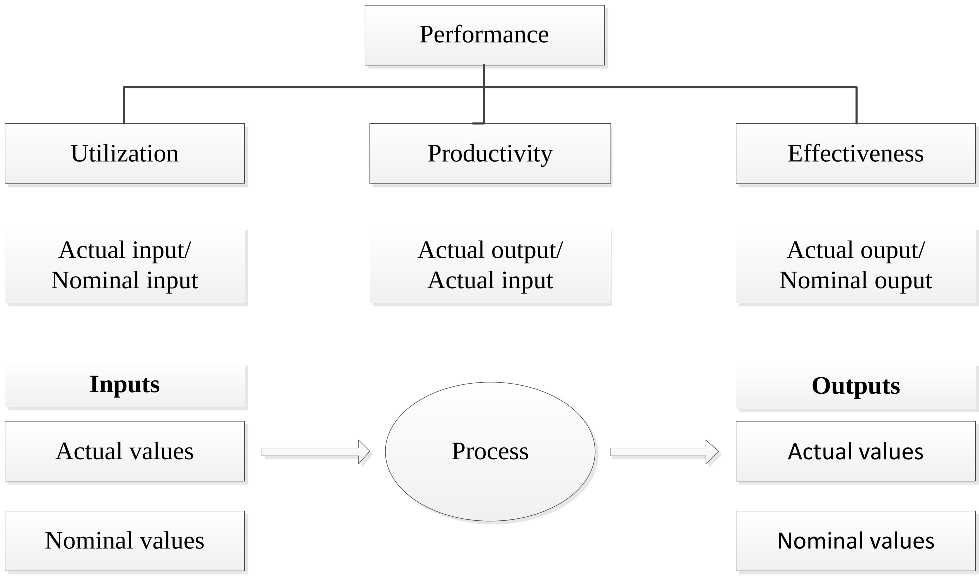


*Figure 2-8 Model of the construct of performance. Redrawn from Caplice and Sheffi (1994)*

Ploos van Amstel and D'hert (1996) present a similar division with input, process and output performance being analogous to utilization, productivity and effectiveness of Caplice and Sheffie (1994). The major difference between the two frameworks is that the former focuses on logistics functions whereas the latter has an explicit process focus. This property contributes to making the framework of Caplice and Sheffie (1994) more adaptable to a pure transportation context as opposed to one based on logistics functions.

*Table 2-1 Dimensions of performance (Caplice and Sheffi, 1994)*

| Dimension | Metric form | Description |
|---|---|---|
| **Utilization** | Actual input/Nominal input | Measures "input usage," e.g., utilized capacity/available capacity |
| **Productivity** | Actual output/Actual input | Measures transformational efficiency, e.g., transport work/traffic work |
| **Effectiveness** | Actual output/Nominal output | Measures quality of process output, e.g., shipments on time/total shipments |

The major difference between the two frameworks is that the former focuses on logistics functions whereas the latter has an explicit process focus. This property contributes to making the framework of Caplice and Sheffie (1994) more adaptable to a pure transportation context as opposed to one based on logistics functions.

There are also frameworks that are developed for the purpose of measuring green or environmental supply chain or logistics performance, e.g., Shaw et al. (2010), Björklund et al. (2012), etc. However, these frameworks are either too broad, i.e., encompassing the entire supply chain, or not applicable to the purpose of this thesis. The focus here is on the transportation network and the relevant environmental impacts that can be measured by the performance dimensions presented above. This is possible if the frameworks are applied correctly to this setting and the measured entities are based on physical properties rather than financial ones (McIntyre et al., 1998).

Working within this framework and in trying to apply it to the context at hand, the literature on performance was surveyed in an attempt to find appropriate measures for evaluating the performance of the transportation network. Very few papers explicitly reported on transportation network performance. The majority of the literature surveyed views transportation in the context of logistics or supply chain management. In many of those cases, transportation was viewed as a function as opposed to a process, the perspective of the shipper was adopted and the main unit of analysis appeared to be cost or some other transferable aggregate, e.g., Mentzer and Conrad (1991), Clark and Gourdin (1991), Kleinsorge et al. (1989; 1991), Stainer (1997), Ploos van Amstel and D'hert (1996), etc. Also, in many cases, the concept of performance did not include all three dimensions identified by Caplice and Sheffie (1994). Van Donselaar et al. (1998) present one of the most comprehensive frameworks that explicitly approaches the subject with the perspective of the carrier and focuses on transportation and distribution. However, here also the unit of analysis is expressed in terms of cost and revenue rather than physical units.

### *2.2.1 Quality of measurements and Key Performance Indicators*

Caplice and Sheffi (1994; 1995) argue that individual performance measures need to be evaluated based on a number of criteria to ensure that they are of sufficient quality (Table 2-2). However, the compound nature of transport performance, particularly when viewed exclusively from the vantage point of physical resources, might create a need for combining several measures to achieve sufficient measurement construct quality. This property is not lost on the originators of this scale, as they themselves identify clear trade-offs between quality criteria.

Increasing the validity of a chosen measure will almost always come at the cost of decreased robustness. This means that the more customized a measure is, i.e., the more apt to capture the specifics of the object of study, the less comparable the measure will become because of that. On the other hand, the more integrative a KPI is, i.e., the more it promotes coordination between different firm functions and processes, the less useful it will become as it probably will become too general or aggregated.

*Table 2-2 Quality criteria for performance measures (Caplice and Sheffi, 1994)*

| Criterion | Description |
|---|---|
| **Validity** | The metric accurately captures the events and activities being measured and controls for any exogenous factors. |
| **Robustness** | The metric is interpreted similarly by the users, is comparable across time, location and organizations and is repeatable. |
| **Usefulness** | The metric is readily understandable by the decision maker and provides a guide for action to be taken. |
| **Integration** | The metric includes all relevant aspects of the process and promotes coordination across functions and divisions. |
| **Economy** | The benefits of using the metric outweigh the costs of data collection, analysis and reporting. |
| **Compatibility** | The metric is compatible with the existing information, material and cash flows and systems in the organization. |
| **Level of detail** | The measure provides a sufficient level of granularity or aggregation for the user. |
| **Behavioral soundness** | The metric minimizes the incentives for counterproductive acts or game-playing and is presented in a useful form. |

Some of the quality criteria are not of concern in the context of this thesis, as the criteria cited are developed for the purpose of application in logistics firms as opposed to scientific inquiry regarding transportation networks. For instance, economy and compatibility are of peripheral interest here as the purpose is not primarily to devise an efficient performance measurement system; rather, it is a tool necessary for evaluating the object of this study, FTN. Similarly, the trade-off between integration and usefulness is easily handled in this context as the study is preoccupied with a specific process rather than the entire supply chain.

In this thesis, the KPI selected need to be valid, useful, of a correct level of detail, robust and promote behavioral soundness. Even if individual measures do not all reach the sufficient level of quality with regard to all the criteria, the combined battery of measures/KPI together need to do so.

### 2.2.2 A construct of transportation performance KPI

Mackinon and Ge (2004) list three KPI for measuring the performance of transportation operations: vehicle loading, empty running and fuel efficiency. The selection of these KPI was predicated on specific circumstances; i.e., cost and commercial performance were excluded as the study focused solely on the transport function rather than on the entire logistics chain. McKinnon (2010) surveys the prevailing measures for the indication of freight transport efficiency. Efficiency is accordingly expressed in loading factors (i.e., utilization: ratio of actual and nominal input) based on units such as weight, volume, deck-area coverage and tonne-km or the level of empty running (i.e., productivity: ratio of actual output and input), none of which is a very good measure for efficiency if used in isolation. This is due to both issues regarding how to measure and the construct of the measures themselves.

In the case of empty running, for instance, the issue of what constitutes an empty unit is difficult to answer without the measure becoming arbitrary or less than valid, robust and useful. Examining the loading factor constructs, it becomes clear that measure can become misleading regarding what it is aimed to assess, i.e., transportation performance. Taken in isolation, load factor as a transportation network performance measure fails the criteria of validity, usefulness and behavioral soundness.

For instance, some systems such as waste management, farming, mining and forestry transports are by default not able to achieve a higher loading factor than 50% when measured on the round-trip. Moreover, the efficiency of transportation is heavily dependent on economies of scale. Using a loading factor measure in isolation would erroneously indicate a suboptimal system consisting of a fleet of smaller units with higher utilization rate as more efficient than a comparable one with sufficiently large units with lower rates of utilization. This flaw in the construct can be partially remedied if the loading factor metrics are used in combination with other complementing indicators such as the number of units, traffic work and/or transport work.

Additionally, as Nanos-Pino et al. (2005) point out, the matter of measuring efficiency is further complicated by the fact that optimum resource utilization from a business perspective does not always perfectly overlap with the optimum physical resource utilization rate. This complication maybe at least partially circumventable if one would regard the physical KPI that drive cost and revenue separately, such as transport work and traffic work that drive revenue and cost, respectively.

When it comes to transport work, another crucial distinction is in order, namely that of the perspective of the shipper and the carrier. The transport work of the network can be divided into two types: internal and external. The external transport work is the total number of tonne-km that is demanded by the shippers in order to satisfy their transport needs, i.e., the product of the goods volume and the O/D matrix (Equation 2-3). The internal transport work is the actual number of tonne-km that is produced by the carrier in fulfilling the transport assignments (Equation 2-4). This distinction is readily apparent in the case of a pure HS setup where almost all of the consignments (barring the ones that have the hub location as their final destination) will lead to higher internal transport work than what the shipper is demanding and subsequently is willing to pay for.

In the equations below, *n* represents all possible relations in the transportation network. $D_i$ is the nominal distance for relation *i*, and $d_i$ is the actual distance transported. $q_i$ is the transported quantity in the relation *i*. $C_i$ represents the capacity that is used in each relation *i*.

$n =$ Number of relations in network
$D_i =$ Nominal distance for relation $i$
$d_i =$ Actual distance transported in relation $i$
$q_i =$ Transported quantity in relation $i$
$C_i =$ Vehicle capacity used in relation $i$

$$\text{Transport work}_{\text{External}} = \sum_{i=1}^{n} D_i q_i \quad [\text{tonnekm}]$$ *Equation 2-3*

$$\text{Transport work}_{\text{Internal}} = \sum_{i=1}^{n} d_i q_i \quad [\text{tonnekm}]$$ *Equation 2-4*

$$\text{Traffic work} = \sum_{i=1}^{n} d_i C_i \quad [\text{tonnekm}]$$ *Equation 2-5*

$$\text{Transport efficiency} = \frac{\sum_{i=1}^{n} D_i q_i}{\sum_{i=1}^{n} d_i C_i}$$ *Equation 2-6*

The traffic work is the total amount of vehicle capacity kilometers that is produced in the network during a period of time which is also measured in tonne-km (Equation 2-5). The quota between external transport work and traffic work is a suggested representation of transportation performance for a given network and a given period in time or analogously a given demand (Equation 2-6). The traffic work is related to the external transport work because that is the main purpose of the transport, to move goods according to a predetermined O/D matrix.

The transport efficiency (Equation 2-6) is thus defined as the quota between the external (nominal) transport work (Equation 2-3) and the performed traffic work (Equation 2-5), for a specific time period (i.e., demand). This transport efficiency construct is subject to the same flaw (risk for sub-optimization) as other loading factor measures discussed above. However, when comparing two setups of the same system, where the unit size is constant, the flaw in the construct will not become an issue. At the same time, it is a more valid, useful and robust construct than the loading factor in McKinnon (2010) due to the fact that this construct penalizes deviation from the shortest possible route. To illustrate, consider any pure HS network as compared to a DS. Using a straight loading factor KPI in isolation would yield that the HS setup is preferable to the DS one in almost every case, where as it is readily apparent that this cannot be the case.

### *2.2.3 Key performance indicators (KPI)*

The KPI utilized in the different studies are presented below (Table 2-3). In the table, the different KPI are related to the different dimensions of performance as described by Caplice and Sheffie (1994). Some of the measures are not stated as ratios, but rather as actual or nominal inputs or outputs. In those cases, they can be used to measure more than one dimension of performance. They are useful in instances where one of the parameters of a ratio is by default constant and a straight comparison of the parameters can be meaningful.

Aside from the last construct included in the table below (Table 2-3), which, as far as is known to the author, is an original construct, all other KPI appear in previous studies by other authors. There are clear indications in the available literature that the performance of a transportation network is closely linked with its flow characteristics. In many comparable studies reviewed, the description or evaluation of the performance of transportation networks is entirely or partially portrayed in terms of flow parameters (e.g., (Buskhe, 1993; Hultén, 1997; Hultkrantz, 1999; Lumsden et al., 1999; Acharajee, 2000; Crainic, 2002; Liu et al., 2003; Caputo et al., 2005; Persson and Lumsden 2006)).

The number of resources in the system and the average fill rate are recurring as measures of performance in numerous previous works (Acharajee, 2000, Buskhe, 1993, Caputo et al., 2005, Crainic, 2002, Liu et al., 2003, Lumsden et al., 1999). In addition to that, Lumsden et al. (1999) make use of the total traffic work. Caputo et al. (2005), alongside Lumsden et al. (1999) and Persson and Lumsden (2006), also take the total transport work into account when evaluating the performance of a transportation network. Mean time and distance between nodes are employed by, e.g., Crainic (2002).

*Table 2-3 KPI for measuring transportation network performance*

| KPI | Description | Performance |
|---|---|---|
| **Number of resources** | The number of loading units, e.g., trucks, required to fulfill the transportation need of each cycle. | Utilization<br>Productivity |
| **Average fill rate** | The ratio of the total utilized and available capacity of the loading units, e.g., fleet of trucks. | Utilization |
| **Transport work** | Total amount of goods shipped multiplied by the distance traveled. | Productivity<br>Effectiveness |
| **Traffic work** | Loading unit capacity multiplied by the distance traveled. | Utilization<br>Productivity |
| **Mean time between nodes** | The average of the transit time of all relations in the network. | Effectiveness |
| **Mean distance between nodes** | The average of the traveled distance of the routes connecting any two nodes in the system. | Effectiveness |
| **Goods flow per link** | The ratio of goods and the number of links in the network. | Productivity |
| **Transport efficiency**[10] | The ratio of external transport work and actual traffic work. | Productivity |

Finally, the research focuses on network performance with regard to physical units of analysis and deliberately delimits economic ones. The reasoning behind this choice is discussed in section 1.3.

## 2.3 Optimization related to FTN

Knapsack Problems are a mathematical formulation of a general group of maximization problems related to capacity constraints. The name "knapsack" is meant to illustrate the capacity constraint property of this group of problems. Suppose there is a knapsack with fixed capacity that can be filled with any number of items from a group of items with varying size and

[10] This construct was developed jointly with Dr. Per-Olof Arnäs.

utility; solving this knapsack problem would be selecting the items from that group that would together yield the highest utility (Martello and Toth, 1990). This is mathematically formulated as follows: the capacity of the knapsack is (c) and constant; the group of items is a binary vector $x_j$ (j=1, ..., n) where:

$$x_j = \begin{cases} 1 \text{ if the item is selected;} \\ 0 \text{ otherwise} \end{cases}$$

Then, if $p_j$ is a measure of the utility (or profit) of the item j and $w_j$ denotes the size (or weight) of the object, the problem would be to select the items so as to satisfy the constraint:

$$\sum_{j=1}^{n} w_j x_j \leq c$$ *Equation 2-7*

So that the objective function is maximized:

$$\sum_{j=1}^{n} p_j x_j$$ *Equation 2-8*

Even though the symbolism of the name "Knapsack Problem" naturally draws the mind to physical capacity, the problem formulation may apply to a wide range of capacity constraint problems, e.g., investment decisions, scheduling of machine time or packing problems, etc.

A special bounded case of the general 0/1 Knapsack Problem where the weight and profit are equal is called the subset sum problem (Martello and Toth, 1990; Kellerer et al., 2004). The object is then to find a subset of weights whose sum is closest to, without exceeding, the capacity, i.e.:

$$Maximize \sum_{j=1}^{n} w_j x_j$$ *Equation 2-9*

$$Subject\ to \sum_{j=1}^{n} w_j x_j \leq c$$ *Equation 2-10*

$x_j = 0\ or\ 1$, j = 1, ..., n.

A Bin Packing Problem is a combinatorial problem of maximizing the use of a limited discrete resource and is referred to by various different names throughout the literature, e.g., cutting stock or trim loss problem, bin or strip packing problem, nesting problem, etc. (Dyckhoff, 1990). One way of looking at a Bin Packing Problem is a special case of a multiple subset sum problem where the capacity is constant. The aim in this case would be to minimize the number of containers or bins used. The mathematical formulation of the general Bin Packing Problem could be as follows, given that an upper bound (m) number of containers and as many binary variables $y_i$ take the value 0 if container (i) is used and 1 otherwise:

$$Maximize \sum_{i=1}^{m} y_i \qquad \text{Equation 2-11}$$

$$Subject\ to \sum_{j=1}^{n} w_j x_{ij} \leq c(1 - y_i) \quad i = 1, \ldots, m, \qquad \text{Equation 2-12}$$

$$\sum_{i=1}^{m} x_{ij} = 1 \quad j = 1, \ldots, n, \qquad \text{Equation 2-13}$$

$$y_i = 0\ or\ 1, \quad i = 1, \ldots, m,$$

$$x_{ij} = 0\ or\ 1, \quad i = 1, \ldots, m, j = 1, \ldots, n.$$

The packing and cutting problems in one, two or three dimensions have an easy-to-recognize connection to their physical applications (the general problem is not limited to three dimensions). For instance, the one-dimensional cutting problem could regard cutting pipes or logs (Dyckhoff, 1990), the two-dimensional problem could be cutting shapes out of sheets of paper or cloth (Lodi et al., 2002) and the three-dimensional packing problem could be filling a transportation unit or bin (Silvano et al., 2000). These examples are by no means exclusive or exhaustive. The mathematical problem could pertain to any planning or scheduling problem to fit the description above (Scholl et al., 1996; Chantzara and Anagnostou, 2006). As the term *Bin Packing Problem* suggests, one of the primary applications lies within transportation (Dyckhoff, 1990; Gehring et al., 1990; Silvano et al., 2000). This issue is connected to general cargo freight transportation in the sense of minimizing the number of loading units through consignment and loading composition.

Another particular bounded case of the general Knapsack Problem is the so-called *change making problem* where the profit is constant and set to 1 and the capacity is constraint (Martello and Toth, 1990; Kellerer et al., 2004). The change making problem imposes equality as opposed to inequality regarding the utility of the items to select, which is the case in the other instances discussed above. This issue is connected to general cargo freight transportation in the sense of maximizing the number of shipments served by one loading unit.

# 3 Methodology

*The methodology employed to carry out the research is presented in this section. Also, research quality and the match between method choices with regard to the posed research questions and the compatibility of the different methods used are discussed.*

The methodological choices made are intimately dependent on the research questions that the inquiry seeks to answer. In any approach, a number of techniques may be employed in order to answer the questions posed as a result of the purpose of the research. The combination, sequence and application of these techniques constitute the methodology.

The concept of Foliated Transportation Networks (FTN) is created in response to an existing real-world problem. Though its constituting parts are individually found in real-world systems, a deliberately designed and systematically implemented FTN does not exist to be studied. Therefore, the study of FTN cannot be complete exclusively utilizing the logic of natural science. The heart of the study is concerned with a desired "out-come" that aims to be achieved by problem solving through design. Thus, the logic of the inquiry will have to deviate from that of pure natural science and also include the science of the artificial (Simon, 1996). What is sought is not limited to the mere description and explanation of how existing systems function but also the normative knowledge of how the network "ought" to be designed to improve the completion of certain goals (Simon, 1996; Denyer et al., 2008). The need for and the application of this type of approach appears to be an increasing trend in management science and engineering (van Aken, 2004).

The purpose of the thesis has been decomposed into two different research questions. Each research question is sought to be answered through a separate study (Figure 3-1). Both studies individually and combined contribute to answering the research questions.

Aside from presenting the methodology used in each study, this section also sets out to present and motivate the overall research design with regard to the purpose of the thesis and the research questions posed to be answered. The methodology employed in individual studies can be likened to pieces of a puzzle that must fit together and make sense when viewed in light of the purpose of the thesis as a whole. Here the argument is made that there exists an acceptable match between the RQs and the chosen methodology employed to answer them; also, that the methodologies of all the studies in combination match the purpose of the thesis reasonably well.

