

Logistics Chains in Freight Transport Modelling

Igor. Y. Davydenko

Delft University of Technology

Research presented in this book has been funded by
the Netherlands Organization for Scientific Research (NWO) and sponsored by
the Netherlands Organization for Applied Scientific Research (TNO)

Logistics Chains in Freight Transport Modelling

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. Ir. K.Ch.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 26 mei om 12:30 uur

Door

Igor Yevgenovych DAVYDENKO, PDEng, M.Sc.
geboren te Kharkiv, Oekraïne

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. L.A. Tavasszy

Samenstelling van de promotiecommissie:

Rector Magnificus

Prof. dr. ir. L.A. Tavasszy

Voorzitter

Technische Universiteit Delft, promotor

Onafhankelijke leden:

Prof. dr. A. Koike

Prof. dr. A.C. McKinnon

Prof. dr. ir. E. de Romph

Prof. dr. F. Witlox

Prof. drs. C.J. Ruijgrok

Jun. prof. dr. H. Friedrich

Prof. dr. ir. P.M. Herder

Kobe University

University of Hamburg

Technische Universiteit Delft

University of Gent

University of Tilburg

Technical University of Darmstadt

Technische Universiteit Delft

This thesis is the result of a Ph.D. study carried out on a part-time basis from 2009 to 2014 at Delft University of Technology, Faculty of Technology, Policy and Management, Section Transport and Logistics

TRAIL Thesis Series no. T2015/10, the Netherlands Research School TRAIL

TRAIL

P.O. Box 5017

2600 GA Delft

The Netherlands

Phone: +31(0) 15 278 6046

E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-190-5

Copyright © 2015 by Igor Y. Davydenko

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the author.

Printed in the Netherlands

Preface

Having background in logistics management systems, I first faced the issue of freight transport modelling in 2008, when we at TNO had to determine the impact of long and heavy vehicles on the functioning of European transport system. At that time I realized two things. First, it was the realization that the aggregate logistics modeling was much less developed than the company-level optimization techniques. Second, I found the subject of freight transport modelling to be really interesting.

A year later, in 2009 I decided to pursue an opportunity to conduct PhD research in the modeling of logistics and warehousing at the macro level. This would not be possible without my promoter's decision to take me on this project. I am deeply thankful to Lori Tavasszy for that decision and for his continuous inspiring and support in this long research effort!

My employer, TNO, has played an important role in the organization of this PhD research project, allowing me to continue developing as research scientist and consultant, while doing PhD part-time at the same time. I am thankful to Arie Bleijenberg and Kees Verweij, who agreed to this construction. My thanks go to Olga Ivanova, Hans Quak, Jaco van Meijeren and Kees Ruijgrok with whom I had the pleasure to work on the content and publish. I was lucky to be surrounded by inspiring colleagues, who helped and motivated me during this journey!

I reserved special thanks to the wonderful colleagues at Statistics Netherlands, the CBS. I am really grateful to Chris de Blois and Peter Smeets for arranging access to the Dutch road transport data. These data are the cornerstone element of this thesis. I would like to thank Thieu van Kasteren, Mathijs Jacobs and many other people at CBS who facilitated and contributed to my research. This collegial relationship opens up new research opportunities that I hope we will discover together.

International applications of the logistics chain model would not be possible without help of Mark Thissen, Hanno Friedrich and Atsushi Koike, who provided me with the European, German and Japanese data respectively. It really enriched my research and let the modeling

techniques to be tested outside of the Netherlands, giving the model an international reach. Thank you for that!

I was lucky to be surrounded by wonderful fellow researchers and friends. I enjoyed the company of Mo Zhang and Ronald Halim; Maureen Lankhuizen had bright ideas on the link to economy; Nilesh Anand was always here to have a talk and to give a push towards meditation. Paul van de Lande brightened up many evenings and ensured that the Dutch version of the summary is readable. Thank you friends!

And finally my family. Parents, you encouraged me not to stop and always go further. Viktoria, you made it possible by your unquestioning love and support. Anna, you were born in the middle of this endeavor and filled my life with joy. Love you all.

Igor Davydenko,
Den Haag, April 2015.