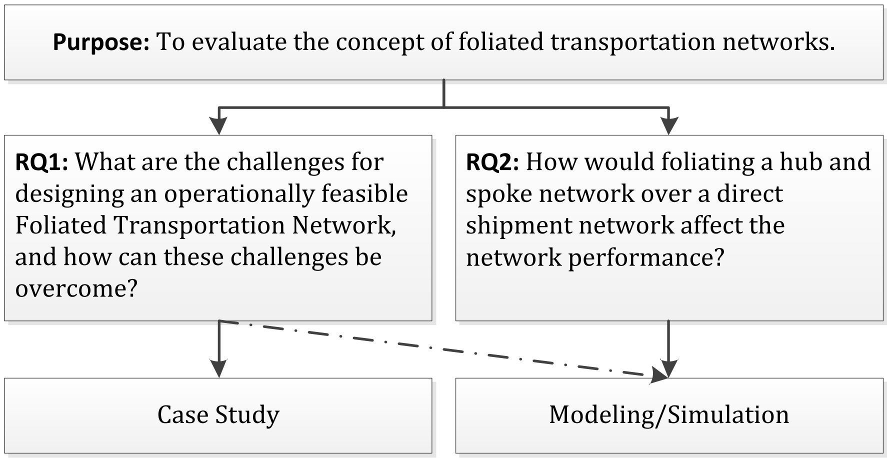


*Figure 3-1 The relationship between research questions and studies*

The studies presented here have resulted in five papers, and each research question draws its answer from more than one paper. With that in mind, this section should gain additional strength if the methods were more clearly linked to each RQ posed.

## 3.1 Study 1 – Case study

The first study is meant to primarily contribute to answering the first RQ which reads: "*What are the challenges for designing an operationally feasible Foliated Transportation Network, and how can these challenges be overcome?"* This question occupies the gap between the conceptual model of FTN and the existing real-world systems. The specific case of FTN studied in this thesis is that of an HS network foliated on a DS network. The first study sets out to find out what it would take to go from an existing DS network to an FTN. Simply put, given the overall task to figure out how to get somewhere, finding out the discrepancy between where you stand and where you want to go as well as a probable trajectories is a reasonable first step.

Decomposing this RQ, several necessary key tasks become clear. The present state of affairs, both in the real systems as well as in the literature, needs to be described. The literature regarding traditional transportation networks as well as mixed model transportation networks in general and FTN in particular need to be examined. The description of the existing networks needs to be juxtaposed to the conceptual goal state of FTN in order to identify the challenges for designing an operationally feasible FTN. Given the challenges identified, the required effort for overcoming them can begin to be formulated.

The characteristics of this inquiry require the collection of qualitative data and methods of analysis. The use of detailed, rich data is necessary for achieving what is sought here; identifying the challenges for changing an existing design to a new one. The description of the existing systems could be based on a case study as long as the selected case meets what Yin (2003) calls the representative rationale. It is important to note that the case study differs from its traditional sense as the case here is not studied to gain insight about the phenomenon or primary object of study, i.e., FTN, per se, but rather to enable the researcher to identify design gaps when the real-world system is compared to the FTN model.

### *3.1.1 Method*

The method employed in the first study, which is exploratory in nature, consists of three parts that have been conducted iteratively; literature study (Hart, 1998), case study (Eisenhardt, 1989; Yin, 2003) and workshop and seminars (Flick, 2006). The literature study and case study are used for describing the existing network and ultimately identifying the challenges for designing an operationally feasible FTN. The workshops and seminars are used for assessing the quality of the findings as well as refining them until saturation is reached. The literature study is performed within the areas of logistics and transportation, information science, mathematics and transportation planning and control. The interplay between the empirical data and theoretical input is consistent with a systematic combining approach (Dubois and Gadde, 2002).

The choice and design of the process used for identifying and analyzing the challenges for designing an operationally feasible FTN is affected by a number of specific conditions:

- no real-world implementation of the concept available
- the inherent complexity of the studied phenomena (transport networks) (Waidringer, 2001; O´Connor, 2009)
- sparse literature on the concept (Persson and Lumsden 2006)
- the interdisciplinary nature of the concept, e.g., logistics and transportation, mathematics, operations research and informatics (Persson and Waidringer, 2006).

These factors led to an iterative process consisting of the three interrelated components: theory (literature study), empirical area (case study) and identified gaps (workshop/seminars). This approach is similar to the so-called “whirlpool approach,” which has been successfully applied in areas such as computer science and information systems (Travisano, 1996; Williams, 1996).

Based on the case study of a representative DS network and conceptual model of FTN, a number of challenges were initially identified. Following this, a literature study was conducted in order to assess if the identified challenges were addressed in previous research, and if so, how these challenges could be overcome in the particular contexts. The results from the first two parts of the study were presented and discussed in seminars and workshops. During this process, additional need for empirical evidence could be identified, prompting the need for further iteration of the previous steps. Several iterations were carried out before saturation was reached.

### *3.1.2 Data*

The Swedish domestic general cargo freight transportation market is dominated by two large companies, each covering the entire country through their own network (Sommar and Woxenius, 2007). This considerably narrows the choice of where to collect empirical data, seeing how FTN is applicable only for large networks carrying an abundance of cargo. In the choice of which company's transport network to use as an empirical starting point, the most important factor was that the studied network should meet the representative rationale (Yin, 2003). The objective of the representative case is to capture the circumstances and conditions of an everyday or commonplace situation. According to Yin (2003): "The case study may represent a typical project among many different projects, a manufacturing firm believed to be typical of many other manufacturing firms in the same industry, a typical urban neighborhood, or a representative school, as examples. The lessons learned from these cases are assumed to be informative about the experiences of the average person or institution." Choosing the network of any of the companies would likely satisfy the representativeness requirement. However, the actual company chosen offers three distinct benefits: access, complexity and write-ups available from previous studies performed by others (Sjöstedt, 2005; Stefansson, 2006).

The studied national network, belonging to a global logistics and transportation provider, consists of a network of more than 30 terminals, covering virtually every corner of Sweden. The goods shipped by the provider are divided into two groups: general cargo, which is handled through terminal(s) and is of specific interest for this study of FTN, and direct goods, which consist of large shipments that do not pass through any terminal. Routinely, consignments totaling a combined equivalent weight of between 30 and 1000 kg are considered general cargo, regardless of their shape and packaging.[11] The empirical data for this study is collected primarily from the general cargo section of the national network.

A very limited number of the trucks are actually owned and operated by the company, which instead makes use of risk and profit-sharing programs with "independent" haulers. The word independent is within quotation marks because the majority of haulers are in fact not independent in the true sense of the word due to the terms of the contract. Based on mutual concessions, some rights and duties are shared between the company and its contracted haulers, e.g., a hauler may have the exclusive rights of a line and in return share the risk of not getting paid for unutilized capacity. This property has two important implications that affect the complexity of the productions system. The management of the resources in the network is decentralized, and the haulers are allowed to allocate underutilized capacity freely.

The information system is fragmented, and different sub-sections cannot and do not communicate with each other in real-time or automatically. For instance, the capacity and profit-sharing system, the goods information system containing EDI (or physical) waybill information and the billing systems are isolated from one another. Counterintuitively, this is an advantage for this decentralized system as the experienced-based low-level control actually outperforms any existing control system in this highly peculiar, fragmented and low-level autonomous setting. This is actually a positive trait of this company as a case to be selected for the study of FTN because the system is comparable to a network of cooperating haulers with no uniform information system, which would have to be the case on the European continent where no one provider has such a dominating position in the market. This in turn could expand the scope of generalization, as much as anyone can generalize from a single case study, regarding the results.

The empirical data employed in the first study is qualitative and has been collected via semi-structured interviews and observation. Two major terminals have been visited, and personnel from senior management to individual operators have been interviewed. The interviewees have been selected through a process of snowballing (Miles and Huberman, 1994), where each new interview reveals the need, identifies the interviewee and creates access to the next one.

[11] This of course does not apply to goods requiring special attention, i.e., chemicals, provisions, etc., that are bound by law or operational conditions to be treated separately.

Semi-structured data collection interviews have been performed with senior managers, sales representatives, information system officers, terminal managers and operations managers. Floor operatives, planners and individual haulers have also been interviewed. The need for additional data gradually arose during the iterative course of consulting the existing theory, collecting empirical evidence and submitting the thus-far results of the analysis to formal evaluating seminars. This process was repeated until saturation was reached (Lindskog, 2008).

### *3.1.3 Validity and reliability*

The validity and reliability of the results could be ensured through the combination of expert group evaluation (Flick, 2006) in successive iterations in a so-called "whirlpool approach" (see Figure 3-2) adopted from research areas of information and computer science (Travisano, 1996; Williams, 1996). This approach is particularly apt when the study precedes a simulation study (Clark et al., 1986).

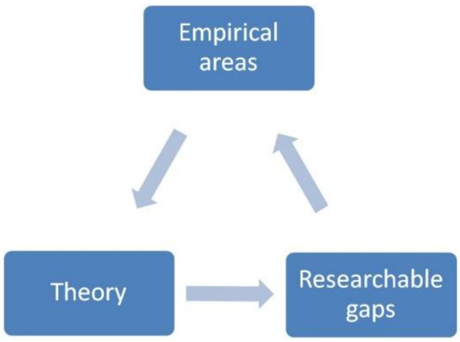


*Figure 3-2 Illustration of the iterative whirlpool approach*

In order to bring additional validity to the results of this study, the identified challenges and their solutions have been discussed, revised and finalized through a series of eight formal seminars. The participants in these seminars have been transportation practitioners and research professionals as well as researchers of other neighboring disciplines such as traffic, logistics and supply chain management. The results have also been put to academic scrutiny, i.e., defended against a senior and a junior opponent, in a review seminar.

The nature of RQ1, i.e., identifying challenges for designing an operationally feasible FTN as well as assessing how these challenges can be overcome, enables the additional testing of the validity of some of the results obtained in the second study, which is modeling and simulation study aimed primarily at answering RQ2. In this context, combining a case study with a simulation study is an appropriate approach as the methods are complementary and contribute to ensuring validity and reliability of the outcome from the respective method (Hellström, 2007). The measures for ensuring overall research quality of all the studies are summarized and discussed under a separate heading (see 3.3).

## 3.2 Study 2 – Modeling and simulation

The second research question reads: "*How would foliating a hub and spoke network over a direct shipment network affect the network performance?*" RQ2 is addressed with a modeling and simulation study. The method of simulation pertains to the act of substituting a real-world system with an abstraction, usually in a computer setting, for the purpose of analyzing the behavior of that real-world system by means of observing the behavior of the model (Banks, 1998). It is a numerical evaluation of a model of a system in order to estimate the true behavior of a real-world system (Law and Kelton, 2000).

Simulation is an appropriate method when the system under investigation cannot be analytically analyzed and experimentation on the real system is not feasible (Goldsman, 2007). It is especially appealing to resort to simulation when alternative system designs or "what-if" scenarios are the object of inquiry (Banks, 1998; Law and Kelton, 2000). With respect to cost and feasibility, simulation can be utilized as a superior substitute to actual on-site experimentation. For instance, a simulation study affords the investigator the ability to compress or expand the real-time of the studied system, allowing the investigator to study longitudinal impacts or detailed workings, respectively, of a system change through the different executions of the model. However, the quality of the results obtained from a simulation study is directly dependent on the model's validity (Sargent, 2004).

The common denominator for all simulation studies are the fact that no analytical solution can be efficiently obtained. There could be many reasons why an analytical solution is out of reach. The real-world system may be so complex that the best alternative analytical solution would be too simplified to cover the aspects that are essential, or of interest, to the inquiry. Exact information may be lacking and unobtainable so that the properties that the interplay between the system sub elements display cannot be analytically described. The real-world system that is the object of the study may even be an entirely new system or a revision of an existing system that is not yet extant and cannot be studied empirically and/or analytically. The input, output and included system sub elements may exhibit a degree of uncertainty or random variation, i.e., be stochastically distributed (Banks, 1998; Law and Kelton, 2000). Carson (2005) amends the condition that the system that is the object of study ought not be "chaotic and out of control." The system should be regularized and its components and their interactions definable.

Regardless of the characteristic or combination thereof that leads to the lack of an acceptable analytical solution, the researcher is basically left with a choice between experimentation on a real system or a simulation. In many cases, simulation may not only prove to be the only viable choice, but also the superior one. Often, experimentation is not feasible either because of the cost or intrusion on the system at hand. In other cases, where experimentation technically would be viable, simulation may still be the superior method to choose because it allows the researcher the opportunity to, e.g., expand or contract the run-time time scale, meaning that simulated system models can be run at higher or lower run-time speeds than real-time experimentation. Along the same lines, a simulation model allows the researcher to, for a negligible marginal cost, test a variety of alternative system setups, revisions or "what-if" scenarios, making the results of the study, which after all are a numerical estimation of the analytically indescribable real-world system, much more robust.

As drawbacks of simulation as a method, Law and Kelton (1991) identify the following. Stochastic simulation is not a very effective tool for optimization because each run of the model is only an estimation of the true behavior of the system and thus several independent runs are required for each system configuration. Also, the risk of the impact of the results of a study overreaching its actual merits is impending because most simulation software, nowadays, is equipped with powerful and credible visualization tools. What the eye sees, the head takes to be true, regardless of the validity or scale of the abstraction from the real system.

Given these study design parameters, the object of study, i.e., the transportation network, needs to be mathematically modeled (Hiller and Lieberman, 1995). Experimentation on real-world systems in order to test and foresee the impacts of new policies or designs is very costly, disruptive and not even always practically possible (Law, 2001b). This is a strong incentive to pursue a modeling approach in general and quantitative modeling in particular that would create the basis for designing a valid simulation model (Hiller and Lieberman, 1995; Hellström, 2007).

Furthermore, results from the simulation model of FTN can be utilized for contributing to the answer for RQ1 particularly regarding the second part of the question. Appropriately designed experiments will reveal not so much how an identified challenge can be overcome operationally but more how large of an impact a specific identified challenge has on the performance of the network. Knowledge about such properties can inform the designer of to what extant challenges need to be addressed for an operationally feasible design to achieve the intended outcome.

### *3.2.1 Method*

The core principle here has been to develop an accurate and valid simulation model in which to run experiments. This effort started off by developing and validating mathematical models of the individual layers of an FTN model. The models are presented in section (4.1) and the validation procedure is discussed below in (3.2.4). The simulation model itself is a discrete event simulation model that has been developed according to the process suggested by Law (2001b). The seven steps are: problem formulation (1), data collection and construction of the conceptual model (2), conceptual model validation (3), simulation model creation (4), simulation model verification (5), experiment design, conduct and analysis (6) and documentation and presentation of the results (7).

As a rule of thumb that would reveals the relative impact of each step on the quality of the end results, Heavy and Ryan (2006) quote the "40-20-40-rule." What this rule of thumb is referring to is that of the total effort spent on a simulation study, 40% should be used for steps 1 through 3 and 5 through 7, respectively, and only 20% on the actual translation of the conceptual system model to an executable computer model. The conceptual model is presented in a separate section (4.2), and the validation and verification effort is presented below in (3.2.4).

Seven distinct variations of the simulation model are developed along the way where each has been tailored for a specific set of experiments. Three of these models are designed for sensitivity analysis. The basic model is illustrated in Figure 3-3.

All other versions are some variation of this basic model. Every run starts with the generation of the demand in the O/D matrix, i.e., the daily volume of goods and amount of shipments in each relation. The goods are then loaded onto trucks and shipped directly until there are not any goods left or the remaining volume is not enough to fill an entire unit. Once this is completed for all relations, the remaining goods are loaded indiscriminately onto trucks and shipped to the hub terminal. There, the goods are sorted and shipped to their final destination. The system key performance indicators are calculated and documented.

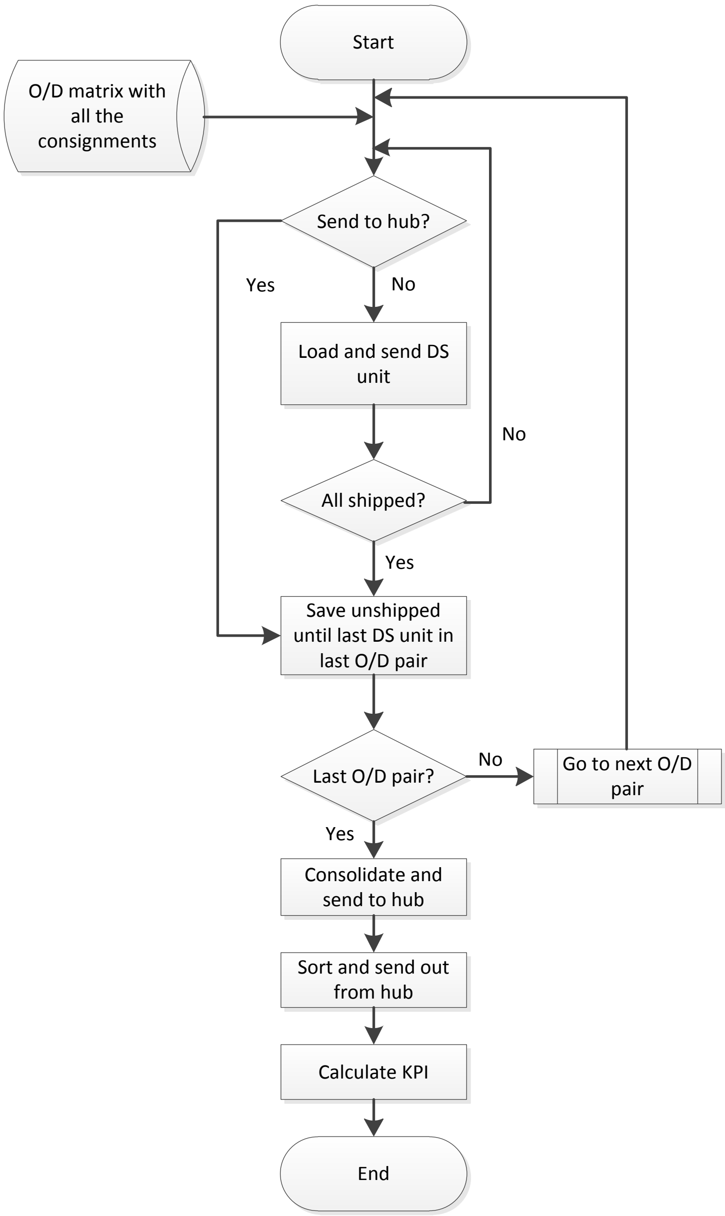


*Figure 3-3 The basic FTN simulation model*

The first variation of this basic model is a reference setup that is pure DS model. Here, the model simply loads the goods in each origin/destination relation on trucks and sends them out directly to their destination terminal. The experiments run in this model are used as a point of comparison. The basic model adheres in its sequence of operation to FTN as described by Bjeljac and Lakobrija (2004), meaning that the hub volumes are identified after the goods going through the DS layer has already been sent. A version of the model was developed to capture FTN as prescribed by Persson and Lumsden (2006), where the hub volumes are identified and sent first, i.e., before goods have been shipped in the DS layer.

Two other variations, one allowing for partial implementation and one for differentiated control, were also developed. The results from the experiment trials have been subjected to relevant statistical analysis. The details of each model and the experiment performed in them can be found in the appended papers.

### 3.2.2 Data

Two separate sets of data from the same reference company make up the empirical element of this study. Both data sets have been data dumps from the ERP system of the company provided to the author upon request. The first set has been used for the validation of the mathematical models of the individual layers of FTN. The second set is used for configuration of the simulation model.

The data sets, which have been verified through comparison with aggregated and historic data obtained from other sources, includes daily freight volumes (payload expressed in kg) for each network relation for a sample of eight consecutive weeks (40 working days). The time period selected is considered representative, as the total flow exhibits signs of high stability in concurrence with available branch statistics and qualitative data. The empirical data has guided the physical network setup (i.e., number and position of terminals), and informed the theoretical distributions for daily goods volumes for each relation.

Considering the domestic Swedish market for general cargo freight, two major operators can be identified. All provide similar services using their respective networks of terminals. Together they fulfill almost all of the demand in that market. The chosen company has the largest share of that market. Even though the business models and production processes of the three major companies differ somewhat in detail, the principal design of their productions systems and demand fulfillment is very similar. For the purpose of this study, key parameters such as number and location of terminals, network coverage, terms of service and distribution of demand in the Origin/Destination matrix are similar enough that choosing any of the two would likely not alter the conclusions that can be drawn. These indications are clear both in terms of official statistics and qualitative data collected from the companies.

The logic behind choosing the reference company is very similar to that of the same choice made above in the first study. Moreover, considering the fact that the empirical data is used to calibrate the simulation of a control principle rather than the current operations of the case company, the requirement for representativeness is met. The choice of the specific company was made partially based on access.

In order to be able to estimate valid theoretical distributions based on the empirical data (eight working weeks of detailed data), the data need to be sequentially independent, i.e., not auto-correlated, homogeneous and modal (Law and Kelton, 2000; Banks et al., 2001). Scatter grams, pivot tables and homogeneity tests have been employed for this purpose. Furthermore, when goodness of fit test is performed with appropriate software, the risk of detecting false positives is sufficiently low in itself (Law, 2001a). None of the tests revealed any reason to doubt the correctness of the data provided (Leemis, 2004). A subset of 15 terminals was selected from the network that consists of less than 30 terminals, yielding an O/D matrix of 210 relations. This means that only O/D pairs that guarantee overnight service are included.