Table of Contents

- 1. Introduction and problem definition 1
 - 1.1. Policy relevance of the modeling on interregional freight and logistics flows 2
 - 1.2. Research questions 5
 - 1.3. Research approach 7
 - 1.4. A note on Research Background 9
 - 1.5. Concise guide to this thesis 10
- 2. Positioning of Aggregate Logistics Modeling 13
 - 2.1. Function of distribution, tradeoffs and optimization 13
 - 2.1.1. Trade-offs involved in distribution-related choices 14
 - 2.1.2. Logistics optimization at the company level 17
 - 2.2. Review of relevant transport modeling literature 20
 - 2.2.1. Review literature 20
 - 2.2.2. Aggregate models 22
 - 2.2.3. Disaggregate models 23
 - 2.3. Detailed review of closely related approaches 26
 - 2.3.1. SMILE / SMILE+ model 27
 - 2.3.2. Physical Distribution Channel Choice Model (PDCCM) 30
 - 2.3.3. Integrated Model for Estimating Emissions from Freight Transport (IMEEFT) ... 32
 - 2.3.4. SYNTRADE model 34
 - 2.3.5. EUNET 2.0 38
 - 2.4. Literature based logistics model requirements 39
 - 2.5. Concluding remarks 42
- 3. Model Specifications 43

3.1. Introduction	43
3.2. Model requirements	43
3.3. Overall model design.....	46
3.3.1. Data-driven choices	48
3.3.2. One echelon and multi-echelon distribution structures	49
3.3.3. Cost functions and location attractiveness.....	50
3.4. Model structure.....	51
3.5. Gravity Model.....	53
3.5.1. Essence of Gravity Models	54
3.5.2. Specification of Separately Calibrated Gravity Model.....	56
3.6. Logistics Chain Model.....	62
3.6.1. Combined Model Calibration	65
3.7. Alternative definition of the logistics chain model	66
3.8. Conclusions on model specification	68
4. Data	71
4.1. Introduction	71
4.2. Data for the Netherlands.....	72
4.2.1. Background information	72
4.2.2. Data structure and properties of location-related information	73
4.2.3. Quality of the location type variable.....	77
4.2.4. Year on year changes in location-type related flows	77
4.2.5. Sample sizes.....	78
4.2.6. OD flows and location attractiveness analysis.....	79
4.2.7. Zero-volume relations	82
4.3. Link between employment and distribution volumes.....	83
4.3.1. Disaggregation at commodity level and smaller regions	87
4.3.2. Conclusions on sectorial employment and distribution volumes	90
4.4. German Data.....	91
4.5. European Data	93
4.6. Discussion and Conclusions	96
5. Model Implementation and Estimation	99
5.1. Separately Calibrated Gravity Model.....	99
5.1.1. Program implementation.....	100
5.1.2. SCGM model runs: looking for the properties of different types of flow	102

5.1.3. Estimation quality for the various flow types	103
5.1.4. All flows put together	104
5.1.5 Distance Decay	105
5.1.6. Estimation of sub-flows	107
5.1.7. Transport demand price elasticity	109
5.1.8. Regional transport production and attraction.....	110
5.1.9. Conclusions on the Gravity Model	111
5.2. Logistics Chain Model for the Netherlands.....	113
5.2.1. Estimation of the model's parameters.....	113
5.2.2. Model implementation and calibration	115
5.2.3. Calibration procedure.....	116
5.2.4. Results of the base model calibration	118
5.2.5. Base model calibration: estimated model parameters.....	124
5.2.6. Alternative logistics chain model implementation and estimation.....	126
5.2.7. Conclusion on logistics chain model estimation.....	130
5.3. European Logistics Chain Model application	132
5.3.1. Implementation of the European Logistics Chain Model.....	133
5.3.2. European Logistics Chain Model calibration	135
5.3.3. European model calibration	136
5.3.4. Conclusions on the European Logistics Chain Model.....	139
5.4. LCM Implementation and estimation on German Data	140
5.4.1. LCM calibration on German Data	141
5.4.2. Conclusions for LCM application on German Data	144
5.5. Discussion and conclusions	144
6. Logistics Chain Model Applications.....	147
6.1. Introduction	147
6.1.1. Main areas of LCM application (spatial logistics, large-scale models).....	147
6.1.2. Influence of the decision variables	148
6.1.3. LCM output indicators for scenario analysis	150
6.2. Logistics Sprawl in Randstad	151
6.2.1. Logistics Sprawl Scenario Elaboration.....	152
6.2.2. Logistics Sprawl Scenario Outcomes	155
6.2.3. Discussion on Quantitative Evaluation of Logistics Sprawl.....	158
6.3. Elasticity of flows for road transport costs.....	158