All 210 relations, in accordance with the hypotheses from theory and logic and results of controlling tests, were able to be matched with at least one theoretical distribution. In all but a handful of relations, more than one theoretical distribution was identified as possible fits. In an overwhelming majority of the cases a lognormal distribution was identified as the top three distributions that did fit the data. A representative sample of relations (21 relations or 10% of total), categorized using size and geographical distribution of the origin/destination pairs, were selected for detailed analysis. In more than 70% of the sample a lognormal distribution was assigned a relative rank of 50% or more. Further analysis revealed that a lognormal distribution in 85% of the sample had a p-value above 50% using Andersson-Darling test. These circumstances coupled with the level of abstraction of the model, the intended experiments and the intended purpose of the use of empirical data, i.e., calibration of the general network model, were deemed enough to warrant the use of lognormal distribution for every relation in the model. The parameters for the distribution for each relation were set based on the empirical data regarding each relation.

The empirical data collected from the reference company has been used not to simulate the operations of the specific firm in and of itself but rather to provide representative input parameters for the comparison of two control models, i.e., a direct shipment network and a Foliated Transportation Network (Clark et al., 1986). The drawback of this approach is that validation can become problematic when there are no historical results to compare with the outcome of the model (Clark et al., 1986). This, however, is not an uncommon problem when using simulation given that simulation is often an appropriate tool for the study of and experimentation on systems that are not available for this purpose in the real world. In this case, simulation is the only viable option, and the use of empirical data for calibration of the model is enough to consider the simulation as empirical (Shafer and Smunt, 2004).

### 3.2.3 Implications of model characteristics

There are different kinds of simulations and many different ways to classify simulation models. Kelton et al. (2010) suggest three dichotomous pairs of properties for this purpose. A model can either be static or dynamic, either continuous or discrete and either deterministic or stochastic. According to Kelton et al. (2010) time does not play a natural role in static models but does in dynamic ones. The model at hand is discrete and stochastic but lacks the dynamic time. More often than not, one of the bases for the dynamism of the model is time, e.g., Clark et al. (1986), de Koster et al. (2004) and Denzler et al. (1987).

In the model included in this study, the only dynamic parameter is that of the distribution of demand in the Origin/Destination matrix coupled with the sequence of operations when matching demand to capacity. The sequence of activities or operations is a representation of time in itself, though not in the dynamic sense expressed in, e.g., Kelton et al. (2010) and Law and Kelton (2000). Dynamic time parameters are not included in the model at all. This condition has some implications regarding experiment design, validation and verification of the model.

It means that in this model only the sequences of activities are important. Neither the time it takes to perform any given activity nor the relative time between the activities is of interest. The only time parameter included in the model is static. This does not in itself make the model static. It just means that the source of dynamism is the daily demand in the O/D matrix and the sequence of activities when matching that demand with capacity and not operational time. This modeling approach is applicable due to the fact that the discrete nature of transportation demand and capacity creates dynamism when the demand is non-deterministic or the match is imperfect, two properties that are true for these types of transportation systems.

Also, the only experiment design parameter that can be altered in order to control the confidence interval of the results is that of the number of runs constituting a trial. In this model each run represents a single day of operations (i.e., time parameter that is independent and static). The model has no memory and is reset before every new run. This is the preferred approach due to two different factors. For one, the model terminates as soon as it has satisfied the transport demand fed into it at the start of a run, meaning that "a day's operation" is not defined with regard to time in the model but rather with regard to a day's workload. Secondly, fleet management aspects are not included in the model, which also reduces the need and use for creating a model with memory; i.e., steady state does not apply in this context.

### *3.2.4 Validity and reliability*

The implication of using simulation as research method is that the validity of the results are strongly connected to the validity of the conceptual model with respect to the objectives of the study and the designed experiment's capability to provide a fair estimate of the true behavior of the modeled real-world system (Leemis, 2004; Sargent, 2004; Robinson, 2006). A number of validations techniques are cited in the literature to be implemented where applicable (Banks, 1998; Law and Kelton, 2000; Banks et al., 2001; Sargent, 2004).

Any validation effort becomes more complicated when it regards system setups or policies that are yet not in effect. Banks (1998) outlines eight different validation strategies for a simulation model. Some of the strategies are not applicable because they require testing the results in comparison to historic outcomes or a specific existing real-world system. This limits the validation strategies that are applicable to face validation, sensitivity analysis and validation of conceptual model. Sargent (2004) stresses that the model need not be absolutely valid over the complete domain of its feasible applicability; rather, it is sufficient to establish its validity for the given experimental conditions. To this end, four different validation approaches are presented: conceptual model validity, model verification, operational validity and input data validity (Sargent, 2004).

Face validation is implicitly performed throughout the modeling and execution of the simulation. To this end, some dummy variables and control statistics have also been included in the model to aid the validation and verification via animation and tracing (Sargent, 2004). In addition, a so-called "independent verification and validation" (IV&V) was employed concurrently during the modeling and experimentation process (Sargent, 2004).

Several sensitivity analyses have been performed (for details of the sensitivity analyses see paper 3). The results of the sensitivity analysis do not add any concerns for the validity or reliability of the model. Furthermore, the system input, conceptual model and behavior has been compared to and confirmed by the broad range of qualitative data. Logical inferences with regard to the generative mechanism that would explain the outcome is also used as means to ensure the validity of the model. Furthermore, tracking individual entities post-hoc or using step-through runs to reveal degenerate behavior or diversion from the conceptual model in the computerized model have been employed to ensure the reliability of the computer model.

The validation effort provides ample evidence for the sufficient validity of the model given its intended use and experimental conditions (Sargent, 2004). First and foremost, the outcomes of models are consistently concurrent with theoretical expectations. Secondly, all the outcomes are readily explained, which also strengthens the argument for validity. Thirdly, it could be argued that the high level of abstraction of the models reduces the requirements necessary for ensuring its validity (Simon, 1996). Although it would be possible to model the real-world DS system in an accurate representation with higher resolution of details i.e., lower level of abstraction, it would be of little use, and would actually create more problems than it solves, due to the fact that the FTN model would lack the real-world counterpart to model. The argument is that this level of abstraction is appropriate given the experiments that are set out to perform.

## 3.3 Research quality

To ensure the quality of the research, the researcher must establish two circumstances. The first is that what is measured is actually that which was meant to be measured, i.e., the results are valid (Riege, 2003; Yin, 2003); secondly, if the first property is true that the measurement is performed correctly, i.e., the results are reliable (Riege, 2003; Yin, 2003). Yin (2003) presents four quality aspects of research: construct validity, internal validity, and external validity and reliability. Lincoln and Guba (1984) contend that these measures are poorly apt for testing qualitative research and introduce corresponding quality aspects of confirmability, credibility, transferability and dependability.

In Table 3-1, a summary of the techniques employed to ensure the quality of research are presented. The techniques are divided into two columns where the first column relates to the study based on qualitative empirical data and the second one to the study based on quantitative empirical data. The qualitative column refers to the first study, and the simulation column regards the second. The measures taken for the qualitative study are more general, whereas the measures corresponding to the quantitative one are more specific to the methodology employed, i.e., simulation.

*Table 3-1 Techniques employed to ensure the quality of research*

| Research quality aspect | Qualitative study | Simulation study |
|---|---|---|
| **Construct validity (confirmability)** | Seminars with inform-ants and colleagues (Lincoln and Guba, 1984) | Conceptual model design (Law, 2001b) |
| **Internal validity (credibility)** | Data triangulation (Yin, 2003)<br>Seminars with practi-tioner and research colleagues (Lincoln and Guba, 1984) | Sensitivity analysis<br>Input data analysis (Leemis, 2004) |
| **External validity (transferability)** | Case study database (Yin, 2003) | Conceptual model validation (Sargent, 2004) |
| **Reliability (dependability)** | Formal final seminar with senior opponent (Lincoln and Guba, 1984) | Experiment design (Banks, 1998)<br>Computer model verification (Sargent, 2004) |

# 4 Model Presentation

*In this section, the different models of the transportation network developed are presented. This includes a mathematical model of HS and DS and a simulation model of FTN.*

The Foliated Transportation Network (FTN) model presented in Persson and Lumsden (2006), that has informed the modeling approach, consists of two network layers: a direct shipment and a hub and spoke network layer (Figure 4-1). The physical network is the same, and the layers referred to are abstract in the sense that they are the product of the principles by which the goods are routed through the physical network. In order to be able to evaluate the performance improvement potential of FTN as compared with its constituting network layers in isolation, FTN and its comprising sub layers need to be modeled in a way that allows quantification of the network performance.

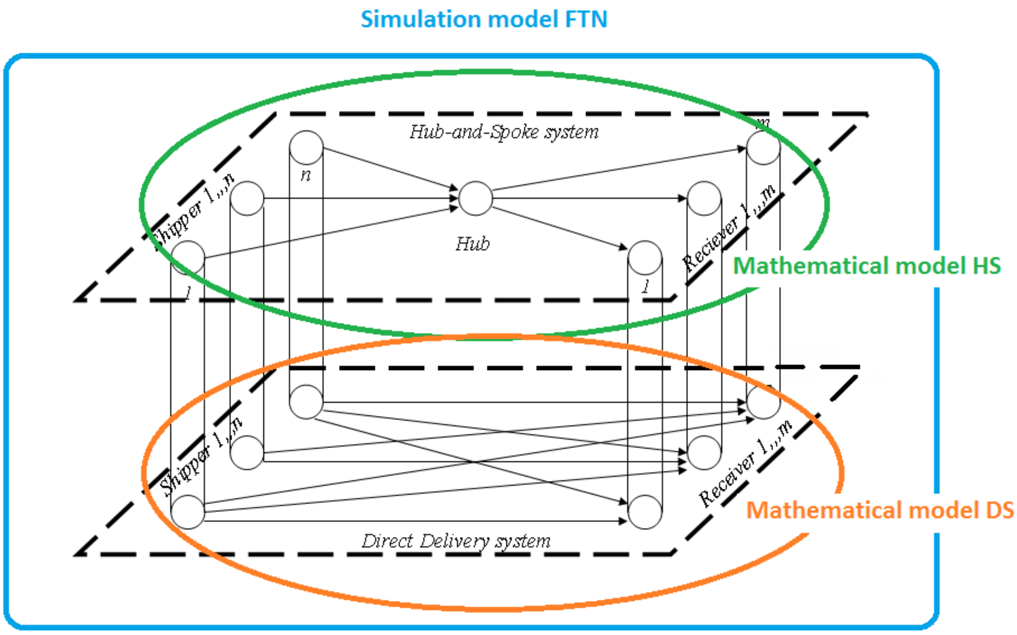


*Figure 4-1 Illustration of modeling approach*

To this end, a mathematical model has been developed. The mathematical model primarily aims to facilitate the quantification of the performance of the sub layers. Based on the mathematical model, a simulation model has been developed that would allow the quantification of the performance improvement potential of FTN based on empirical data. In the following sections both of these modeling approaches are presented in detail.

## 4.1 Mathematical model

Throughout this research, a general mathematical model has been developed and used in order to allow for the quantification of capacity requirements of the network both regarding loading capacity and transit time. This model serves as a basis on which the performance of DS, HS and FTN setups has been evaluated.

### *4.1.1 Capacity utilization*

The total number of resources (U), i.e., trucks or trailers, is in the DS case calculated according to Equation 4-1.

$$U_{DS} = \sum_{i=1}^{n}\sum_{j=1}^{n} int\left({}^{q_{ij}}/_{C}\right) \qquad \textit{Equation 4-1}$$

In Equation 4-1, (n) denotes the number of nodes, ($q_{ij}$) is the amount of goods to be shipped from node (i) to node (j) and (C) represents the maximum capacity of the physical carrier. The operator (int) simply stands for the result of the operation inside the parentheses rounded up to the nearest integer, e.g., 5.1 = 6. The basis for this model is that the total number of resources necessary in the network is equal to the sum of the minimum number of trucks/trailers that satisfy the transportation capacity need of each relation. This approach assumes “overnight deliveries,” i.e., that each unit can be utilized only once during each cycle. This assumption is applied for all parts of the model as a whole.

The same model for the HS structures is not as straightforward, mainly due to the fact that all relations do not make use of just one link, i.e., all goods in a relation first are shipped in the intra-hub links toward the hub and then, if necessary, in the inter-hub links and/or then from the hub to the destination terminal. Moreover, the hub-satellite links are utilized in a single direction during each sequence, i.e., all the goods are first shipped from the satellite to the hub and then in the reverse direction. However, the inter-hub links are used in both directions simultaneously.

These inherent routing characteristics of the HS network make the balance of intra-hub and inter-hub resource flows within each cycle an important factor in modeling the number of resources necessary. Before arriving at an expression describing this dependence, the foundation of the model needs to be established as follows:

The number of nodes, (n) = j

$$n = \{n_1, n_2, n_3, \dots, n_j\}$$

where, j ≥ 3

The number of hubs, (h) = r

$$h = \{h_1, h_2, \dots, h_r\}$$

where, r ≥ 1

Also:

$$n_j = h_r \text{ and } n_3 \leq h_1$$

$$n_k \in h_1 = \{n_1, n_2, \dots, n_{h_1-1}, n_{h_1} = h_1\}$$

$$n_k \in h_2 = \{n_{h_1+1}, n_{h_1+2}, \dots, n_{h_2-1}, n_{h_2} = h_2\}$$

$$\vdots$$

$$n_k \in h_r = \{n_{h_{r-1}+1}, n_{h_{r-1}+2}, \dots, n_{h_r-1}, n_{h_r} = h_r\}$$

The consolidation process at the departing terminals leads to Equation 4-2 for ($\Sigma H_I$), which symbolizes the total number of intra-hub trucks arriving at the hub ($H_I$):

$$\sum H_I = \sum_{k=\min\{n \in H_I\}}^{n \in H_I} int\left(\sum_{j=1}^{n}\left({q_{kj}}/{C}\right)\right) \qquad \textit{Equation 4-2}$$

Here, ($n_j$) denotes all the nodes in the network and ($n_k$) represents all the nodes within the scope of the hub ($H_I$). Moreover, ($q_{kj}$) is the amount of goods destined from node ($n_k$) to node ($n_j$). The expression captures the fact that all shipments from any satellite terminal, regardless of their final destination, are initially transported to the hub to which the originating terminal is directly connected. Accordingly, all intra-hub traffic from ($H_I$) to each of its satellites is equal to the sum of all the goods destined for that satellite, irrespective of the shipment origin in the network.

To be able to model the entire inbound and outbound traffic at each hub, the inter-hub traffic also needs to be expressed. $\sum I_{HI}$ and $\sum U_{HI}$ refer to the total number of trucks arriving at and departing from the hub ($H_I$).

Equation 4-3 and Equation 4-4 model the fact that the sum of all traffic to and from any hub is equal to the sum of intra-hub and inter-hub traffic to and from that hub. The total number of inter-hub trucks arriving at ($H_I$) and the total number of inter-hub trucks leaving from ($H_I$) are displayed in the last half of the right side of the expressions below, respectively.

According to the same logic as in the case of intra-hub traffic, i.e., all shipments to and from all satellite terminals pass through each respective hub(s), the sum of all inter-hub traffic is expressed. However, the aggregation of flows at the departing hubs needs to be accounted for. This is obtained by first accumulating the total amount of goods that are to be shipped between two regions at the departing hub, before the minimum number of trucks required is calculated.

$$\sum I_{H_I} = \left(\sum_{k=\min\{n\in H_I\}}^{n\in H_I} int\left(\sum_{j=1}^{n}\left(\frac{q_{kj}}{C}\right)\right)\right) + \sum_{\forall h_r \neq H_I} int\left(\left(\sum_{n_j\in h_r}\sum_{n_k\in H_I} q_{jk}\right)/C\right)$$

*Equation 4-3*

$$\sum U_{H_I} = \left(\sum_{k=\min\{n\in H_I\}}^{n\in H_I} int\left(\sum_{j=1}^{n}\left(\frac{q_{jk}}{C}\right)\right)\right) + \sum_{\forall h_r \neq H_I} int\left(\left(\sum_{n_j\in h_r}\sum_{n_k\in H_I} q_{kj}\right)/C\right)$$

*Equation 4-4*

With this framework set, the minimum total number of resources, i.e., trucks and/or trailers that are required in the HS case to satisfy the transportation capacity needs ($U_{HS}$), is modeled as follows:

$$U_{HS} = \sum_{\forall h}\left(\left|\sum I_{H_I} - \sum U_{H_I}\right| \times \delta\left(-1\left(\frac{\left|\sum I_{H_I} - \sum U_{H_I}\right|}{\left(\sum I_{H_I} - \sum U_{H_I}\right)}\right)\right) + \sum H_I\right)$$

*Equation 4-5*

Equation 4-5 is based on the balance of resources flowing in and out of one hub. If the total number of outbound trucks is fewer than or equal to the total number of trucks arriving at the hub, then the minimum quantity of trucks necessary to satisfy the capacity needs at that hub is equal to the sum of the number of trucks heading from the satellite terminals of that hub, to that hub. However, if the imbalance is reversed, additional units equal to the difference of incoming and outgoing traffic is required. This imbalance, in cases where it occurs, is explained by the imbalances in goods flows and consolidations, splitting up and reconsolidation effects at the different nodes of the network.

### *4.1.2 Network throughput time*

The minimum time required to travel between any two nodes in the DS network is set to the ($D_{ij}/t_{\mu}$), where (D) is the distance between node (i) and node (j) and ($t_{\mu}$) is the mean speed of the fleet of trucks through the network. Even though a timetable for departures exists, in theory, the only constraints in this network are the scheduled time of arrival and the time it takes to drive the distance. This definition makes the calculation of transportation time between two nodes in the system trivial and the mean transportation time between all the nodes in the entire network as portrayed in Equation 4-6. Here, (R) is the number of relations in the network.

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{D_{ij}}{t_{\mu}}\right)\Big/R \qquad \text{Equation 4-6}$$

The models describing the transportation time between two nodes and the times of departure from every node and hub in the HS network is a bit more complicated. Also, the level of complexity grows with the number of hubs in the system. This difficulty is a result of the fact that the arrival of the physical shipments, both intra- and inter-hub, at each hub, irrespective of the final destination, needs to coincide, creating limited windows of opportunity for the agglomeration operation. Accordingly, the time schedule needs to be set with respect to both the inter- and intra-hub constraining connections, i.e., the connections where the time required for transport is the greatest.

Another defining cutoff value is the predefined system time of delivery. All shipments are to have reached their final terminal by a predefined time. These conditions are formulated in the following equations for the HS setups. Equation 4-7 expresses the minimum transportation time between two nodes, and Equation 4-8 establishes the time of last departure from every node:

$$t_{n_i n_j} = t_{n_i h_j} + t_T + \max\{t_{h_i h_k}\}^{\forall k} + t_T + t_{h_j n_j} \qquad \text{Equation 4-7}$$

$$ToD_{n_i} = T - \max\{t_{h_i n_m}\}^{\forall m \in h_i} - t_T - \max\left\{t_{h_i h_j}\right\} - t_T - t_{n_i h_i}$$

Equation 4-8

(ToD) is the latest time of departure from the node ($n_i$). As all shipments from a node in the HS setup, regardless of destination, are initially sent to the hub, each node needs only one time of departure. (T) represents the latest time of delivery of the system. The different indexes of (t) stand for either the transportation time between the indexed origin-destination pairs, e.g., the left side of the first equation or handling time at terminals, i.e., ($t_T$). The mean transit time between the nodes in the HS setups is then computed in the same way as the DS case, i.e., the sum of the transit times of every node pair divided by the number of relations.

## 4.2 Simulation model

Based on the empirical data and results from the second study, a direct shipment network has been chosen as the point of reference for comparing the performance of FTN. The illustration in Figure 4-2 depicts a smaller network than the one used in the simulation model for reasons of clarity. For the same reasons, only two nodes are used in this example even though all the procedures and properties of those two nodes are valid for all the nodes in the network simulated. The task of keeping this in mind henceforth is left up to the reader.

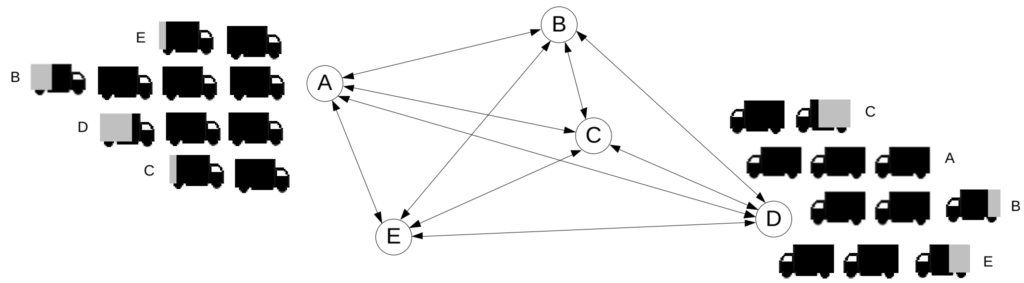


*Figure 4-2 Illustration of a direct shipment network (only two nodes depicted)*

As depicted in the illustration (Figure 4-2), the goods are consolidated in the origin terminals (A) and (D) in dedicated trucks and are set to be shipped the shortest way to the destination terminal. The letter at the back of each row of trucks denotes the destination for all the trucks in that row. The boundary for the study is drawn at the gates of the terminals in the network, i.e., the goods collection and distribution to/from terminals is not included in the model. The use of the term “network” as opposed to “system” ought to have implicitly conveyed this clarification. The gray areas of the trucks illustrate unutilized capacity. Trucks that are all black are hence fully utilized.

The principal idea of FTN is to only send the full trucks directly and consolidate the remaining volumes via hub transshipment. In Figure 4-3 the trucks that are to be shipped in the HS layer of the FTN are boxed in and redrawn as consolidated shipments in the lower half of the illustration. These volumes include every last truck that is not full. These units are intermittently destined for the hub terminal, which in this illustration is terminal (C), for consolidation and further transport to their final destination. This is of course in line with standard operations in an HS network.

Due to this same property, the goods can here be consolidated irrespective of their final destination, which was not the case in a pure DS setup.

This is illustrated in the bottom half of Figure 4-3 where the trucks that are routed via hub (C) carry goods for more than one destination. Trucks leaving (A) for hub (C) in the HS layer carry goods that have (B, C, D and E) as the final destination, and trucks leaving (D) for the hub carry goods with (B, C and E) as the final destination. In the simulation model, this same principle is applied for all the nodes in the network.

An overview of this simplified illustration reveals that the total number of trucks necessary is reduced. The same amount of goods (upper half of Figure 4-3) that required four trucks to be dispatched from (A) and three trucks to be dispatched from (D), by foliating, needs only three and two trucks, respectively, to satisfy the same demand (see lower half of Figure 4-3). This illustration also exposes two challenges, both of which are explored in the simulation study.

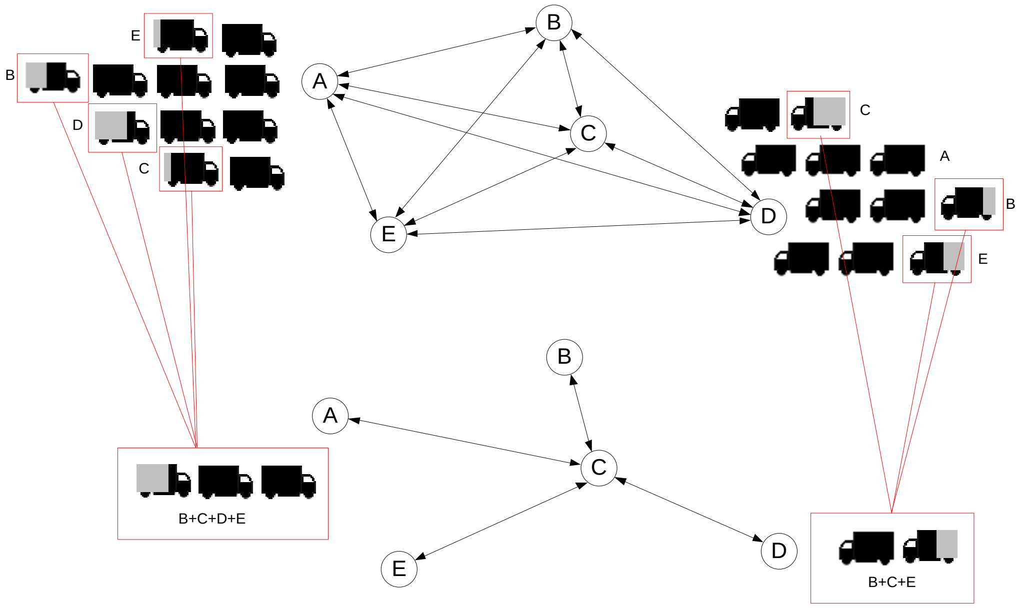


*Figure 4-3 Illustration of a Foliated Transportation Network (only two nodes depicted)*

For one, because the volumes transshipped through the hub will by default require a longer time in transit, it is of interest to be able to identify and send these shipments as early as possible. Doing so may offset the negative impact that the hub detour and additional consolidation/-deconsolidation steps in the hub might have on the network's transit time or mean time between nodes.

Secondly, it is apparent from Figure 4-3 that the cutoff fill rate that is set to consider a truck full is not unproblematic. This limit is dependent on a number of different parameters that need to be explored further. For instance, looking at the last truck in the A-E relation in Figure 4-3, it is not clear that the gain in trying to utilize the excess capacity in that relation is outweighed by the additional distance and handling operations incurred by rerouting it via the hub. Also, the question of the balance of incoming and outgoing volumes to the hub complicates this choice.

# 5 Appended Papers

*This thesis is based on the five appended papers that will be briefly introduced below. The relation between the studies, the papers, the research questions and the purpose is also presented here.*

In order to answer the research questions, two studies have been undertaken. The results from the studies are presented in five papers. These papers are briefly presented in the following section. The relationship between the RQs, studies and papers are presented below in Figure 5-1.

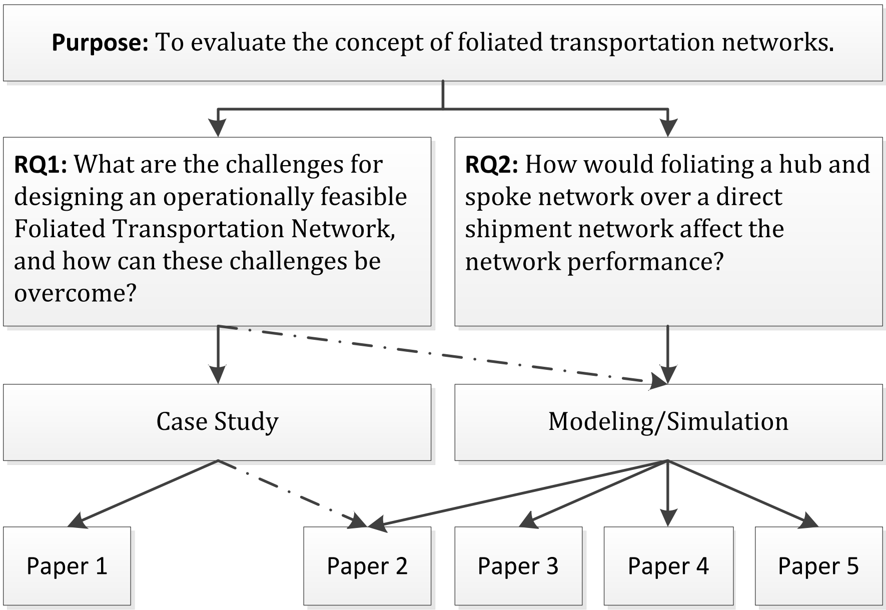


*Figure 5-1 The relationship between RQs, studies and papers*

## 5.1 Paper 1 – Research outlook on a mixed model transportation network

The primary aim of the first paper is to contribute to answering the first research question. The results from the first study are presented in paper 1. Also, papers 2 through 5, presented below, contribute to answering the first research question.

### *5.1.1 Paper outline*

The stated purpose of the study reads in the paper as: "…to present a research road map for developing the concept of FTN to an operational model. The road map contains both an overview of the empirical as well as theoretical gaps that need to be filled in order to establish the concept of FTN."

A qualitative approach composed of three key elements has been adopted in paper 1. These three elements are literature review, case study of an existing system and identification of the design challenges via seminars and workshops. These steps have been repeated in numerous iterations until saturation has been reached.

That which is referred to as "researchable gaps" in the first paper is analogous to what is referred to as "challenges for designing an operationally feasible FTN" in RQ1. Identifying these design challenges is both the primary purpose and the main contribution of this paper to the thesis. The findings of this study inform the remaining design challenges with regard to developing a feasible operational model of FTN.

### *5.1.2 Results and conclusions*

The identified challenges can be broadly categorized as pertaining to transportation planning and control, transportation operations and transportation network optimization (see Table 5-1). The details of the identified design challenges are discussed below.

*Table 5-1 Identified design challenges*

| Design challenge | Specific issues | Theoretical domain |
|---|---|---|
| **Transportation planning and control** | – Governing rule issues<br>– Bin Packing Problem | Transportation management<br>Mathematics |
| **Transportation operations** | – Presorting<br>– Identification | Transportation management<br>Information and communication |
| **Transportation network optimization** | – Change Making Problem<br>– Heterogeneity | Mathematics<br>Information and communication |

FTN is described by Bjeljac and Lakobrija (2004) as the portion of the goods in each relation that does not fill up a full unit, and thus travel through the HS layer of the network is identified and shipped last, i.e., after all the full units in each relation have departed. The result of this design is that the goods that are to travel the farthest total distance and require extra handling, i.e., extra time in transit, will depart last. Persson and Lumsden (2006) handle this issue by proposing that the goods destined for the hub layer of the network be identified and shipped first, thereby affording the most time-consuming portion of the total shipping volume the largest time window. This approach, however, will require a high level of capacity planning detail (Sternberg and Stefansson, 2007; Sternberg, 2008). That detailed level is difficult to achieve for homogeneous goods in the current systems, not to mention the implications of attempting to do so regarding heterogeneous goods.

Exploring this path, the quality and level of detail in the current customer orders needs to be established and compared to what level and quality of order data is absolutely necessary, preferable and/or feasible to expect from the transport buyers. Furthermore, the question of the impact of the need for information and accuracy, and the methods to obtain those, on the flexibility and robustness of the system, deserves attention. This also relates to another question, the level of rigor of the governing rules of the system; e.g., latest time of order entry, tolerance for accommodating last-minute changes, the parameters for decentralized decision making and the ability of the system to cope with deviations from the plan need to be determined.

The heterogeneity of the goods contributes additional complexity to the problem. For instance, the heterogeneity of the physical properties of the goods leads to a setting where the loading composition and loading sequence of the goods within the transportation unit will affect the loading capacity required. This means that even if the physical utilization rate regarding, e.g., volume, weight or length could be measured precisely (which may not even be the case) the reverse relation, i.e., planning for capacity based on aggregated weight, volume or length parameters cannot be taken for granted.

It is not, then, only a matter of information quality/requirements on its own, but one that is complicated with a multi-dimension, multi-choice, multi-constraint Bin Packing Problem. Crassly, it means that given perfect information and absolute ability to simulate the operations necessary, the issue is still unsolved. Mathematically, there is a non-trivial Change Making Problem to be addressed, as well. This Change Making Problem will need qualitative inputs from the refined system design for its specific solution. For illustration, the profit variable in the Change Making Problem could be as easily handled as the number of shipments put through the hub (minimize function) or a compound measure to be defined later in the lines of "handleability," i.e., ease of terminal handling (maximize function).

A seemingly more trivial matter concerning the time aspect is one of reducing the total time in transit for the goods within the HS layer of the

FTN. The extent of the detour that hub transshipment entails can considerably affect the additional time in transit required. In some instances it may be necessary to reduce the handling time. Several approaches have been pointed out in the past, e.g., selecting only easily handled goods for the HS layer (Bjeljac and Lakobrija, 2004), pre-sorting in cages (Acharajee, 2000), use of RFID (Persson, 2006a) or any combination of these and other possible operational solutions.

As mentioned several times before, the goods that travel through the HS layer of the system will not only be in transit longer than the ones in the DS layer, but they will travel a greater distance as well. The point being made is that by sending a slightly less than 100% full truck into the DS layer, the extra distance and handling for that entire almost-full truck will be eliminated compared to its traveling though the HS layer. This then begs the question: What is the filling rate of a "full" truck? Could that be situation-dependent? Or, is there a set estimated value to follow? How substantial is the effect of these decisions on the system-wide performance?

In summary, it is not only a matter of data quality, but it also concerns the resolution of the data (i.e., the level of detail and richness of data) in combination with the specific requirements of heterogeneous goods and the knapsack problem, i.e., the ability to successfully transform the data into a solid plan. The tricky part is that even success in doing so does not clear the fog and automatically lead to an uncomplicated state of affairs. There will be a point of balance between where central planning will enable better resource utilization and where it will inhibit flexibility and robustness. In order to be able to make a business decision about this trade-off, this interrelation needs to be cleared, and, to the greatest extent possible, quantified. The combined effect of the planning errors, i.e., the knapsack problems and the error of the available/feasible information obtained from the consigner, like inadequate or faulty information, last-minute changes, etc. need to be examined.

### *5.1.3 Contribution to the thesis*

The literature review revealed that FTN is well anchored in transportation network theory. What is lacking is partially new theoretical knowledge in mathematics, application of new technology and a detailed applied design for FTN. In the context of the domain of this thesis, FTN research finds itself in the intersection between transportation management, information and communication and mathematics. Three interrelated areas where central design gaps exist have been identified: transportation planning and control, transportation operations and transportation network optimization.

The identified gaps are not only in application of existing technology or lack of necessary but not yet existing technology, but also in knowledge. In the case of the mathematical issues, new previously unsolved non-trivial mathematical problems need to reach a concluding solution for theoretical advancement of this concept. However, as far as developing an applicable FTN, these obstacles may be circumvented by approximate rule-of-thumb or heuristic-based solutions. This notion is further examined in papers 3 through 5.

## 5.2 Paper 2 – The stepwise replacement of direct shipment network with a hub and spoke system

The second paper contributes to the answer of both of the RQs. The model presented in this paper along with the means to measure the transportation network performance is meant to provide the basis for the simulation model developed in the second study. The primary contribution of this paper to the thesis is that of providing a validated mathematical model on which to build the simulation model.

### *5.2.1 Paper outline*

In the paper, the purpose of the study is presented as: "…to model and evaluate the impact of change of network structure from a direct shipment to a hub and spoke system on the performance of the transportation system."

Conclusions drawn from the results of this paper are in part the basis for the hypotheses to be explored in the next one. This paper presents a consequence analysis of the impact of this alteration on the performance of the transportation network. In transportation textbooks and in general network theory this impact is anticipated and described. However, in order to be able to model FTN, one needs to be able to model its constituting parts. This study provides such a model that also is tested and validated using empirical data.

This study utilizes mathematical modeling for evaluating the impacts of the choice of network structure on the KPI and subsequently the performance of the transport network. To achieve the goals of the study, key performance indicators for the different setups, based on the empirical data, have been calculated using the developed models. The mathematical models are developed in accordance with existing transportation network theory and available empirical data (Hiller and Lieberman, 1995).

The study draws empirical data from a representative transportation network, but does not aim to model that specific network per se. Instead, the empirical data is used as input for the mathematical models of the general, ideal typical networks developed in the paper. Hence, even though the modeled flow of goods in the DS network is principally accurate, it differs in detail from the real case. The results obtained serve as indication of the validation of the models developed.

### 5.2.2 Results and conclusions

The results of the consequence analysis offers support for the validity of the models as the results adhere to existing theory. The results are based on comparisons of outcome of a direct shipment network with the different system configurations of a hub and spoke system containing three, two and one hub, respectively. The outcome is expressed with respect to the predetermined KPI. The outcome of this study may be summarized as follows:

- The minimum number of resources, i.e., trucks and trailers required, is directly influenced by the number of links in the network. A reduction in the number of links, with a sustained number of nodes, i.e., the shift from DS to HS, yields a decline in the minimum number of resources required. This relation is enhanced by the number of hubs in the network, i.e., the same number of links and nodes but a higher number of hubs would yield a greater reduction.
- Reducing the number of links in the network increases the flow per link, which improves the average filling rate of the trucks and resource utilization of the system. The number of hubs has little to no effect on this relation.
- Given the sustained number of nodes, a reduction in the number of links in the network causes an increase of the minimum total transport work required, to handle the same amount and composition of load. The number of hubs in the network enhances this relation where an additional hub generates added transport work.
- The rise of the number of hubs in the network increases the mean distance and transit times between the nodes where a significant part of the additional time is due to the need for coordination of ingoing and outgoing flows.

### 5.2.3 Contribution to the thesis

The major contribution of the paper to this thesis is providing a valid representation of the individual layers of FTN that can be transferred to a simulation model. Also, the identification of specific key performance indicators (KPI) for measurement of transport network performance and their operationalization are crucial for the continued effort.

The results concur with existing theory, which in itself could be considered an argument for the validity of the models produced. The mathematical models developed and utilized for this consequence analysis are of greater interest in this context because they actually provide a solid modeling basis for the development of the simulation model.

Finally, the findings regarding the mean time between nodes in the case of the single hub setup indicate that the handling performance issues are relevant to address, even though they are shown to only marginally adversely affect the performance. The implications of these findings are addressed primarily using the results presented in paper 5 but also in paper 4.

## 5.3 Paper 3 – Quantifying the performance improvement potential of Foliated Transportation Networks

The third paper primarily contributes to the answering the second research question. This paper also contributes to the answer of the first RQ. The second RQ also draws from the results from papers 4 and 5 for its answer.

### *5.3.1 Paper outline*

The stated purpose of the study in the paper is: "...to quantify the performance improvement potential of foliated transportation networks (FTN) compared to a traditional direct shipment network (DS) with respect to key performance indicators (KPI) that are identified to express the physical performance of a transportation network." The potential of FTN as compared to a DS is quantified and subjected to sensitivity analysis. Particularly, the impact of the prognosis error and the cutoff value for deciding which layer to send goods through are explored.

In this paper the potential of the principle of FTN has been quantified using a discrete event simulation (DES) model. These results have also been subjected to a sensitivity analysis in order to determine the order of magnitude of the impact of planning and control error as described in the first paper. The ambition has been to model and compare general, ideal-typical networks as opposed to actually modeling a specific existing network.

First, the performance of the two models, i.e., DS and FTN, based on the same set of data has been compared. This comparison is the basis of the analysis to determine the impact of introduction of FTN on the network performance. To ensure that the identified potential is not mistakenly attributed to FTN as opposed to some other factor that is the real reason for the obtained results, a two-level five-factor factorial design has been performed. In this part of the analysis, aside from the change in structure, the size of the trucks in the system, the volatility of demand, the density of the network and planning and control precision have been included.

Moreover, in order to reveal the sensitivity of the outcome regarding the prognosis error, a series of tests with increasing margins of error have been run. The results of these tests are used to confirm the robustness of the FTN with respect to the necessary prognosis procedures as identified by Persson and Lumsden (2006) and also in paper 2.

Finally, the first performance tests have been rerun, with altered fill rate limit from 100% to 75% for trucks to be sent directly. The 75% limit is a judgment based rough estimate and is invoked to identify whether the hypothesized potential due to optimization of governing rules of the system is valid. This step only provides an indication regarding the abovementioned hypothesis as that regards dynamically set levels for each relation, based on a number of factors both local and global in the system.

### *5.3.2 Results and conclusions*

The results referred to are true at a significance level of $p<0.01$ and are quantified at the confidence interval of 99%. The main experiment shows that in the FTN setup as described by Bjeljac and Lakobrija (2004), the system level average fill rates of the trucks were increased by 14.5 % (±0.2), the minimum number of trucks required was reduced by 10.5% (±0.4), the total transport work increased by 5.2% (±0.5) and the traffic work was not affected compared to the DS setup.

However, when the fill rate of the trucks to go directly was reduced to more than 75% instead of 100%, the results were affected. The improvement potential regarding the number of trucks required and the average fill rate of trucks were marginally diminished at the same time as the traffic work was drastically reduced and total transport work was also marginally reduced (see Table 5-2). The convergence of these results with the results presented in paper 1 regarding the research outlook is strongly indicative of still untapped potential that would be a result of the successful design of the governing rules for FTN and an operative optimization effort. The modification above is not dynamically defined; indicating that optimization with respect to each relation, each run or other appropriate criteria could likely yield additional performance benefits. This hypothesis is in part tested in paper 4.

*Table 5-2 Comparison of two different cutoff values for rendering a truck "full"*

| | FTN potential (DS=100%) | FTN potential (DS>75% ) | Difference |
|---|---|---|---|
| **Average fill rate** | +14.5% ± 0.2 | +13.65% ± 0.23 | -0.75% ± 0.11 |
| **Number of trucks** | -10.5% ± 0.4 | -9.60% ± 0.41 | +0.95% ± 0.08 |
| **Transport work** | +5.2% ± 0.5 | +2.57% ± 0.45 | -2.63% ± 0.08 |
| **Traffic work** | No significant difference | -13.62% ± 0.44 | N/A |

As for the planning and control precision presented as a crucial aspect in the implementation of FTN, the sensitivity of the FTN setup regarding those issues has also been investigated. In effect, the model has been modified to adhere to the principles put forth by Persson and Lumsden (2006) for these tests. The FTN system is shown to be fairly robust regarding the effects of the prognosis error where the error needs to reach unrealistic levels of size and variation before the FTN performance is lowered to the same level as the DS setup. This is true for all KPI except transport work where the FTN outcome, in compliance with theory, actually is higher than DS in all cases. This is due to the fact that a portion of the goods, i.e., the goods being shipped through the HS layer of the network, do not travel the shortest way to their destination, resulting in a utilization of the overcapacity in the network.