6.3.1. Elasticity estimated by the LCM in the case of logistics sprawl	159
6.3.2. Discussion on Gravity Model and LCM Elasticity Estimations.....	160
6.4. Discussion.....	160
7. Conclusions	163
7.1. Conclusions on available knowledge	164
7.2. Conclusions on the mathematical model formulation	165
7.3. Conclusions on data requirements and data availability	166
7.4. Conclusion on model implementation, estimation and application.....	167
7.5. Outline of further research.....	168
Summary	169
Samenvatting.....	177
References	185
Appendix 1: NSTR goods classification	194
Appendix 2	195
TRAIL Thesis Series.....	196
About the author.....	198

1. Introduction and problem definition

The research presented in this PhD thesis has been motivated by the fact that the Netherlands, and the Randstad region in particular, are affected by the large transport flows and extensive operations of the logistics sector. These operations create welfare for those people who work in the sector, who own the companies, and for the Dutch and European societies as the whole. The strong transport and logistics sector has also negative impact on the public infrastructure, air pollution, greenhouse gas emissions, causes accidents and other negative effects on the societal welfare.

The strength of the Dutch logistics sector, the benefits and negative effects that it brings, require a thorough understanding of the logistics systems at the regional level. A quantitative tool, which is capable of an analysis of freight flows via the logistics and distribution systems in an empirically valid way, would be of substantial help for the decision makers at the national and regional levels. The quantitative tool (or a model) should also be able to study the system in a scenario-wise way, such that the impact of certain changes in the system can be quantified, studied and assessed.

The Dutch logistics system is complex, with a large number of various national and international players. Understanding the logistics system in a holistic way requires an appropriate degree of aggregation and abstraction, as it would be unpractical to study each company and each private decision maker in the Netherlands: there are too many for this approach to be feasible. Therefore, this research concentrates on regions, as opposed to the individual companies, and describes the transport demand patterns realized by the logistics systems, as opposed to optimization of those systems, see FIGURE 1.1 for a graphical positioning of the focus of this research. This research is also limited to the transport flows realized by the road transport mode. This thesis looks at the core product of the logistics sector: linking production and consumption together, or how the goods are shipped from the places where they are produced to the places where they are consumed.