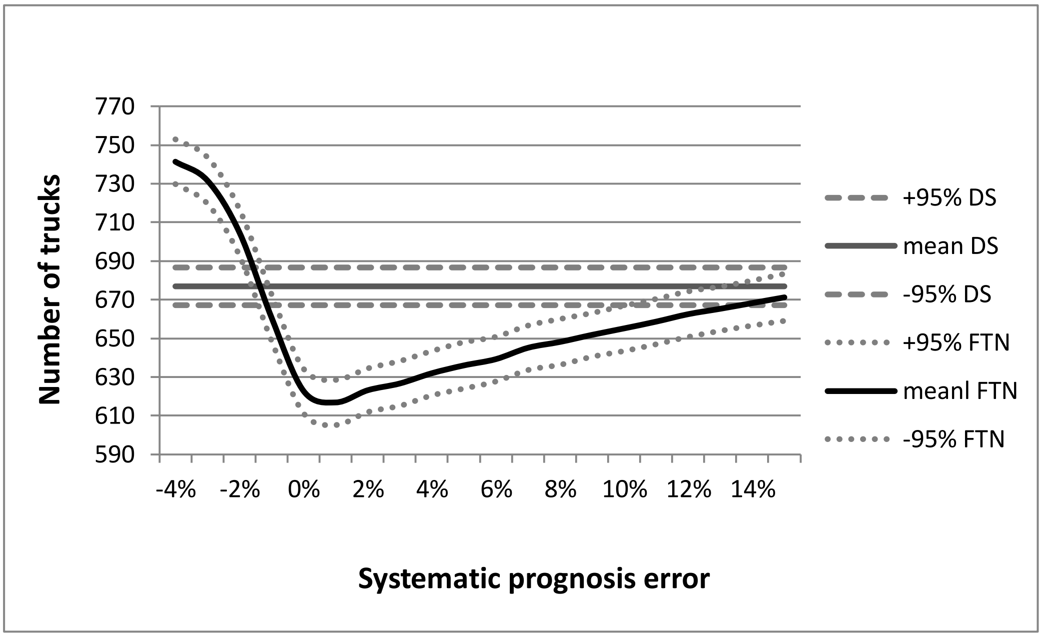


*Figure 5-2 The impact of systematic prognosis error on minimum number of trucks (+/- 95% lines illustrate the confidence interval)*

The systematic prognosis error is easily detected and manageable. It would be unlikely for the extreme levels included in the inquiry to be permitted to continue in a real system. However, it is apparent from Figure 5-2 that negative prognosis error, i.e., underestimating the real capacity needs, has a more significant impact on the results than the positive error, i.e., overestimating the need for capacity.

The systematic error is fixed as opposed to the distributed error, which is randomly distributed as a triangular distribution. The base of the triangle is enlarged from both ends with fixed increments in 20 steps, starting with 0 (presented as [a] on the horizontal axis of Figure 5-3) and finally reaching [-5%, 0, 30%] (presented as [u] on the horizontal axis of Figure 5-3).

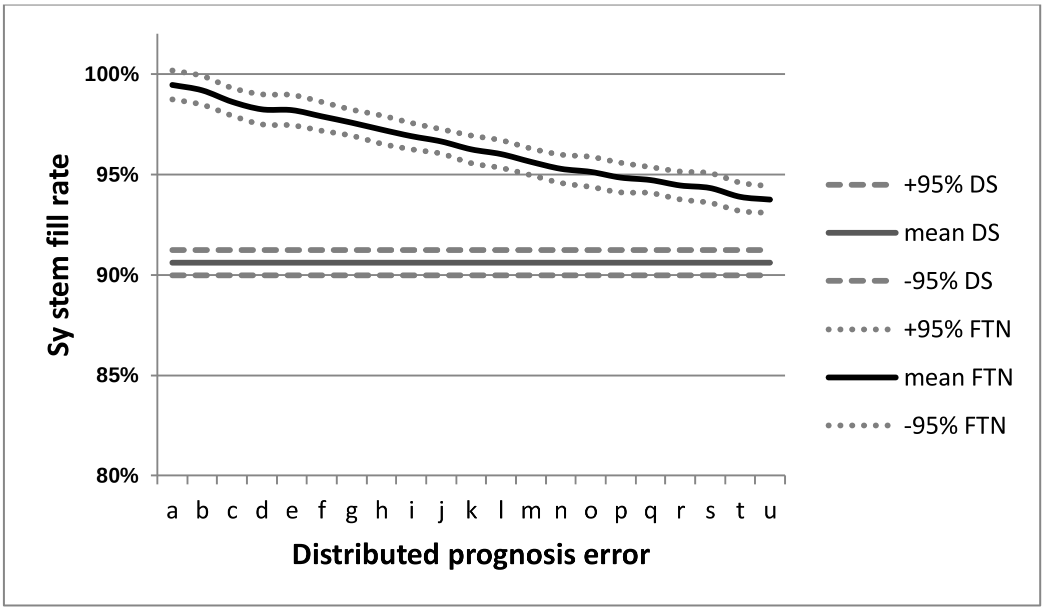


*Figure 5-3 The impact of distributed prognosis error on fill rate (+/- 95% lines illustrate the confidence interval)*

The relation between prognosis error and performance, as shown in Figure 5-3, is linear. This suggests that robustness is a result of the size of the identified potential rather that some other property of FTN that would require further explanation. The distribution of the error is assumed and would require further empirical studies to determine how likely the occurrence of levels of up to 30% error are in current and/or future systems.

### *5.3.3 Contribution to the thesis*

The findings from the third paper establish that a significant performance improvement potential exists in implementing FTN as compared to a DS. The identified potential is shown to be fairly robust. The sensitivity analysis also reveals that the obtained improvement is chiefly the result of change in setup from DS to FTN. Furthermore, some of the challenges identified in the first paper are shown to have limited impact on the identified potential.

It is demonstrated that error in capacity planning and control impact the result disproportionately negatively only when the capacity required for the HS layer is underestimated. In the case of overestimation, the impact is only marginally negative. These findings indicate that FTN can be successfully implemented while the identified planning and control challenges are not addressed at the level of detail and accuracy that was previously anticipated.

## 5.4 Paper 4 – The impact of differentiated control on the performance of Foliated Transportation Networks

The fourth paper contributes mainly to the answer of the second research question. The results presented in paper 4 are also relevant for answering the first research question. RQ2 also draws from the results presented in papers 3 and 5 for its answer.

### *5.4.1 Paper outline*

In the paper, the purpose of this study is stated as "…to explore the impact of differentiated control on the performance of a Foliated Transportation Network (FTN)." In this study the notions of when a unit can be considered "full" is examined, partially with regard to whether it is possible to statically find a value that is valid and in terms of how different strategies affect the performance of FTN.

Furthermore, the notions of real-time dynamic planning and control of the distribution of goods between the different sub layers of the network and its impact on performance are studied. This is one of the main challenges identified in the first study. The findings are important for the evaluation of feasibility and performance of FTN, two key concepts of the overall purpose of the thesis.

The DES model from the previous paper has been modified and further developed for the purpose of running the experiments designed for this one. The empirical data collected for the previous paper, along with the measures for ensuring validity and reliability of the simulation model, are near identical in both papers. The experiment requires optimization in some steps. The optimization has been performed using a commercial optimization suit included in the simulation software.

The experiment design compares FTN configurations of different levels of differentiation with regard to controlling the distribution of goods in the different sub layers of the network. Six different levels of differentiation are compared to each other with regard to their impact on the performance of the network.

One of the results from the first paper, highlighted the fact that determining when a unit is full, in order to be able to make the decision about through which layer to route said unit, requires dynamic, real-time optimization and control. This means that each individual relation in the network would need to have a unique and dynamic cutoff value at which a unit would be considered full. While this holds true at a conceptual level, it is of interest to be able to quantify how much of the identified performance improvement potential of FTN is dependent on the ability to effectively do so.

A network where the cutoff value for each individual relation is uniquely and dynamically determined would be considered to have the maximum level of differentiation. Conversely, a setup where one cutoff value was to be implemented for the entire network would be considered not to have any differentiation at all. In fact, this was the configuration used in the study where the potential of FTN was sought to quantify (paper 3).

In the main experiment of this study, six levels of differentiation were used. In the one extreme, a single cutoff value was used for the entire network, and in the other, 12 different values were used. In the configurations where differentiation was implemented, relations were, based on the empirical data, clustered together in categories and controlled using the same cutoff value. This measure was necessary because the number of relations (210) made the individual treatment of each relation too computationally demanding. The six different setups consisted of one, five, six, seven, eight and 12 categories. The different levels of differentiation are set based on the empirical data.

### 5.4.2 Results and conclusions

The results from the experiments are shown in Figure 5-4, where the x-axis denotes the number of categories, or levels of differentiation by which the FTN setup is controlled, and the y-axis denotes the network performance as a ratio of the minimum required transport work and produced traffic work. It is clear that differentiating the cutoff value for directing the flows of goods between the two layers of the network has a statistically significant effect on the network performance. However, the level of differentiation does not need to be very high to reach this potential. More importantly, the results indicate that more is not better. In fact, there is no statistical difference between the performances of the configurations with five to 12 categories. This conclusion rests on two observations.

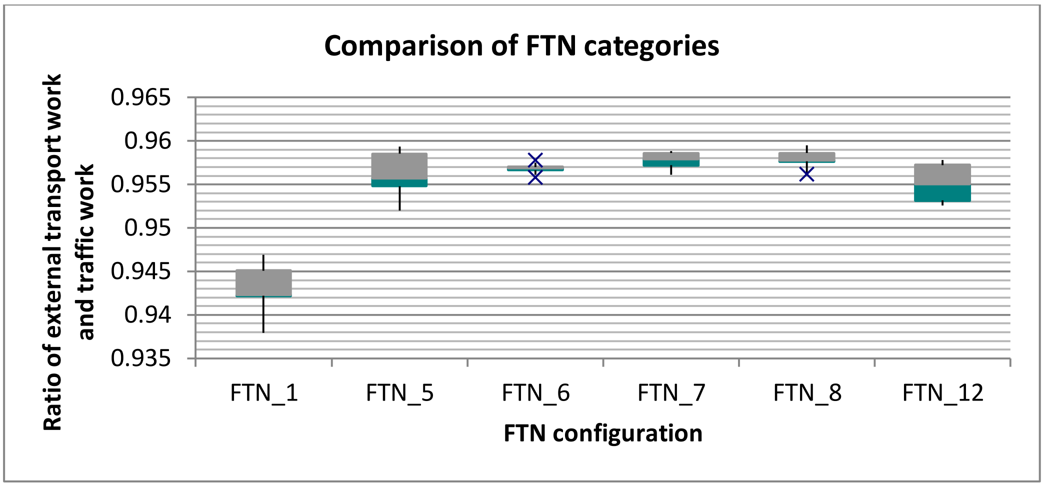


*Figure 5-4 Box and whiskers graph of the results of the trials of all FTN configurations where the number denotes the level of differentiation*

First, the effort necessary to optimize the distribution of goods across the network layers grow exponentially with each additional level of differentiation and quickly surpassing what would be operationally feasible. It is highly doubtful that the additional effort needed can be motivated with the additional potential that can feasibly be realized. Already at lower levels of differentiation, the diminishing returns of additional efforts are apparent. The relatively meager outcome of the most differentiated setup, i.e., 12 categories, is an indication of the limits of the optimization suit employed in this study. Moreover, even this result was made possible through an optimization process that required runs over a period of time that would be operationally infeasible (several days). Naturally, this time can be shortened if higher computational or optimization efficiency were to be utilized. The point remains, however, whether the additional cost of this operation would be covered by the additional improvement of network performance.

Secondly, the maximum theoretical potential that remains at this point is limited. This is further indication of the diminishing returns of real-time dynamic optimization or even continued differentiation. These results lend support to the "low-hanging fruit" phenomenon hypothesized in paper 3.

The sensitivity of the results was tested i. a. for the distribution of the size of the consignments and the mechanism used for grouping relations into categories. The tests did not reveal any cause for concern regarding the validity of the results. It can also be concluded that future studies of networks of comparable size probably can assume the volume of goods as continues without deteriorating the validity of the results.

### *5.4.3 Contribution to the thesis*

The findings from this paper bring another of the main challenges for designing a feasible FTN model identified in paper 1 under question. These results demonstrate that the marginal impact of real-time optimization and control would be negligible and likely exceed the cost of achieving it.

Although the identified design challenge remains interesting, its impact on the identified performance improvement potential of FTN seems to be marginal. The operational feasibility of implementation of FTN is hence strengthened as it is shown that the bulk of the identified potential is obtainable using rule-of-thumb-based approximation.

## 5.5 Paper 5 – In-transit services and foliated control: the use of smart goods in transportation networks

The primary aim of the fifth paper is to contribute to the answer of the second research question. The results of the study are also relevant for answering the first research question. The second RQ also draws from the results from papers 3 and 4 for its answer.

### *5.5.1 Paper outline*

This paper describes how introducing smart goods and tracking-based information management practices brings direct service improvement to customers and incremental transport efficiency improvements to transporters on the transport network level. The paper examines the dynamics of a partial/stepwise implementation of FTN and its performance-improving potential. The result regarding partial implementation of FTN constitutes the major contribution of this paper to the thesis.

The research approach utilizes design theory to develop interface modeling and discrete event simulation methodology. An empirically grounded simulation demonstrates the mechanisms generating incremental efficiency improvements for transporters as customers adopt in-transit services.

The DES model from paper 3 has been modified and further developed for the purpose of running the experiments designed for this one. The empirical data collected for the previous paper, along with the measures for ensuring validity and reliability of the simulation model, are near identical in both papers.

In order to be able to introduce foliated control, the network operator needs to be able to identify and control individual consignments in the network. At the same time, the technological solutions that would enable such an improved level of control also enable the provider to offer additional value adding services for the transport customers. Providing additional value for the customers would likely not only create demand for such services but also the incentive for customers to carry some of the additional cost.

The concept of smart goods is used for the technological solution that would enable the improved level of control. The experiment has its vantage point in the idea that introduction of smart goods in the system would likely occur stepwise, i.e., only portions of the goods going through the network would be endowed with these additional capabilities. Assuming that only that share of the total that is smart would be eligible for foliated control, it becomes of interest to know, both with regard to feasibility and potential, what impact the varying amount of smart goods would have on the ability to implement FTN.

The fact that only a subset of the total amount of goods in the system are eligible for distribution between the different network layers, two conditions need to be met for a unit to be routed via the hub and not directly. For one, the amount of goods remaining for the last unit in a relation needs to be less than what would constitute a full unit, and the amount of smart gods in the same relation needs to be greater than what needs to be rerouted. This captures the limitation that leads to a partial implementation of FTN.

In an experiment where the share of smart goods is incrementally increased from 0 to 100% some interesting question about feasibility and potential can be answered.

### *5.5.2 Results and conclusions*

The results show that to realize the full potential of foliated transportation network, only half of the total volume needs to be available for individualized planning and control, i.e., as smart goods (Figure 5-5). Also, it is evident that the earlier increments of the available amount of smart goods result in a larger impact on the system efficiency potential than the later ones. Figure 5-5 also illustrates that about 80% of the total potential of Foliated Transportation Network is feasible to realize with a 20% share of smart goods in the system. This means that the introduction of smart goods will readily make new efficiency potentials available for the transporter without any notable threshold effects.

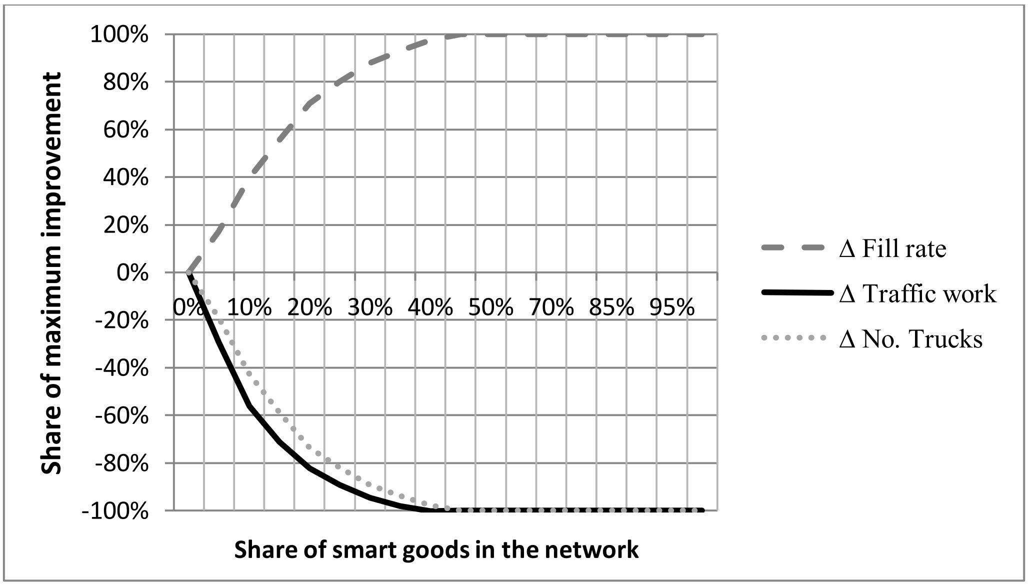


*Figure 5-5 Impact of increasing the share of smart goods available on the transport network's key performance indicators (expressed in percent of maximum improvement potential)*

### 5.5.3 Contribution to the thesis

The interface in-transit concept was originally developed to illustrate how a number of customer-focused services rely on a common set of interaction patterns. The very same approach to handling smart goods enables the stepwise/partial introduction of FTN. The same technological platform, e.g., smart goods, necessary for improving customer service can at the same time also be harnessed to increase network level efficiency.

These results also establish that substantial improvement can be realized with a partial implementation of FTN. This is important to note for two primary reasons. For one, reducing the amount of goods shipped through the hub and still obtaining substantial performance improvement will likely have implications for the implementation of FTN. This conclusion has implications regarding the identified and partially validated challenge regarding handling performance. Intuitively, the impact of handling performance on the overall performance of the system will be limited as the volumes that are routed through the hub are substantially reduced. Secondly, this property will likely diminish the obstacles for achieving an empirical test in an existing real network, which should be the next appropriate step for furthering the research about FTN.

## 5.6 Summary of paper findings

The outcome of each paper relevant for addressing the research questions is summarized in Table 5-3.

*Table 5-3 The outcome of the research questions*

| Paper | Aim | Findings |
|---|---|---|
| **P1** | Identify design challenges | ● Identification of design challenges<br>● Hypothesizes the potential for use of approximated rule-of-thumb-based control in lieu of analytical solutions |
| **P2** | Model and evaluate individual layers | ● Validates model with empirical data<br>● Provides the basis for the simulation model<br>●Supports deduction of performance improvement potential hypothesis<br>●In part validates handling performance challenges identified in P1. |
| **P3** | Quantify performance improvement potential of FTN | ● Confirms deduced hypothesis of performance improvement potential<br>● Hypothesis of additional potential to be gained by optimizations of system design<br>● Reveals the sensitivity of FTN to prognosis error to be relatively limited |
| **P4** | Investigate impact of differentiated control on FTN network performance | ● Indicates that static approximation in control is sufficiently effective<br>● Indicates diminishing returns regarding the real-time dynamic control (i.e. in part rejecting hypothesis from paper 3) |
| **P5** | Investigate the impact and feasibility of partial implementation | ● Indicates the partial/stepwise introduction of FTN is feasible<br>● Indicates that a majority of potential is attainable with minor implementation |

The results from the first study, presented in paper 1, primarily contribute to answering first RQ. In this paper, the challenges for designing an operationally feasible FTN are defined and identified.

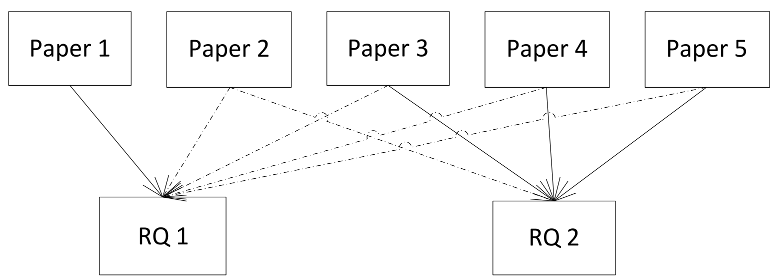


*Figure 5-6 The contribution of results presented in different papers to answering the RQs*

The second study, in part presented in paper 2, has modeling and evaluation tools and hypotheses that may be viewed as intermediaries, which enable the analysis of the second RQ. The results from this study also highlight some of the dynamics of the two network layers of which FTN consists and some of the challenges identified in the first study.

The results from the second study, which is the most extensive one, are also presented in papers 3 through 5. Even though the model on which the experiments are performed is redesigned for the purpose of each individual paper, the fact that all three models are based on the same empirical data makes it reasonable to view papers 3 through 5 as the results of the same study. Paper 3 provides evidence for and attempts to quantify that the hypothesized improvement potential, which is crucial for answering the second RQ.Papers 4 and 5 indicate further that the benefits of FTN are attainable even with partial/stepwise implementation and static approximation as opposed to dynamic real-time control. The results presented in papers 4 and 5 also contribute to answering the first RQ. The relationship between the results presented in the different papers and the RQs that they help answer are presented in Figure 5-6.

# 6 Results and discussion

*To reach the purpose of this thesis, two research questions were devised and addressed with two studies. The results of these studies will be discussed in this section.*

The purpose of this thesis makes clear two principle areas of interest regarding the study of Foliated Transportation Networks (FTN), namely, evaluation of the feasibility of implementing FTN and the performance improvement potential of the concept. This section follows, in part, the same logic in its structure. An additional heading is added regarding the network characteristics that contribute to explaining the mechanisms yielding the identified potential of a feasible FTN.