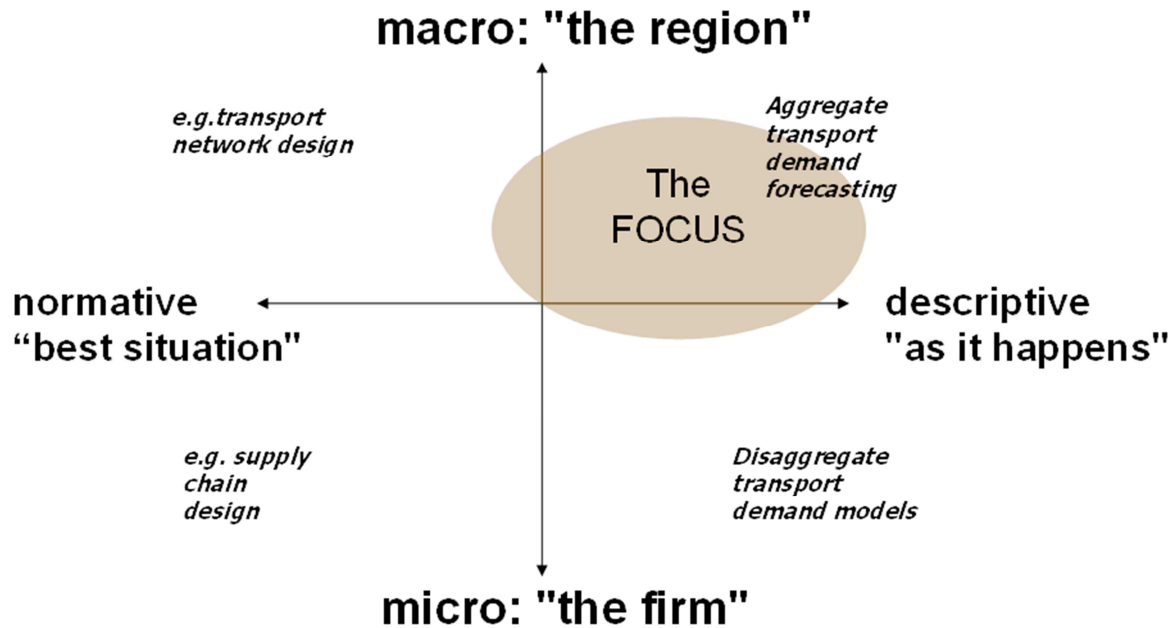


FIGURE 1.1. Focus of the research presented in the thesis

1.1. Policy relevance of the modeling on interregional freight and logistics flows

Considering policy relevance of the research presented in this PhD thesis, four distinct policy areas can be identified. These areas are as follows.

1. The problem of freight generation, resulting in infrastructure load, pollution and externalities.
2. The problem of employment in the wholesale, distribution and warehousing sectors
3. Spatial planning and spatial patterns of distribution and warehousing facilities
4. Transport and distribution structures linking the Dutch mainports and hinterland.

1. Freight generation, infrastructure load, pollution and external effects. The problem of pollution, CO₂ emissions and other external effects is one of the most important problems in the realm of transportation policies. The European Commission has issued own rules and guidelines on the policy impact assessment (European Commission Impact Assessment 2014). The logistics chain model presented in this thesis provides a more adequate estimation for the flows generated by the distribution facilities than the dominant modeling practices (e.g. TRANS-TOOLS). Therefore, the impact on infrastructure, pollution and road accidents can be estimated more accurately than it had been possible before incorporation of the logistics into the freight modeling frameworks.

2. Employment in the wholesale and distribution and warehousing sectors. Employment in the warehousing and distribution, and in the related sector of wholesale, is an important factor in the Netherlands. The sector provides for low education level jobs, such as order picking,

fork lift truck operators, as well as for administrative labor. Employment in the warehousing can be considered as a stabilizing factor at the regional level, as the facilities are not moved around or often relocated. Chapter 4 shows the relationship between distribution throughput and employment and Chapter 6 shows the case of logistics sprawl in the Randstad region, a phenomenon of a spatial spread from relatively concentrated logistics clusters in the 1970s to geographically more decentralized patterns around large urban areas. These two examples provide a tool for assessment of spatial distribution of employment in the wholesale, and warehousing and distribution sectors of the economy.

3. *Spatial planning and spatial patterns of distribution and warehousing facilities.* The Logistics Chain Model developed in this thesis is capable of assessment of the spatial organization of the warehousing and distribution facilities. An illustration of this capability is the logistics sprawl case of Chapter 6. The policy makers at the national and regional levels may be interested in making some regions more attractive for logistics facilities, while some other regions make consider this industry to be a burden. The model provides an empirically grounded tool for quantitative assessment of the measures necessary to take in order to achieve the desired result in respect to spatial organization of the distribution, thus opening up the area of policies on spatial organization of warehouses and distribution facilities.

4. *Transport and distribution structures linking the Dutch Mainports and hinterland.* The sea port of Rotterdam and the Amsterdam's Schiphol Airport are defined as the Dutch Mainports, with a corresponding place in the society and regulations, see for instance van Gils et al. (2009). The port authorities are interested in how goods are transported from the port to the goods' final consumption or rework points. This understanding is important for the port development plans in respect to the competing ports (e.g. ports of Hamburg and Antwerp), and for the extra services potentially provided at the ports. The competitive position of the sea ports is influenced by the organization of the distribution infrastructure, which serves the flows from the ports, grounding the flows to Rotterdam and not to competing ports. Consequently, some extra services can be provided at the port's grounds, such as distribution and value added logistics. By being able to analyze logistics chains linking the port to the final destination of the goods, the port authority can better position services of the port, make rational plans for ground allocation at the port, and better protect goods flows via the port from the competition.

To make the results of research in the field of freight transport modeling practically applicable and interesting for the target audience of the potential users, it is necessary to understand the target users of the research results and look at their needs. Four major classes of the (potential) users of the research on incorporation of logistics choice modeling into freight modeling tools have been identified. They are the following:

1. Government bodies. This includes national, international and regional governments, which are involved in policy making and policy execution. These bodies are interested in decisions on new infrastructure projects, facilitation of less polluting transport solutions (modal shift, technological innovations within the transport modes, and most relevant, organization of logistics chains), regional development and regional externalities of transport, future transport policies (e.g. whether to harmonize European policy in respect to long and

heavy vehicles). There are a large number of questions that freight transport models can help answer for the government bodies of different levels. An overview of the main EU-level policy areas related to the road transport can be found in EC publications (e.g. European Commission, 2012); at the national Dutch level, the questions on congestion, safety, competitive position of the transport and logistics sector play the main role (e.g. NTG Beleid goederenvervoer, 2014). As a part of transport modeling instrumentation, the logistics chain can help better answering the questions of policymakers. As a standalone model, it is relevant for the the questions of employment in distribution sector, competitiveness of the sector, spatial organization of logistics chains, of which the logistics sprawl case is an example.

2. Private business. The target group consists of large (multinational) businesses, which are interested in calculation of “what-if” scenarios for decisions on strategic questions. The businesses may need decision support systems that can accurately model future scenarios in respect to transport and freight flows. For instance, a big company might be interested in congestion situation, assessment of regional attractiveness for distribution centers, as factors in search of an optimal location for a distribution center or a warehouse. Large businesses are interested in development of future scenarios, which include strategic decisions on location of the facilities and the structure of transport chains (e.g. DHL, 2012). Another example is the port authority who may be interested in what happens to the goods transshipped through the port it manages, and how the logistics chains look like for the purpose of better positioning of the port with respect to value added services at the port premises, as well as understanding of the functioning of the hinterland connections. There is also a question of the future need for port capacity and the possibility of a gridlock in hinterland connections, which will impact the pattern of distribution chains, see for instance BCG (2007).

3. Non-Governmental Organizations (NGO). This type of organizations is very similar to the governments in their modeling needs, though goals that they pursue are different from those of the governments. These are related to argument preparation for certain specific (lobby) activities and setting up of the public agenda. For instance, the Brussels based Transport and Environment promotes transport policies based on the principles of minimization of the harmful impacts and maximization of effectiveness of resource use; the European Intermodal Association (EIA) aims to deliver concrete tools to improve, among others, logistics infrastructure of which distribution and warehousing play a role.

4. Research and consultancy. This group of users will work directly with the modeling methods presented in this thesis. The researchers and consultants will be the major link between the modeling world and “real” users of the modeling outcomes, such as the governments and private businesses. For example, The Netherlands Ministry of Infrastructure and the Environment have developed a roadmap for strategic traffic and transport model development, which includes a logistics model, see Rijkswaterstaat (2012). Transport modeling requires specific skills that the governments and private organizations rarely have in house, but researchers and consultants continually work on acquiring innovations in their knowledge area. Thus the consultancy and research bureaus will further expand application areas and assist the end users with actual modeling work.

1.2. Research questions

In the first year of the reported in this thesis PhD research project, a number of research questions have been defined, thus providing a scientific and practical mandate for the execution of the project. The following list formulates the questions of this PhD project.

RQ 1: How can a freight model be designed and implemented with the aim to estimate empirically valid transport flows necessary to ship the goods from production to consumption locations, emphasizing a proper modeling of flows related to warehouses and distribution centers?

RQ 2: What are the alternative approaches to modeling of distribution structures in a macro level freight model?

RQ 3: What are data requirements for logistics model of *RQ 1* and what are the data availability, quality and update policies?

RQ 4: How do changes in transport and warehousing costs influence interregional freight flows?

In practical terms, the research goal of this project has been to design and construct a logistics model capable of predicting spatial locations of distribution and warehousing on the one hand, and also capable of prediction of transport flows that production-consumption relations generate with an emphasis on distribution centers and warehousing as the nodes in production-consumption chains. No less important goal of the research is the empirical validity of the model, which is the scientific core of this research.

Logistics model is a broad notion: there are company level models, which deal with the business needs of the companies. The company level models are micro normative models such as the RESPONSE model (TNO, 2004)), which optimize properties of logistics networks controlled by a company or a collection (consortium) of collaborating companies. The main question that these models answer is on how to reduce the total logistics costs while maintaining the desired customer service level. The outcome of this type of logistics modeling is a representation of “an ideal reality”, not a present state of networks, but a state which a company should be willing to achieve. The other type of logistics modeling is the descriptive modeling, which reflects on the current state of the logistics system and describes its behavior in a quantitative way. This second type of the modeling was the research goal of this project (see FIGURE 1.1): not to optimize costs of a logistics system of a company or a consortium of companies, but to describe the functioning of the *regional logistics systems*, which may encompass thousands of businesses altogether.