The first part of this chapter focuses on the aspects of the feasibility of implementing a foliated transportation network. Secondly, the network characteristics are discussed, which is meant to provide the explanatory underpinning of the results regarding both feasibility and potential. The final part of this chapter focuses on the evaluation of the potential impact of an FTN implementation on the performance of the network. The evaluation is mainly based on the simulation studies but also qualitative results and theoretical considerations.

## 6.1 Evaluating feasibility

One aspect of evaluating the feasibility of an FTN implementation is to identify the design challenges that exist and the theoretical domains within which these challenges can be overcome. The result from such an effort cannot be expected to be comprehensive and exhaustive on all levels. The choice of what to include at this stage is based on what needs to be addressed in order to enable feasible operational design of an FTN with regard to existing comparable systems. The evaluation of feasibility will then be based on the size of the identified design gaps and the necessary effort to sufficiently bridge them. In some cases, detailed theoretical knowledge might be lacking and design challenge might end up being left unaddressed within the theoretical domain in which they reside, or be tackled using rough approximations or rules of thumb. To be able to assess these shortcomings' practical importance for the operational feasibility of FTN and impact on the identified potential is also critical for the evaluation at hand.

The identified design challenges can be organized into three categories: transportation planning and control, transportation operations and transportation network optimization. These broadly labeled categories are rooted in three principal theoretical domains, namely, transportation management, information and communication sciences and mathematics.

In the following, the identified categories of design challenges will be discussed. In order to bridge the existing design gaps, both the need for developing new knowledge as well as new applications for existing knowledge have been identified. Also the possibility and consequences of bridging identified gaps with rule-of-thumb and approximated solutions have been discussed.

Transportation management research is a cross-disciplinary, applied field of science where, e.g., network theory and information sciences cannot be considered to be completely foreign to the field of transportation as is. Similarly, the identified design challenges are all in their application intertwined, and they all affect each other. In favor of providing a comprehensible structure, some distinctions are made between the theoretical domains even though the cross-disciplinary nature of transportation management research allows for the inclusion of three theoretical domains without the distinction presented.

### *6.1.1 Transportation planning and control*

The major divide in terms of transportation planning and control, between the transportation networks of today, i.e., DS and HS networks, and FTN, is rooted in the purpose or aim of the planning and control operations. In this context, the planning and control referred to is that of load capacity and routing. In DS and HS networks, routing is statically predetermined and is not much of an issue as opposed to FTN. Currently, the principal purpose of transportation planning and control is preventing capacity shortages. Simply put, as long as there are not any shipments left behind in the terminals when the last truck departs, the planning and control effort has fulfilled its purpose, regardless of the capacity activated in doing so. This simplification is purposeful and intends to clarify the fundamental difference in the planning and control effort necessary in the different setups. In reality, there always exists a trade-off between utilization and productivity on the one hand and customer service (i.e. effectiveness) on the other.

In contrast, in an FTN setup, the objective of the planning and control effort is the prevention not only of capacity shortages but also of underutilized capacity. In effect, a perfect match is sought between the capacity required in each network layer and the capacity allocated. This, combined with the inherent discrete property of transportation capacity, is where the need for development arises. These statements are based on the assumption that the network operator will continue the practice that is widespread today, namely, that customer orders are almost never turned down, at least not based on capacity utilization concerns.

Naturally, operators of traditional DS networks also seek to minimize underutilized capacity. This is normally accomplished by trading off delivery precision (effectiveness) with capacity utilization (efficiency), as few other options are available in such systems. This is in part accentuated by the practice that operators almost never turn down assignment requests from customers. In light of this, the distinction between FTN and its comprising layers in isolation becomes clear, when considering FTN's goal of sustained or improved delivery precision. Trading efficiency for effectiveness could be a zero sum game with regard to overall performance, whereas sustaining or improving the later while increasing the former ought to always lead to superior performance.

Another distinction needs to be made at this point because there are two different aspects of the planning and control operations that are interesting with regard to FTN design. One aspect of allocating capacity and goods to the different layers is that of optimization, which will be dealt with below under a separate heading. The other—the one in focus in this section—is that of being able to execute the optimized plan, i.e., the ability to minimize the error between the prognosis for capacity need and actual outcome. The applications in use today are not designed to handle planning and control at that level of detail (Sternberg, 2011). In order to be able to minimize this error, new applications are needed. These new applications and their consequent impact on the design and implementation of FTN are heavily dependent on what will be feasible to achieve with new and existing information, communication and identification technologies and how the size of the error affects the performance of FTN.

The results from the simulation studies, which are based on empirical data, reveal that the FTN setup is relatively insensitive with regard to planning and control error. This indicates that even though design and development of new applications or adaptation of old ones to FTN are useful, much of the identified potential is obtainable within the limits of existing technology and approximate rule-of-thumb-based planning and control. This is promising because even given new and more effective tools to collect, process and communicate necessary information, the customers' willingness and ability to provide reliable and accurate information might not be on par with what would be required. The same applies to real-time network optimization and dynamic control of the distribution of goods between the different network layers. This is particularly true for systems with a relatively stable flow of goods, which is not uncommon for general cargo freight transportation networks.

Until now, the perspective of this discussion has been on what is/would be technically feasible. However, it is worth mentioning that what is feasible or technically most efficient does not automatically translate to that which is desirable. There are business considerations that fall outside the scope of this thesis but nonetheless cannot be completely ignored. Given an enhanced precision in capacity planning, control and allocation, it might, from a business point of view, make sense to operate the system in another way than prescribed in the FTN model today. For instance, new opportunities may arise where excess capacity could be preserved in order to widen the time window for accepting orders or to provide special services within the system that would require deviation from the technically superior design. In a detailed design, these aspects also need to be addressed.

### 6.1.2 Transportation network optimization

The dimension of transportation network optimization can be divided into three different domains where two regard the optimization of the planning and control of the network and one the actual execution.

The first set of issues finds its solutions in the network theory and mathematics. Two special cases of non-trivial presently unsolved combinatorial mathematics problems will have to be involved in the design of FTN. The first is a special case of a multi-dimensional multi-constraint Bin Packing Problem, the development of an application of the solution of which is required for the optimization of the use of the total loading capacity. With properly chosen optimization parameters, the loading composition and sequencing of individual trucks with the goal of overall system optimization can be addressed. The second problem is a multi-dimensional multi-constraint special case of a Change Making Problem, the development of an application of the solution of which is required for the optimization of the use of hub facility resources. Similarly, the choice of optimizing parameter is not given and needs to be determined in the research for such applications. A third problem, with roots in mathematics and network theory, is the globally optimal allocation of load capacity and goods to the different layers of the FTN.

It is safe to say that none of these issues is likely to find an analytical solution; rather, a heuristic or rule-of-thumb-based solution is called for in this context. However, existing theory and solutions to general forms of the problems can help provide a theoretical basis for any attempt at a heuristic or rule-of-thumb-based solution. Furthermore, given an analytical solution, any application based on such methods would for its success likely require access to detailed and reliable information about the goods. Inherently, the information required is to be supplied by the customers, who historically have demonstrated a poor track record in providing accurate information. This is especially true when the requested information only provides more efficient transportation operation with a marginal price incentive for the customers as the primary upside.

The second set of issues related to network planning and control optimization is that of transportation management. As pointed out earlier, the planning and control precision sought after in an FTN design might afford new business opportunities and services. Optimization based on those grounds is highly relevant but falls outside the scope of this thesis. However, they may not be ignored in the design of an operational FTN model.

Finally, as the ability to execute dynamic rule over FTN and perform the planning and control effort called for here, new IT applications are a necessity. In effect, the operational FTN needs to be re-optimized upon arrival of every new order. This will not be possible without developing new ITC applications for this purpose. The issue of customer-provided information remains an obstacle in this context also, as elaborated on above.

Although all three areas mentioned above carry significant value and interest in their respective theoretical domains, the findings of the experiments indicate that they might only impact the feasibility and performance potential of FTN marginally. In fact, the rule-of-thumb-based approach to optimization, planning and control and static approximation of dynamic elements of the network produce results that suggest that the point of diminishing returns will have been reached already at this stage.

### *6.1.3 Transportation operations*

The additional consolidation step that the HS layer of FTN entails creates new operational challenges that need to be dealt with. Primarily, the time required for the terminal operations in general and hub terminal operations specifically need to be minimized. Though outside of the scope of this research, the pickup and distribution operations ought to also be included in the design of an operationally feasible FTN. The pickup and delivery sequence of orders, the arrival time of pickup trucks and departure time of distribution vehicles and the terminal handling times required all impact the design of detailed operating rules of the network. Time windows for different activities are directly dependent on these operations. In a detailed design of an operational FTN model, these aspects cannot be ignored.

In this respect, which is closely intertwined with the planning and control issues discussed above, information, communication and identification technologies could potentially play an essential role. New identification technologies such as RFID and other automatic identification technologies are likely to reduce the terminal handling time as well as improve real-time planning and control of the network. Route planning for pickup and distribution operations, load sequencing and presorting of goods destined for the HS layer are other aspects that are dependent on new applications of ITC and hold a major potential to affect the detailed design of an FTN model. Furthermore, the same technological platforms will likely enable the transportation service provider to produce additional value-adding services that customers could potentially demand and be willing to pay for. This would carry some of the additional cost at the same time as it provides value for the customers and creates efficiency in operations for the provider.

Such considerations also raise questions regarding the centralization/decentralization of decision making in the network. New emerging areas like complexity and concepts such as smart freight are promising new areas where some of the answers to these types of questions may be sought. It is hypothesized that the paradigms of centralization/decentralization have fundamental influence on the FTN design principles. However, the indications regarding the feasibility of implementing rule-of-thumb-based control applies to centralized and decentralized systems alike.

### *6.1.4 Too wide a gap to bridge?*

Based on the discussion above, it can be concluded that design gaps exist in three principle theoretical domains, i.e., transportation planning and control, transportation operations and transportation network optimization. Theoretical knowledge as well as new applications for existing theory in order to arrive at a set of design principles for an operational model of FTN is in part lacking. However, the results from the simulation experiments based on empirical data utilizing rule-of-thumb planning and control and static approximation of the dynamics suggest that the impact of the identified gaps might not be a critical obstacle for the design and implementation of a feasibly operational FTN.

In light of this, and also in light of the overall purpose of this thesis, it becomes interesting to investigate the potential of FTN for improving the physical performance of a transportation network compared to existing systems. Evaluating the potential of FTN holds isolated value on its own, and explaining the mechanisms yielding the identified potential is valuable for practitioner and academics alike.

## 6.2 Network characteristics

The results from papers 2, 4 and 5 contribute to the explanation of the networks' characteristics and the mechanisms that yield the performance potential of FTN. They also contribute to providing the explanatory underpinning for some of the conclusions drawn regarding the feasibility of an FTN implementation. In addition, the results converge with extant theory and support the validity of the developed network model. This has provided a compelling argument for reusing the same model description as the starting point for the simulation model to be developed for the simulation studies.

### *6.2.1 Layers in isolation*

Inherent with the HS system characteristics, the number of links in the network is significantly reduced when the model setup is altered from DS to HS. This fact, along with the results revealing that the required number of resources in the network decrease at the same time as the minimum transport work increases, are coherent with the increasing average filling rate and resource utilization rates that the models yield.

Under the assumption that all but the last truck, from any origin to any destination, in each relation, is fully loaded, i.e., the maximum possible number of trucks in the system with a fill rate less than 100% is equal to the number of physical relations, i.e., the number of links multiplied by two, the findings above can easily be explained, considering the significant decrease of the number of links in the network. The reduction of the number of links in the network produces the same effect on the maximum possible number of trucks in the system that might not be fully loaded. The subsequent effect of the decline of the share of the total number of trucks that may not achieve 100% fill rate is not only a reduction of the number of resources necessary, but also a boost of the average fill rate. Both of these impacts are driven also by the consolidation process inherent in the HS network.

The minimum required transport work and the mean distance and transit time between nodes in the network also proves to increase in the HS setups. These results are coherent with the existing theory as the goods in such networks seldom travel the shortest way, as opposed to the DS networks, where this always is the case. In addition to the rise in the average traveling distance between the nodes of the network, the consolidation at the hubs requires coordination of shipments to and from the hubs, resulting in stalls for “quicker” shipments, forcing them to adhere to subsystem “worst times.” The discrepancies between the change in the average distance and the alteration of the transit time suggest that the time-consuming effect of the increased distance is less significant than the impact of the heightened coordination needs related to the increasing number of hubs, on the mean transit time.

The increase in mean transit time reduces the transportation systems’ quality and flexibility of service. In order to comply with the latest time of delivery, the time window for submitting goods to the network at the nodes will shrink; in certain instances it will shrink considerably. For the HS systems to function properly, the hub terminals are required to have the capacity to handle surges of arriving/departing loads in a very limited period of time. This will most certainly lead to the necessity of extensive overcapacity at the hub terminals regarding facilities and access to a workforce, the size of which can be adapted to the overtime fluctuating workload, whereas the flow of goods through the terminals in the DS system is much more leveled.

### *6.2.2 The impact of parts on the whole*

Three main contentions from the discussion above are central to the design and potential of FTN. These include, first, the impact of the number of links on the system’s capacity utilization; second, the impact of additional consolidation steps on the network mean time and distance and third, the impact of hub coordination on the total time in transit. This provides a solid basis for the development of an experimental model.

The capacity utilization gains on system level when replacing a DS with an HS are a result of the reduction in the number of links. Assuming that it is possible to utilize 100% of every truck’s capacity as long as there are enough goods available means that the number of trucks that risk not having enough goods available to reach a 100% fill rate is equal to the number of relations, i.e., twice the number of links in a network. However, it is also evident that because the goods being shipped through an HS network almost never travel the shortest[12] way, the amount of transport work increases in an HS network.

The superiority in performance expected in FTN compared to DS and HS is rooted in this observation. If trucks that are 100% full are sent directly, the

[12] Goods that have the hub terminal as the final destination are naturally not subjected to this condition.

system fill rate will not be negatively affected due to the potentially less-than-full last trucks in every relation. The portion of the goods that fall outside of this criteria are consolidated in the HS layer of the FTN, resulting in the minimum number of trucks risking departure with underutilized capacity at the price of additional transport work and mileage for that portion of the goods. Furthermore, the mean time between nodes in the network will increase partly because of the additional distance to cover and partly because of the coordination effect where goods spend more idle time at the hub terminal. These observations translate into a host of implications for the design and potential of FTN.

The inference of these circumstances implies that the maximum efficiency improvement potential of a Foliated Transportation Network is inversely proportional to the number of trucks per origin/destination relation. Foliated transportation network is hence only applicable for a "Goldilocks" transportation network that is not too big or too small. If the volume in each relation is small, the network setup would ideally be a hub and spoke network optimally with the ability to shortcut (Lumsden et al, 1999). If the volumes are large, the impact of removing the inefficiencies of the last unit in each relation would become marginal to negligible on the system level. The importance of the unit size also needs to be highlighted, as the link described above is between number of units per relation rather than goods volume. On the other hand, transportation systems are characterized by efficiency properties that are driven by scale economy, which implies that a system utilizing smaller units than necessary is, in this context, inherently less efficient than one utilizing units that are as large as possible (Hultén, 1997).

In direct opposition to the point about scale economy and efficiency, reducing the unit size in a direct shipment network would positively affect some of the indicators used for measuring network performance, i.e., diminishing the improvement potential of Foliated Transportation Network regarding fill rate and traffic work. At the same time, this tactic would obviously affect the number of trucks inversely. However, it is evident that such a system cannot be considered more efficient. What is revealed here is the complex nature of transport efficiency and the difficulty of capturing it with a single measure. All of the included performance indicators are necessary simultaneously in order to be able to assess the efficiency of the system.

Especially when viewing the system from a resource consumption perspective, the number of trucks is a significant driver of crucial parameters such as fuel consumption, human resources, road congestion and so on. The arguments above can thus be summarized as follows: the most efficient direct shipment network utilizes as large a unit as possible, thereby minimizing the number of units per relation. In this context, the impact of Foliated Transportation Network is significant and readily available.

On the same theme, it ought to be desirable, if not vital, to (with respect to the increased network mean time between nodes in the HS layer) identify and ship the goods destined for the hub as quickly as possible. In fact, this is the very reason why the need for research on the enhancement of the precision of transportation planning and control has been identified as one of the important design challenges, where there exists a gap. It is reasonable to assume that even upon improving the accuracy of prognosis, planning and control of the allocation of capacity and goods between the different layers of FTN, the discrepancy between plan and outcome will likely not amount to zero. This in turn highlights the significance of the ability to evaluate the sensitivity of FTN for such errors.

## 6.3 Evaluating performance improvement potential

Being able to handle the challenges that an increase in network mean time between nodes would bring about has been cited above as one of the prerequisites for a successful design and implementation of FTN. This applies regarding both operational issues such as terminal throughput time and consolidation operations and planning and control issues regarding the distribution of goods between the layers and departure time for goods intended for different layers of the foliated networks. These constitute constraints on the ability to quantify the performance improvement potential of FTN under the assumption of improved or sustained service quality (i.e. network effectiveness). Additionally, the possibility of offering potential additional services that would be enabled due to the full-scale implementation of FTN also affects the performance of the network. Nonetheless, it is possible to quantify the performance improvement potential of FTN with the condition of sustained level of service. The service quality in focus here is the latest time for accepting an order, pickup and delivery timeframes and the order information provided by the customer.

Bjeljac and Lakobrija (2004) studied the feasibility of implementing FTN, as they define it to be, within the limits of an existing network, with positive results. These results, i.e., the feasibility of routing goods via a hub within the existing time constraints, have been duplicated in the second paper. Invoking those results, it is possible to both investigate and quantify the performance improvement potential of FTN and to subject those results to sensitivity analysis regarding the challenges specified above, particularly those regarding the need for and impact of new planning and control applications. In line with this reasoning, the delimitation of the potential additional improvement, due to the resolution of currently remaining design gaps identified, poses no detriment to the validity of the results produced. The validity of the evaluation is also enhanced on the back of thorough and comprehensive sensitivity analysis. The cornerstone of the evaluation and quantification effort consists of experiments run in the simulation model of an ideal typical network model developed based on empirical data. The layers of the network models are based on the models presented in Chapter 4, and the model is modified for the purpose of each experiment run.

### 6.3.1 Results from the main experiment

The main experiment (presented in paper 3) was meant to test how the performance of the network, measured using the KPI number of trucks, transport work and traffic work, when the network was modified from a pure DS network to an FTN. The first run of the experiment highlighted one of the predicaments anticipated, namely, how full is a full unit?

Setting the cutoff fill rate for redirecting a truck via the hub to 100% in the first trial of this experiment yielded the results presented in the first column of the table above (Table 6-2). As evident from the table, this approach yielded a significant improvement with regard to the number of vehicles. However, with the traffic work unaffected and the transport work increasing, the only discernible impact with regard to the negative external effects of the network stems from fewer trucks with higher fill rates, driving just as far as previously, i.e., a marginal deterioration with regard to environmental impact.

*Table 6-1 Results from the main experiment*

| | FTN potential (DS=100%) | FTN potential (DS>75% ) | Lowering cutoff for hub routing |
|---|---|---|---|
| **Average fill rate** | +14.5% ± 0.2 | +13.7% ± 0.23 | Slight decrease in fill rate |
| **No. of trucks** | -10.5% ± 0.4 | -9.6% ± 0.41 | Slight increase in no. of trucks |
| **Transport work** | +5.2% ± 0.5 | +2.6% ± 0.45 | Slight decrease in transport work |
| **Traffic work** | No significant difference | -13.6% ± 0.44 | Significant decrease of traffic work |

A closer look revealed that setting the cutoff to 100% was a modeling mistake due to the fact that achieving 100% fill rate in any unit in any relation, given the discrete nature of the load units and the goods, is extremely unlikely. This meant that, in this configuration, practically every last truck in every relation was being sent to the hub regardless of its fill rate. To control for this flaw in the original model, a new cutoff value was chosen. The new value of 75% is a judgment-based rough estimate. Any choice would have been a simplification as the conceptual model is based on dynamically set cutoff values and not a single static one for all relations in the network.

However, even with this simplification, the results demonstrate the potential of FTN as compared to DS with regard to network performance. The KPI measuring number of vehicles, fill rate and transport work are largely unaffected at the same time, as the necessary traffic work has decreased significantly. Simply put, by using FTN, fewer trucks with higher fill rates need to drive a considerably shorter total distance to fulfill the same transportation demand in the case of FTN than if a DS structure were used. This change has positive economic and environmental implications.

### *6.3.2 Sensitivity analysis*

The impact of the factors, system setup, number of terminals, volatility of demand, loading error[13] and truck capacity, were examined in a two-level five-factor complete factorial design. The five factors were selected based on qualitative reasoning and previous results presented in the literature. The factors and their levels are found in Table 6-2.