The modeling goals have led to the choice of a macro scope of the model and to the choice for the descriptive nature of the model. The research questions do not necessarily mean that the model should be a macro one, as for instance the GOODTRIP model (Boerkamps, 1999) is based on modeling of the behavior at the micro level through simulation of a population of

actors. However, the input and output of the model have been set to be at the level of interregional goods flows, and given data-related constraints, the choice has been made for a descriptive macro model. Chapter 3 explains in more detail the tradeoffs involved in making this choice.

The descriptive macro nature of the model, fits into the 5-step modeling framework, where the logistics model extends the classical 4-step modeling approach. In transport modeling, the classical 4-step modeling framework can be extended with a 5th-step, the logistics model (Tavasszy, 1998 and 2006, see FIGURE 1.2). The reader may also notice the similarity between freight modeling framework of FIGURE 1.2 and the design of the DBR freight modeling effort of FIGURE 1.5 presented later in this chapter.

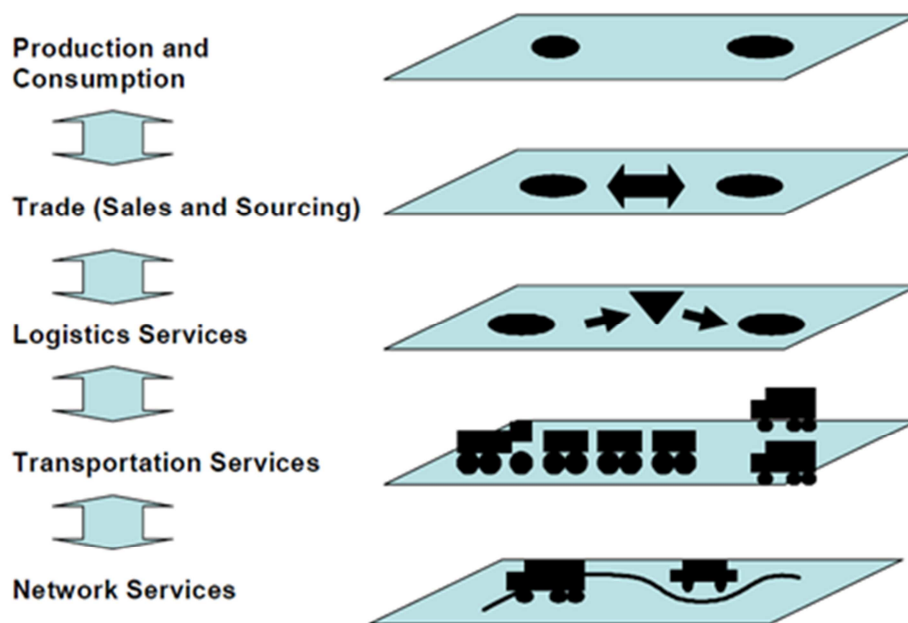


FIGURE 1.2. Freight modeling framework with explicit emphasis on warehouses (Tavasszy, 2006)

This research treats production and consumption locations as well as the volume of goods produced and consumed as a given. Depending on the dataset, a match between production and consumption locations may need to be made in the form of trade flow estimation. The essence of the modelling effort on distribution structures is captured in FIGURE 1.3. Given the trade flow between production and consumption points (or terminals and places of rework, where goods are consumed for further production), the logistics model translates the trade flow into transport flow. The transport flow consists of direct shipments, or shipments via one distribution center or a number of distribution centers.

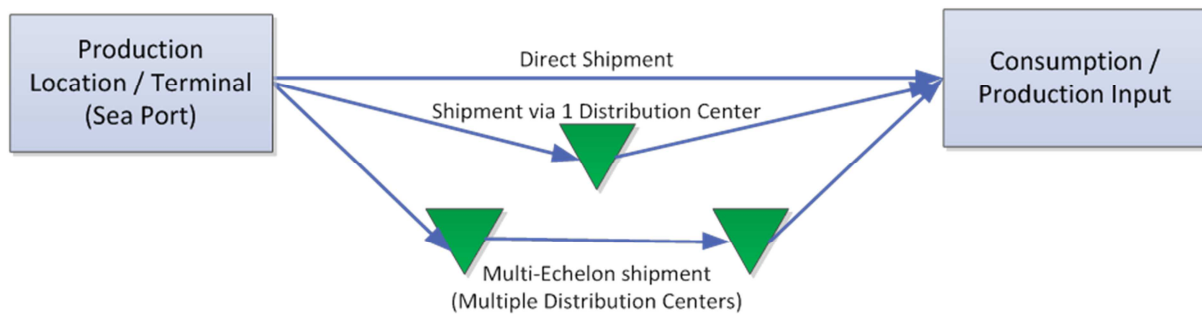


FIGURE 1.3. Conceptual representation of distribution systems

Setting out on the course of designing and implementation of a logistics choice model for interregional trade and transport flows, two major aspects had to be considered together. First, what method can be employed to represent the logistics choices? And second, what data are available to allow application and testing of the methods? The search for methods required understanding on the logistics systems and distribution within supply chains on the one hand and the scientifically proven transport and freight modelling techniques on the other hand.

The essence of a logistics choice model is to take Production-Consumption (PC) trade flows as the input and estimate transport Origin-Destination (OD) flows as the output. This can be realized in a comprehensive modeling suite, such as for instance TRANS-TOOLS (TRANS-TOOLS, Deliverable 6 (2008)) and SMILE (Tavasszy 1998), where the input of logistics model is generated in upstream modules and the output of the logistics model is used in the downstream modules. Incorporation of the logistics model into a comprehensive model, therefore, solves the data input problem, as trade distribution functionality of these models provides the necessary data.

However, incorporation of the logistics choice module into a sophisticated modeling suite would make it difficult to show the empirical validity of the logistics layer, precisely the problem that the SMILE and TRANS-TOOLS models suffer from. A standalone logistics choice model would have to rely on “external” data sources, preferably empirically observed ones: a substantial part of this research project has been devoted to the search and understanding of the available data sources (see Chapter 4); the data availability has influenced the modeling design (see Chapter 3 and 5), leading to the need of implementing not only the logistics choice layer, but also combining it with the trade and sourcing layer of the FIGURE 1.2 framework.

1.3. Research approach

The work presented in this thesis has followed a classical research approach, see FIGURE 1.4. First, the research questions have been identified and confirmed given the practical needs for a logistics choice model, as well as the state-of-the-art in the scientific literature. The research questions led to an identification of the modeling methods, as well as to an understanding of the general data needs for the required logistics choice model. The data needs have been researched together with the question on data availability and sustainability of the available data. The data availability and data quality are the central questions in this research project as they define the scope of the logistics modeling. During the course of the project, some

expected data sources did not prove to be viable, while the other data sources became (unexpectedly) available. The understanding of the data availability has led to a specification of the model design.