The result variables for this experiment are the ratio of total volume and the number of trucks (R1), the average fill rate (R2) and the ratio of total transport work and total traffic work (R3). The reason for including two ratios is the fact that the number of terminals (factor B) affects the number of trucks and total traffic work, rendering the results incomparable between the two levels of that factor. By introducing a ratio, the relative indirect impact on the number of trucks and the traffic work could possibly be revealed.

In the factorial design experiment factors A, D and AB have the highest impact on all three result variables consecutively. Factor A, i.e., the system setup, has by far the highest impact in all three cases. The system setup has twice the impact of factor D and almost five times the impact of factor AB

[13] Loading error refers to the discrepancies that arise in fill rate due to the composition and loading sequence of the goods on a truck.

(see Table 6-3). The variable (K) denotes the reference interval of experimental error with 0.95 confidence interval.

*Table 6-2 Five-factor factorial design*

| | Factor | High (1) | Low (-1) |
|---|---|---|---|
| **A** | System setup | FTN | DS |
| **B** | No. of terminals | 15 | 8 |
| **C** | Demand | Empirical standard deviation | Empirical standard deviation + 50% |
| **D** | Loading error | U[0.95 , 1] | U[0.85 , 1] |
| **E** | Truck capacity | 40 ton | 25 ton |

The relatively high impact of factor D on the overall results further indicates the importance of some of the issues identified in the first paper regarding network planning and control, operations and optimization. The ability to mix and match goods in a way that minimizes the impact of factor D is dependent on the results of all three identified design dimensions. The combined effect of factor AB shows that FTN favors the denser alternative network in this analysis, further supporting the argument put forth about when FTN is more appropriate to implement, based on network characteristics.

*Table 6-3 Result of factorial design*

| Factor | Volume/no. Trucks (R1) | Average truck fill rate (R2) | Transport /traffic work(R3) |
|---|---|---|---|
| **A** | 0.103911 | 0.072164 | 0.074589 |
| **D** | 0.047858 | 0.045098 | 0.045364 |
| **AB** | 0.026314 | 0.022663 | 0.020262 |
| **±K**[14] | 0.003436 | 0.000706 | 0.000702 |

In the model description of FTN presented by Persson and Lumsden (2006), it is required to on the basis of prognosis identify and ship the hub volumes before the direct volumes depart from the origin terminals. This approach inserts an uncertainty into the setup, the effects of which are difficult to foresee. Therefore, this aspect has been included in the sensitivity analysis of the results.

The outcome of the FTN model according to Persson and Lumsden (2006) has been compared to the DS model outcome with an incremental fixed

[14] Reference interval of experimental error ($p<0.05$)

prognosis error.[15] The error has been increased by increments of 1% from (-4%) to (+15%). A negative error indicates that the capacity need has been underestimated in the plan, and conversely, a positive error indicates an overestimation of the capacity need. With this approach the impact of just the size of the prognosis error on the results has been clarified. It is, however, unrealistic to assume that a fixed systematic error would be sustained without correction. Therefore, the same sets of tests have been run with a randomly distributed error with incrementally increasing standard deviation. A triangle distribution has been used for this end where the extremes have been increased in 20 steps to go from trig [0.9525, 0.95, 0.935] to trig [1, 0.95, 0.65]. The result of this analysis is meant to highlight the robustness of the results of the FTN with regard to the size and spread of the prognosis error.

The results show that FTN is fairly robust in this respect. The impact of the prognosis error on the overall performance of the network is limited in the sense that the error must reach unrealistically high or volatile levels before the performance improvement of FTN (as compared to DS) is erased. This robustness implies low-hanging fruit benefits, which in turn implies that FTN implemented with support of even a crude planning and control system would be likely to yield a performance gain.

### 6.3.3 Effectiveness/efficiency trade-off

An assumption that all the models share is that of perfect delivery precision, i.e., all the consignments get delivered within the promised time window, which in this context means overnight. The impact of this assumption is not negligible on the capacity utilization results especially regarding the DS network, because what it ultimately means is that regardless of the utilization rate of the last truck in every relation, the truck will be dispatched to accommodate this requirement. Taking into consideration the fact that the size of the units in the fleet is uniform, this could result in grossly underutilized units in some relations. The reason for opting for this approach is the difficulty of modeling a rule for handling the trade-off between efficiency and service level that would be general and at the same time not be arbitrary.

In a straight comparison between different setups, this condition does not produce an adverse impact. The problem arises from the fact that the existing theory and empirical evidence strongly suggest that the cost for achieving perfect delivery precision will exponentially increase the closer

[15] Prognosis error refers to the variation in fill rate due to the discrepancies between forecasted need for capacity and actual outcome on a truck-by-truck basis.

to 100% it comes. In fact, the empirical data collected from the reference companies inferred as much.

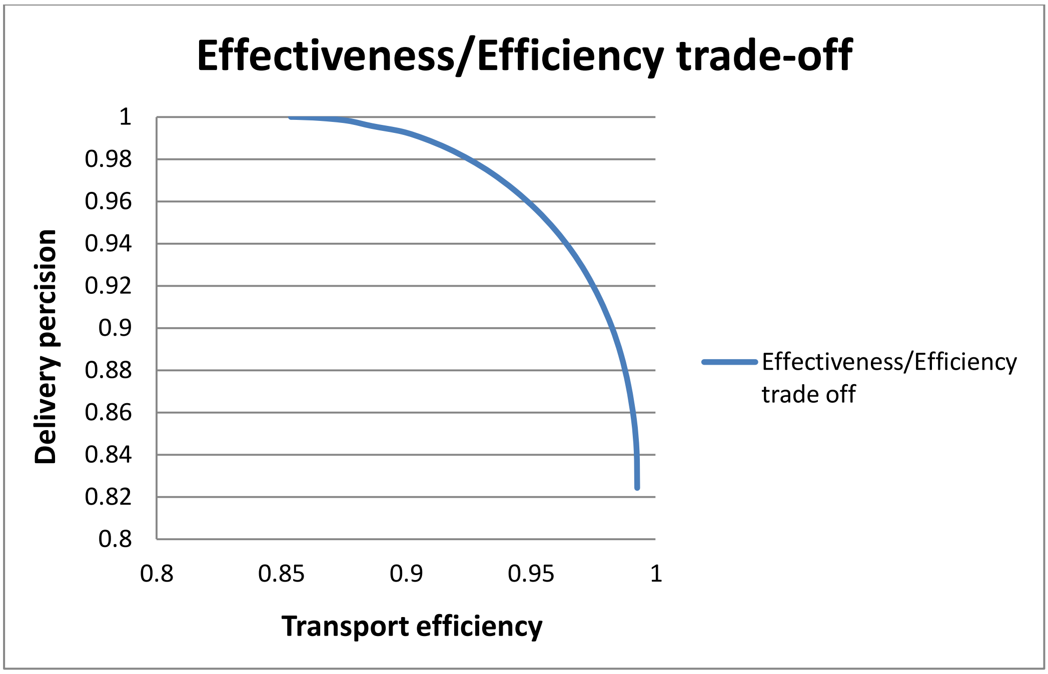


*Figure 6-1 Trade-off between delivery precision (effectiveness) and transport efficiency obtained from an experiment run in the simulation model*

However, much of the evaluation effort is actually comparative in nature, and the purpose of the model developed was to represent an ideal typical network and not to mimic any specific network already in existence. In order to test the sensitivity of the results regarding this trade-off, the following experiment was run (Figure 6-1). The experiment is run on a DS setup, and the vertical axis in the figure represents delivery precision expressed in percentage of goods delivered overnight, while the horizontal axis represents transport efficiency expressed as the ratio of external transport work and total traffic work. The results, which are based on empirical data, correspond very well with the established theory on the subject.

What this means is that in cases where the identified potential of FTN does not appear to apply due to the fact that the existing DS network does don't display the low levels of capacity utilization present in the modeled DS network, the trade-off makes it possible to increase service levels with sustained or improved efficiency, hence improving the overall performance of the transportation network. This would i.a. affect the maximum cost that an FTN implementation would be allowed to cause as reducing inefficiency is more unambiguously valued than improving service quality.

### *6.3.4 Last truck fill rate*

The FTN setups modeled here are based on the existing cited literature. In these, no specific level of fill rate is determined for the decision of sending the last truck through the different network layers, i.e., DS or HS. Based on this, any truck that is not 100% filled has been sent through the HS layer, i.e., via the hub terminal. Reviewing the results of the original experiment reveals a conceptual gap in the model. In reality, one would probably prefer to send, e.g., a 90% full truck directly instead of shipping it via the hub. The underutilized capacity may not warrant the additional time, distance and terminal handling operations that a hub detour would entail. Based on the results from the first paper, this design gap was partly anticipated. In addition to the detour, issues of flow balance at the hub and hub truck fill rates could also come into play, i.e., if adding additional volumes to the hub layer would significantly deteriorate the flow balance at the hub or fill rate of (and by default the number of trucks) the hub-bound trucks. Furthermore, the size of the impact of these dynamic properties are of interest for being able to evaluate the feasibility of implementing FTN given the limits of existing analytic solutions and technologies.

Based on this, it is hypothesized in the conceptual model that the cutoff value would likely be different for different relations both due to static conditions of different nodes of the network and the dynamic interplay of goods distribution within and between nodes and network layers. The answers to these questions are sought by trying to assess the impact of the cutoff value on the overall performance of the network. Hence, the sensitivity of the results to the impact of different levels of fill rate that would determine a truck's rerouting through the hub terminal has been investigated.

The experiments show, upon optimizing the system based on single cutoff value, that, the optimal level for considering a unit full deviates from 100% and is closer to 75%. The fact that the result of the effort to optimize a single static cutoff value has yielded the same number as the one used in the first experiment, which was a more or less arbitrary level estimated based on judgment, is purely coincidental. Statically approximating the dynamics of network and differentiating the cutoff value for directing the flows of goods between the two layers of the network for different relations clearly produces a statistically significant effect on the network performance. However, the level of differentiation does not need to be very high to reach this potential. More importantly, the results indicate that more is not better. In fact, there is no statistical difference between the performances of the different levels of differentiation that are included in the experiment. This conclusion rests on two observations.

First, the effort necessary to optimize the distribution of goods across the network layers grows exponentially with each additional level of differentiation, quickly surpassing what would be operationally feasible. It is highly doubtful that the additional effort needed can be motivated with the additional potential that can feasibly be realized. Already at lower levels of differentiation, the diminishing returns of additional efforts are apparent. The comparatively meager outcome of the most differentiated setup is an indication of the limits of the optimization suit employed in this study. Moreover, even this result was made possible through an optimization process that required runs over a period of time that would be operationally infeasible (several days). Naturally, this time could be shortened if higher computational power or more powerful optimization tools were to be utilized. The point remains, however, whether the additional cost of this approach would be covered by the additional improvement of network performance. Secondly, the maximum theoretical potential that remains at this point is limited. This is further indication of the diminishing returns of real-time dynamic optimization or even continued differentiation.

Finding the analytical solutions identified as lacking in the first paper is of course still a valid pursuit. However, these results make clear that designing and implementing an FTN is feasible given the existing tools and technologies. Rule-of-thumb-based approximations seem to yield results that account for a vast majority of the available theoretical maximum of the potential available.

### *6.3.5 Partial implementation*

To implement foliated control in an existing network, it is reasonable to assume one way to proceed would entail a gradual stepwise implementation. Experiment results reveal that a so called 80/20 rule applies from a performance perspective, i.e., about 80% of the identified potential will be realized once 20% of the total volume of goods is available for foliated control. Saturation, i.e., the realization of the full potential, would be reached once about 50% of the total volume is available for foliated control.

These findings are explained by the circumstance that as the shares of the total volume of goods that are available for foliated control grow, the most underutilized units are the ones that first become available for redirecting through the hub. Simply put, when only a small portion of the goods in each relation is possible to reroute, only units containing corresponding amounts or less can be redirected, making the units removed from the DS layer grossly underutilized. As the share of goods available for foliation grows, the impact of each additional unit reallocated will diminish. Finally, surpassing some level, in this experiment around 50%, the available amount of goods for foliation ceases to be the deciding factor.

The transport unit (i.e., size of the truck) will have an impact, not on the shape of the outcome, but on its cutoff values and magnitude. For instance, given a fleet of smaller vehicles, the saturation point of the potential of Foliated Transportation Network would be reached before the levels of goods available for foliation reach 50%, and conversely, given a fleet of larger units, the need for the same in the system to realize the full potential of Foliated Transportation Network implementation will exceed the current level of 50%. Plainly, in a less efficient version of the system modeled, i.e., in one where the units are of smaller capacity and thus the version has less room for improvement based on the mechanism of foliation, the full potential would be feasible with a smaller share of goods available for foliated control, and vice versa.

There is another way to interpret these results and their practical implications for partial implementation. A network operator seeking to implement foliated control merely to tap into its efficiency potential would be able to do so without extensive initial investment in new identification and information applications. The share of the total amount of goods that would require enhanced control, tracking and identification would be limited enough for it to be possible to achieve with additional manual operations and contingency management. Furthermore, partial implementation reduces the workload of the hub. This means that a significant portion of the identified potential can be realized with a fraction of the additional handling cost that a full-scale implementation would entail.

# 7 Conclusions and future research

*In this section, the research questions posed are addressed based on the combined results of the studies. Also, the practical, theoretical implications of the results are discussed. Moreover, the need and direction for future research will be discussed.*

The results from the studies are synthesized in this section in order to address directly the answers to the research questions posed. Conclusions are presented in the form of practical and theoretical implications of those results. The contribution from each research question is summarized in Table 7-1.

## 7.1 RQ1: What are the challenges for designing an operationally feasible Foliated Transportation Network, and how can these challenges be overcome?

A number of specific issues regarding the distribution of goods between the different layers of the network, the combination and sequence of loads in units, efficiency of terminal operations and network optimization are some of the critical challenges to address. The identified challenges for designing an operationally feasible FTN pertain to transportation planning and control, transportation operations and transportation network optimization. Even though any specific design gap signals the need for continued research within the relevant theoretical domain, a operationally feasible model of FTN can be achieved utilizing approximations and rule-of-thumb-based approach.

These design challenges find their scientific base primarily within the theoretical domains of transportation management, information and communication science and mathematics. No single theoretical domain can be singled out as the most important one. The nature of the research called for here is cross-disciplinary and requires input from all three areas. The further development of FTN requires both new theoretical knowledge and the development of new applications within each area regarding all the identified dimensions.

However, bridging the existing gaps of theoretical knowledge and practical applications with design based on approximations and rule-of-thumb approach appears to produce sufficiently improved performance for it to be operationally feasible. The cost of achieving this potential, i.e., enhanced control and additional consolidation operations, needs to be determined on a case-by-case basis. Given the size of the identified potential, the benefits ought to outweigh the costs, especially regarding the changing cost structure of transportation operations.

## 7.2 RQ2: How would foliating a hub and spoke network over a direct shipment network affect the network performance?

The two network layers within an FTN allow for a more fitting match between allocated capacity and demand within the network. By allowing only perfect match between capacity and demand in the DS layer and additional consolidation for the remaining volumes within the HS layer, the overall network performance regarding capacity utilization and productivity will be improved.

However, unintended impacts such as increased network mean time and distance between nodes may arise as a result of the combined impact of the sub layers on the FTN if not handled properly. Similarly, other sought-after impacts may be unattained if appropriate design and operation measures are not introduced.

In order to counter the potential negative impacts and to pursue additional benefits not obtained by default, strategies ought to be devised. The design effort can be based in part on the gaps in knowledge and application identified in the previous question. More importantly, design based on approximations and rule-of-thumb approaches appears to suffice for unlocking the bulk of the performance improvement potential identified regarding the implementation of a feasible FTN.

It has been shown that a relatively large potential for performance improvement exists in implementing FTN. A switch from a pure DS to a foliated network would realize more than half of the maximum theoretical potential efficiency. This conclusion is valid for networks where FTN is applicable as discussed in section 6.3. As it has been pointed out, the detailed design and implementation of FTN is partially dependent on knowledge and applications that are currently lacking. However, the sensitivity analysis of the results and the relative success of approximations and rule-of-thumb-based control shows that the majority of the identified potential is attainable through adaptation of existing systems, procedures and applications. FTN outperforms the reference system even given high levels of error in planning and operations as well as when real-time dynamic control is substituted with static rules of thumb.

Two points regarding this robustness need to be made. Firstly, even though FTN outperforms the DS network, the improvement potential deteriorates proportionally to the increase of prognosis error. Secondly, the deterioration in the overall performance improvement potential of the system is almost perfectly linear. This means that the robustness mentioned is a result based entirely on the size of the original potential identified and is not an inherent property of FTN.

It is further concluded that the design, control and optimizing efforts regarding an FTN implementation do hold the key to additional improvement potential. It is difficult to precisely estimate the impact of the gaps still existing on the performance of the feasible operational design's performance. The size of the remaining potential as compared to the theoretical maximum alongside the relatively small impact of optimization and differentiation in control on the performance of FTN suggests that the point of diminishing returns might already have been reached.

## 7.3 Summary of results and contributions

The contribution from each research question is summarized in Table 7-1.

*Table 7-1 Contribution of each research question*

| RQ | Paper | Outcome | Contribution |
|---|---|---|---|
| **RQ1: What are the challenges for designing an operationally feasible Foliated Transportation Network, and how can these challenges be overcome?** | 1-5 | • Design challenges<br>• Gaps in knowledge and application<br>• Feasibility given the gaps | Design challenges regarding network planning and control, operations and optimization were identified as key. The majority of the maximum theoretical potential is reachable via approximations and rule-of-thumb approach. |
| **RQ2: How would foliating a hub and spoke network over a direct shipment network affect the network performance?** | 2-5 | • Network model<br>• Improvement potential<br>• Sensitivity<br>• Feasibility of rule-of-thumb-based control | A substantial performance improvement is likely to be achieved with the implementation of FTN. The results are found to be robust, and feasible implementation will likely lead to the realization of a major part of the identified potential. |

### 7.3.1 Practical implications

The substantial improvement potential identified based on the principle of FTN along with the indicated impact of partial implementation and rule-of-thumb-based foliated control point to a number of practical implications.

For one, the identified potential and its magnitude suggest low-hanging fruit benefits, i.e., even a partial or simplified implementation with contained costs and intrusions within existing systems and networks would be likely to yield a relatively high return in terms of improved physical performance or service level. Implementing FTN in an existing network would imply an oversight of the amount and use of terminal operations resources in the would-be hub terminal. These results indicate that even with less sophisticated and simple rule-of-thumb-based governing rules, FTN could outperform a traditional direct shipment network in the context at hand.

Secondly, a case can still be made for future research on the subject, primarily on design and implementation of new identification, information and communication technologies. The same platform of technologies would likely enable the production of additional value-adding services for the transport service buyers. Implementation of FTN would likely also provide new opportunities for the development of existing business models and operations.

Finally, the concept of foliated control here is applied to a very specific setting. The same principal concept can be applied to other areas where dimensions other than network structure are foliated, e.g., transportation modes, service providers, production segments etc. The collective results of this research provides support for one of the hypotheses on which this concept was built, namely, that the implementation of two systems in one, when compatible, can create a new system that outperforms any of its constituting parts in isolation.

### 7.3.2 Theoretical implications

The results presented above strongly indicate that the identified potential is an outcome of the combination of two network principles in a hybrid, mixed model network. These results gain additional validity as they concur with what was deduced from the existing theory. The further implication of this is that the same principle, i.e., the one of foliating two-system structures in a mixed model hybrid, would be valuable in contexts other than the current one.

Additionally, it could be concluded that these results are not universally valid outside of a specific spectrum of goods flows. Transpiration networks where the demand in each relation exceeds one truckload per day are ample candidates for FTN implementation. There also ought to exist an upper boundary where the expected returns from an FTN implementation would diminish. This upper boundary is not investigated within the scope of this thesis.

The empirical input in the studies and the convergence of the results with existing research strengthens the case for mixed model transpiration networks. Hopefully, these results will lay the foundation for a large-scale experiment or trial implementation to be included in the future continuation of this research.

## 7.4 Future research

Below, some reflections about the opportunities for future research are presented. Three main areas for future research are suggested: continued research aimed at further bridging the existing design gaps regarding the operational design of FTN, exploring new business opportunities that the implementation of FTN would enable and research on the broader application of the phenomenon of foliation. On a more general note, the existing theory on transportation efficiency also needs further development.

So far in this research, the increased goods handling at the hub that would result from foliation has not been addressed in depth. This is an important area to further explore as some of the cost for the performance improvement comes for this property. Coupling this issue with the fact that about 80% of the identified performance improvement potential can be realized if around 20% of the total volume of the goods in the network is available for foliation raises interesting questions. The construct for measuring performance of the network could be further developed to include the terminal operations and hence create opportunities for determining the level of foliation that would be optimal.

Moreover, the impact of the pickup and delivery operations on the networks' performance has not been addressed in this thesis as they fall outside of the scope of the study. From a practitioner point of view, these aspects need to be explored further. Another issue that requires further attention is that of the centralization/decentralization of the decision to route goods between the different network layers. Even though the studies performed evaluate the feasibility and potential of FTN without regard to this issue, from both a theoretical and practical point of view, being able to explain the impact of the different approaches on the final result is interesting enough to warrant future research.

The results presented in this thesis decisive enough to warrant experiments, pilot runs or partial implementation in a real-world system. Even though the avenues of future research mentioned above are possible to address using similar methodology as previously, e.g., simulation, the continued research would benefit from closer ties to empirical evidence. If future studies could be performed on a physical implementation of FTN, to some degree, in a real-world system, the results would probably have increased the relevance and validity.

The research presented in this thesis has been focused on the technical aspects of FTN. However, as argued in papers 1 and 5, the application of FTN would likely create opportunities for offering additional services to the shippers and as a result new business opportunities for the service providers. New business models need to be explored in this context.

In this thesis a special case of a general phenomenon has been studied. The idea of foliating more than one system in order to obtain a system that outperforms any of its constituting parts individually is a more general concept than the specific application studied here, i.e., the foliation of HS and DS network structures in a new and better performing system. The broader phenomenon of foliation is likely a very promising avenue for future research. The research could be conducted both on existing occurrences of the phenomenon and continued effort for designing new foliated systems.

The former approach would entail identifying the phenomenon of foliation in existing systems. In the special case of FTN here, the foliation is done on the routing of goods through a terminal network based on two different principles. Foliation could possibly be performed on other dimensions of a system, e.g., load units, traffic mode, business units, etc. The latter approach would likely entail identifying opportunities for designing new foliated systems much like the work that is done in this thesis.

Finally, this thesis makes clear the need for developing meaningful constructs for measuring transportation network performance that are focused on the carriers, their operations and the utilization of physical resources. A limited attempt is made here in order to accommodate the studies designed. A more comprehensive effort is needed for developing universal measures that would be possible to sustain via official statistics. This line of research is much broader than the scope of this thesis. Successes in doing so would likely have great implication for practitioners and researchers alike.

# References

Acharajee, P. (2000). Capacity Utilization of the Belly-Hold in Line-Based Air Traffic. Department of Transportation and Logistics. Göteborg, Chalmers university of Technology. **Lic. Eng. thesis**.

Arnäs, P. O. (2007). Heterogeneous Goods in Transportation Systems - A study on the uses of an object-oriented approach. Department of Technology Management and Economics. Gothemburg, Chalmers University of Technology. **Ph. D. thesis:** 262.

Aykin, T. (1995). "Networking Policies for Hub-and-Spoke Systems with Application to the Air Transportation System." Transportation Science **29**(3): 201--221.

Banks, J. (1998). Principles of Simulation. Handbook of Simulation - Principles, Methodology, Advances, Applications and Practice. J. Banks. New York, NY, John Wiley & Sons, Inc.**:** 3-30.

Banks, J., J. S. I. Carson, et al. (2001). Discrete-Event System Simulation. Upper Saddle River, NJ, Prentice Hall.

Bjeljac, R. and D. Lakobrija (2004). Foliated Hub Systems, the Necessary Requirements for Implementation at Schenker. Division of logistics and transportation. Gothenburg, Chalmers University of Technology. **M. Sc. thesis:** 70.

Björklund, M., U. Martinsen, et al. (2012). "Performance Measurements in the Greening of Supply Chains." Supply Chain Management: An International Journal **17**(1): 29-39.

Brandt, B. F., G. Bångman, et al. (2007). Infrastrukturplanering som en del av transportpoltiken - Underlag till inriktningsplanering 2010-2019. P.-Å. Vikman. Östersund, SIKA**:** 102.

Bryan, D. and M. O'Kelly (1999). "Hub-and-Spoke Networks in Air Transportation: An Analytical Review." Journal of Regional Science **39**(2): 275-295.

Buskhe, H. (1993). Resursutnyttjande i linjebaserad vägtrafik. Department of Transportation and Logistics. Göteborg, Chalmers University of Technology. **Lic. Eng. thesis**.

Caplice, C. and Y. Sheffi (1994). "A Review and Evaluation of Logistics Metrics." International Journal of Logistics Management **5**(2): 11-28.

Caplice, C. and Y. Sheffi (1995). "A Review and Evaluation of Logistics Performance Measure Systems." International Journal of Logistics Management **6**(1): 61-74.

Caputo, A. C., L. Fratocchi, et al. (2005). "A Framework for Analysing Long-range Direct Shipping Logistics." Industrial Management & Data Systems(105): 876-899.

Carson, J. S. I. (2005). Introduction ot Modeling and Simulation. Winter Simulation Conference. M. E. Kuhl, N. M. Steiger, F. B. Armstrong and J. A. Joines. Orlando, FL**:** 16.

Chan, F. T. S. (2003). "Performance measurement in a supply chain." International Journal of Advanced Manufacturing Technology **21**: 354-548.

Chantzara, M. and M. Anagnostou (2006). MVRC Heuristic for Solving the Multi-Choice Multi-Constraint Knapsack Problem. International Conference on Computational Science. V. N. Alexandrov. University of Reading, UK, Springer-Verlag**:** 579-587.

Chow, G., T. D. Heaver, et al. (1994). "Logistics Performance: Definition and Measurement." International Journal of Physical Distribution & Logistics **24**(1): 17-28.

Clark, R. L. and K. N. Gourdin (1991). "Measuring the Efficiency of the Logistics Process." Journal of Business Logistics **12**(2): 17-33.

Clark, T. D., G. W. Arnett, et al. (1986). "The Use of Simulation in System Design: A Case Study Using the Precision Location Strike System." Journal of Operations Management **6**(3–4): 363-379.

Cowie, J. (2010). The Economics of Transport - A theoretical and applied perspective. Oxon, UK, Routledge.

Crainic, T. (2003). Long-haul Freight Transportation. Handbook on Transportation Science. R. W. Hall. New York, Kluwer Academic Publishers.

Crainic, T. G. (2002). A Survey of Optimisation Models for Long-Haul Freight Transportation. Handbook of Transportation Science. R. W. Hall, Kluwer.

Crainic, T. G., N. Ricciardi, et al. (2004). "Advanced Freight Transportation Systems for Congested Urban Areas." Transport Research Part C. **12**: 119-137.

de Koster, R. B. M., T. Le-Anh, et al. (2004). "Testing and Classifying Vehicle Dispatching Rules in Three Real-World Settings." Journal of Operations Management **22**(4): 369-386.

Denyer, D., D. Tranfield, et al. (2008). "Developing Design Propositions Through Research Synthesis." Organization Studies **29**(3): 393-413.

Denzler, D. R., W. J. Boe, et al. (1987). "An Experimental Investigation of FMS Scheduling Rules under Uncertainty." Journal of Operations Management **7**(1–2): 139-151.

Dobson, G. and P. J. Lederer (1993). "Airline Scheduling and Routing in a Hub-and-Spoke System." Transportation Science **27**(3): 281-297.

Dubois, A. and L.-E. Gadde (2002). "Systematic Combining: an abductive approach to case research." Journal of Business Research **55**: 553-560.

Dyckhoff, H. (1990). "A Typology of Cutting and Packing Problems." European Journal of Operational Research **44**: 145-159.

Eisenhardt, K. M. (1989). "Building Theories from Case Study Research." The Academy of Management Review **14**(4): 532-550.

European Commission, D. E. a. T. (2006). Keep Europe Moving - Sustainable mobility for our continent. Luxembourg, Office for official publication for the european communities.

Flick, U. (2006). An Introduction to Qualitative Research. London, Sage Publications.

Flodén, J. (2007). Modelling Intermodal Freight Transport - The Potential of Combined Transport in Sweden. Department of Business Administration, School of Business, Economics and Law. Göteborg, Göteborg University. **Ph. D. thesis:** 272.

Gehring, H., K. Menschner, et al. (1990). "A Computer-Based Heuristic for Packing Pooled Shipment Containers." European Journal of Operational Research **44**: 277-288.

Goldsman, D. (2007). Introduction to Simulation. Winter Simulation Conference. S. G. Henderson, B. Biller, M.-H. Hsiehet al. Washington, DC**:** 26.

Hart, C. (1998). Doing a literature review - Releasing the social science imagination. London, UK, Sage.

Heavey, C. and J. Ryan (2006). Process Modelling Support for the Conceptual Modelling Phase of a Simulation Project. Winter Simulation Conference. L. F. Perrone, F. P. wieland, B. G. Lawson, D. M. Nicol and R. M. Fujimoto. Monterey, CA**:** 801.

Hellström, D. (2007). On Interactions between Packaging and Logistics - Exploring Implications of Technological Developments. Department of Design Sciences. Lund, Lund University. **Ph. D. thesis**.

Hiller, F. S. and G. J. Lieberman, Eds. (1995). Introduction to Operations Research. McGraw-Hill Series in Industrial Engineering and Management Science. New York, McGraw-Hill.

Hultén, L. A. R. (1997). Contatiner Logistics and Its Management. Department of Transportation and Logistics. Göteborg, Chalmers University of Technology. **Ph. D. thesis**.

Hultkrantz, O. (1999). Networks in Road-Based Traffic. Department of Transportation ans Logistics. Göteborg, Chalmers university of Technology. **Lic. Eng. thesis:** 59.

Kellerer, H., U. Pferschy, et al. (2004). Knapsack Problems. Berlin, Germany, Springer.

Kelton, D., R. P. Sadowski, et al. (2010). Simulation with Arena. New York, McGraw-Hill.

Kleinsorge, I. K., P. B. Schary, et al. (1989). "Evaluating Logistics Decisions." International Journal of Physical Distribution & Logistics **19**(12): 3-14.

Kleinsorge, I. K., P. B. Schary, et al. (1991). "The shipper-carrier partnership: A new tool for performance." Journal of business logistics **12**(2): 35-57.

Lammgård, C. (2007). Environmental Perspectives on Marketing of Freight Transports - The Intermodal Road-Rail Case. Department of Business Administration, School of Business, Economics and Law. Göteborg, Göteborg University. **Ph. D. thesis:** 290.

Law, A. (2001a). How the Expertfit Distribution -Fitting Software Can Make Your Simulation Models More Valid. Winter Simulation Conference. B. A. Peters, J. S. Smith, D. J. Medeiros and M. W. Rohrer. Arlington, VA**:** 256-261.

Law, A. (2001b). How to Build Valid and Credible Simulation Models. Winter Simulation Conference. B. A. Peters, J. S. Smith, D. J. Medeiros and M. W. Rohrer. Arlington, Va**:** 22.

Law, A. and D. Kelton (2000). Simulation Modeling & Analysis. Singapore, McGraw-Hill.

Leemis, L. M. (2004). Building Credible Input models. Winter Simulation Conference. R. G. Ingalls, M. D. Rossetti, J. S. Smith and B. A. Peters. Washington, DC**:** 29.

Lincoln, Y. S. and E. G. Guba (1984). Naturalistic Inquiry. Beverly Hills, Ca, Sage Publications.

Lindskog, M. (2008). Systems Thinking - Hard(ly) Core? NOFOMA, Helsinki, Finnland.

Liu, J., C.-L. Li, et al. (2003). "Mixed Truck Delivery Systems with both Hub-and-Spoke and Direct Shipment." Transport Research Part E. **39**: 325-339.

Lodi, A., M. Silvano, et al. (2002). "Two-Dimensional Packing Problems: A Survay." European Journal of Operational Research **141**: 241-252.

Lumsden, K. (1995). Economy of Transportation Systems. Lund, Studentlitteratur.

Lumsden, K. (2006). Fundamentals of Logistics. Lund, SWE, Studentlitteratur.

Lumsden, K., F. Dallari, et al. (1999). "Improving the Efficiency of the Hub and Spoke System for the SKF European Distribution Network." International Journal of Physical Distribution & Logistics **29**(1): 50-64.

Manheim, M. L. (1979). Fundamentals of Transportation Systems Analysis. Cambridge, MA, MIT Press.

Martello, S. and P. Toth (1990). Knapsack Problems - Algorithms and Computer Implementation. Chichester, England, John Wiley & Sons.

McIntyre, K., H. A. Smith, et al. (1998). "Logistics Performance Measurement and Greening Supply Chains: Diverging Mindsets." International Journal of Logistics Management **9**(1): 57-68.

McKinnon, A. C. (2006). "Life without Trucks: The impact of a Temporary Disruption of Road Freight Transport on a National Economy." Journal of business logistics **27**(2): 227-250.

McKinnon, A. C. (2009). "Benchmarking Road Freight Transport - Review of a Government-Sponsered Programme." Benchmarking: An International Journal **16**(5): 640-656.

McKinnon, A. C. (2010). European Freight Transport Statistics: Limitations, Misinterpretations and Aspirations. 15th ACEA scientific advisory goup meeting.

McKinnon, A. C. and Y. Ge (2004). "Use of a Synchronised Vehicle Audit to Determine Oppertunities from Improving Transport Efficiency in a Supply Chain." International Journal of Logistics: Research and Application **7**(3): 219-238.

Mentzer, J. T. and B. Konrad (1991). "An Efficiency/Effectiveness Approach to Logistics Performance Analysis." Journal of business logistics **12**(1): 33-62.

Miles, M. B. and A. M. Huberman (1994). Qualitative Data Analysis. London, Uk, Sage.

Miljödepartementet (2008). Regeringens Klimatpolitik. Miljödepartementet. Stockholm, Regeringskansliet.

Nanos-Pino, J., G. Carrera-Gómez, et al. (2005). "Technical Efficiency of Road Haulage Firms." Transportation Research Record: Journal of the Transportation Research Board **1906**: 26-32.

O'Kelly, M. (1998). "A Geographer's Analysis of Hub-and-Spoke Networks." Journal of Transportation Geography **6**(3): 171-186.

O'Kelly, M. and D. Bryan (1998). "Hub Location with Flow Economies of Scale." Transport Research B. **32**(8): 605-616.

O´Connor, K. (2009). "Global City Regions and the Location of Logistics Activity." Journal of Transport Geography **In press**(doi:10.1016/j.jtrangeo.2009.06.015).

Persson, P.-O. (2006a). Identification Technologies in Transportation - In the Context of Foliated Transportation Networks. 29th Australian Transport Research Forum. Gold coast, Australia.

Persson, P.-O. (2006b). Traceability of Goods in a Transportation Network and the Potential Impact on Logistics And Transportation Services. 4th International Logistics and Supply Chain Congress. Izmir, Turkey.

Persson, P.-O. and K. R. Lumsden (2006). Foliated Transportation Networks. Logistics Research Network, Newcastle, United Kingdom.

Persson, P.-O. and J. Waidringer (2006). Foliated Transportation Networks. 18th Annual NOFOMA Conference, Oslo, Norway.

Ploos van Amstel, R. and G. D'hert (1996). "Performance Indicators in Distribution." International Journal of logistics management **7**(1): 73-82.

Regeringen (2008). Regeringens proposition 2008/09:50 - Ett lyft för forskning och innovation. D. o. education. Stockholm, Sveriges Regering.

Riege, A. M. (2003). "Validity and Reliability Tests in Case Study Research: A Literature Review with "Hands-on" Applications for Each Research Phase." Qualitative Market Research: An international Journal **6**(2): 75-86.

Robinson, S. (2006). Conceptual Modeling for Simulation: Issues and Research Requirements. Winter Simulation Conference. L. F. Perrone, F. P. Wieland, B. G. Lawson and R. M. Fujimoto. Monterey, CA**:** 792.

Rodrigue, J.-P. (1999). "Globalization and Synchronization of Transport Terminals." Journal of Transport Geography **7**: 255-261.

Rodrigue, J. P., C. Comtois, et al. (2006). The geography of Transport Systems. New York, Routledge.

Ross, D. F. (1996). Distribution Planning and Control. Boston , MA, Kluwer Academic Publishers.

Roy, J. and T. G. Crainic (1992). "Improving Intercity Freight Routing with a Tactical Planning Model." Interfaces **22**(3): 31-44.

Sargent, R. (2004). Validation and Varification of Simulation Models. Winter Simulation Conference. R. G. Ingalls, M. D. Rossetti, J. S. Smith and B. A. Peters. Washington, D.C.**:** 17.

Saxin, B., C. Lammgård, et al. (2005). Meeting the Demand for Goods Transports - Identification of Flows and Needs among Swedish Companies. 17th Annual NOFOMA Conference, Copenhagen, Denmark, Department of Operations Management, Copenhagen Business School.

Scholl, A., R. Klein, et al. (1996). "BISON: A Fast Hybrid Procedure for Exactly Solving the One-Dimentional Bin Packing Problem." Computer Ops Res. **24**(7): 627-645.

Shafer, S. M. and T. L. Smunt (2004). "Empirical Simulation Studies in Operations Management: Context, Trends, and Research Opportunities." Journal of Operations Management **22**(4): 345-354.

Shaw, S., D. B. Grant, et al. (2010). "Developing Environmental Supply Chain Performance Measure." Benchmarking: An International Journal **17**(3): 320-339.

Shepherd, C. and H. Günter (2005). "Measuring supply chain performance: Current research and future directions." International journal of productivity and performance management **55**(3/4): 242-258.

Silvano, M., D. Pisinger, et al. (2000). "The Three-Dimensional Bin Packing Problem." Operations Research **48**(2): 256-267.

Simon, H. A. (1996). The Science of the Artificial. Cambridge, Ma, MIT University Press.

Sjöstedt, L. (2005). A Conceptual Framework for Analysing Policy Maker's and Industry Roles and Perspectives in the Context of Sustainable Goods Transportation Barriers to Sustainable Transport, Spoon Press.

Sommar, R. and J. Woxenius (2007). "Time Perspectives on Intermodal Transport of Consolidated Cargo." European Journal of Transportation and Infrasturcture Research **7**(2): 163-182.

Stainer, A. (1997). "Logistics - A Productivity and Performance Perspective." Supply Chain Management **2**(2): 53-62.

Stefansson, G. (2006). "Collaborative Logistics Management and the Role of Third-Party Service Providers." International Journal of Physical Distribution & Logistics **36**(2): 76-92.

Stefansson, G. and H. Sternberg (2011). "Supply Chain Inefficiencies: Transportation - the Forgotten Factor." Journal of business logistics.

Stern, N. (2007). The Economics of Climate Change: The Stern Review, Cambridge University Press

Sternberg, H. (2008). "Transportation Visability and Information Sharing - A Case Study of Actors' Requirements." World Review on Intermodal Transportation **2**(1): 54-72.

Sternberg, H. (2011). Waste in Road Transport Operations - Using Information Charing to Increase Efficiency. Technology Management and Economics. Gothemburg, Sweden, Chalmers University of Technolgy. **Ph. D. thesis**.

Sternberg, H. and G. Stefansson (2007). Freight Transport Operations Management - Coordination and Information Needs from a Logistics Servicee Provider Perspective. The 5:th International Conference on Supply Chain Management and Information Systems. Melbourne, Australia.

Taha, T. and D. Taylor (1994). "An Integrated Modeling Framework for Evaluating Hub-and-Spoke Networks in Truckload Trucking." Logistics and Transportation Review **30**(2): 141-166.

Taylor, C. (2007). "Are You Ready?" CSCMP's Supply Chain Quarterly(2): 41-50.

Taylor, G. D., S. Harit, et al. (1995). "Hub and Spoke Networks in Truckload Trucking: Configuration, Testing and Operational Concerns." Logistics and Transportation Review **31**(3): 209-238.

Trafikanalys (2011). Statistikunderlag rörande tomtransporter och fyllnadsgrader. B. Saxton. Stockholm, Trafikanalys.

Trafikverket (2011). Redovisning av: Regeringsuppdrag att analysera och föreslå tgärder för minskad tomdragning och ökad fyllnadsgrad. Trafikverket.

Travisano, J. J. (1996). Lessons Learned: The Object-Oriented Design of StarView. Astronomical Data Analysis Software and Systems V, A.S.P. Conference Series.

Waidringer, J. (2001). Complexity in Transportation and Logistics Systems : An Integrated Approach to Modelling and Analysis. Technology Management and Economics. Göteborg, Chalmers University of Technology. **Ph. D. thesis:** x, 125 s.

van Aken, J. E. (2004). "Management Research Based on the Pradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules." Journal of Management Science **41**(2): 219-246.

van Donselaar, K., K. Kokke, et al. (1998). "Performance Measurements in Transportation and Distribution Sector." International Journal of Physical Distribution & Logistics **28**(6): 434-450.

Wandel, S., C. Ruijgrok, et al. (1992). Relationships Among Shifts in Logistics, Transport, Traffic and Informatics. Lund, Studentlitteratur.

Williams, J. D. (1996). "Managing Iteration in OO Projects." Computer **29**: 39-43.

Vinnova (2004). Nationell strategi för transportrelaterad FUD. P. Eriksson. Stockholm, Sweden, Vinnova, swedish Governmental Agency for Innovation Systems**:** 39.

Woxenius, J. (2007). "A Generic Framework for Transportaion Network Design: applications and treatment in intermodal freight transport literature." Transport Reviews **27**(6): 733-749.

Yin, R. K. (2003). Case Study Research - Design and Methods. Thousend Oakes, Ca, Sage Publications.

Zäpfel, G. and M. Wasner (2002). "Planning and Optimization of Hub-and-Spoke Transportation Networks of Cooperative Third-Party Logistics Providers." International Journal of Production Economics **78**: 207-220.