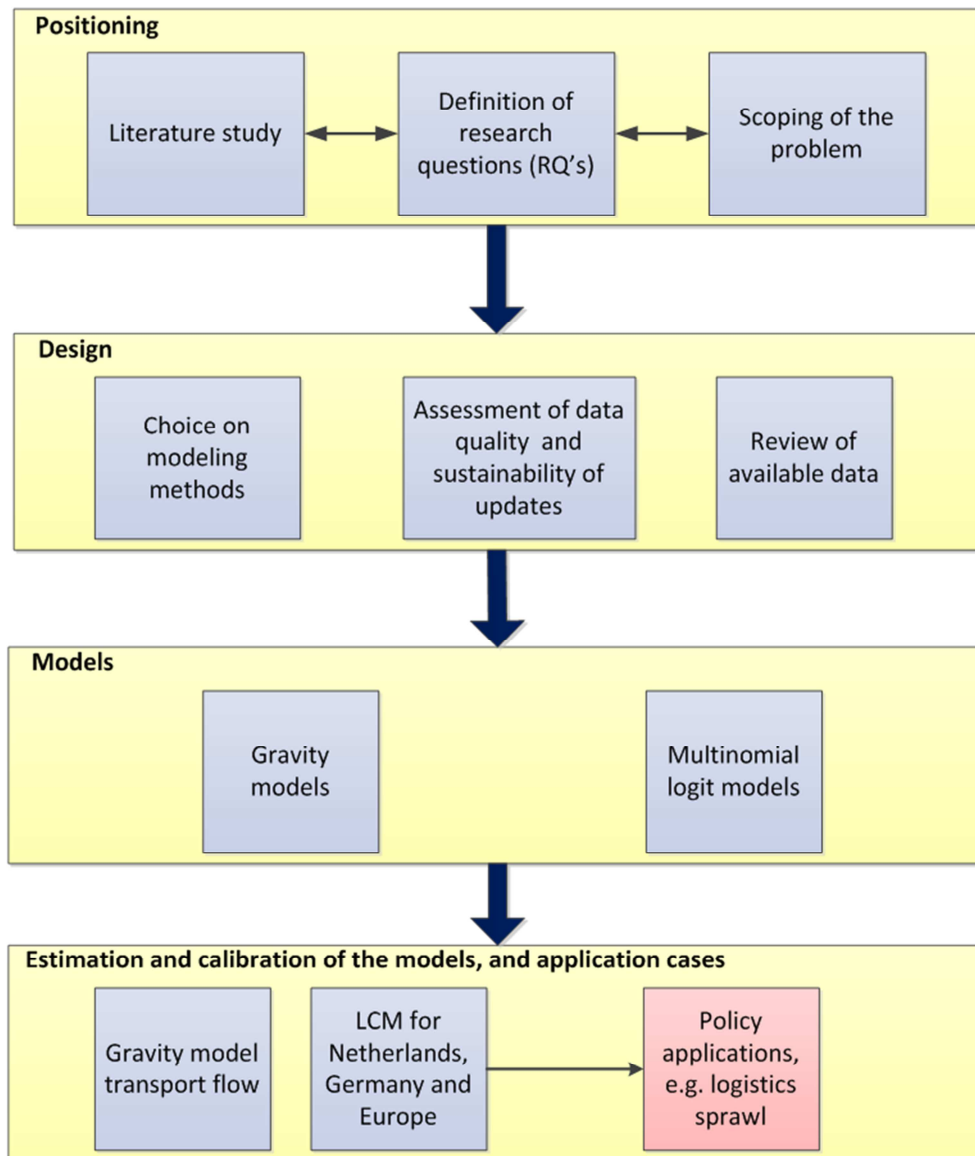


FIGURE 1.4. Schematic presentation of the research approach

Essentially, two main model classes have been developed. The first one is the gravity model, which has been used for two purposes: (1) estimation of transport origin-destination flows, where estimations on distance decay and transport demand price elasticity can be concluded and (2) for the estimation of the trade flows, which are the necessary input for the logistics model. The logistics is modeled using a nested logit model, where top level choice (direct or via distribution) is modelled using a binary logit and the nested level is modelled by a multinomial logit model. The logistics model got two implementations: one is based on the estimation of the components of the total logistics cost, another one is based on the estimation of shipment sizes (transport batch sizes). The cost-based model is deemed to be the primary model implementation in this thesis.

The research project ensures empirical validity of the models by the means of model calibration on the observational dataset for the Netherlands. The German and European data are only partly directly observed, though the model is calibrated on those data as well. The thesis presents a policy-related logistics model application case, where the effectiveness of policy-related scenarios are assessed for the phenomenon of logistics sprawl in the Netherlands.

1.4. A note on Research Background

The negative effects of the strong position of the Randstad region in national and international logistics networks can also be seen as a part of yet broader problem dealt within the Sustainable Accessibility of the Randstad region program (Dutch abbreviation: DBR). The Netherlands Organization for Scientific Research (NWO) has recognized the importance of the sustainability and accessibility of the Randstad region. Therefore NWO has provided funding for the research that would lead to the development of instruments that help understanding, provide insights and decision support tools to achieve the goal of a more sustainable and accessible systems in the region. This PhD thesis is a part of the DBR research program (DBR, 2013), dealing with the sustainable transportation of freight and making the emphasis on the role of distribution and warehousing in the freight systems.

This research effort is a part of an ambitious freight modeling part of the DBR program, see FIGURE 1.5. The four-step freight modeling framework can be recognized in the program design (see FIGURE 1.2 for the comparison). Sub-project 1a has provided for regionalization of interregional trade flows. Research presented in this thesis has been carried out in project 1b, namely, translation of interregional trade flows into transport flows by the means of empirically proven logistics models has been accomplished. Other parts of the research program dealt with optimization of multimodal networks and the problematic of city logistics.

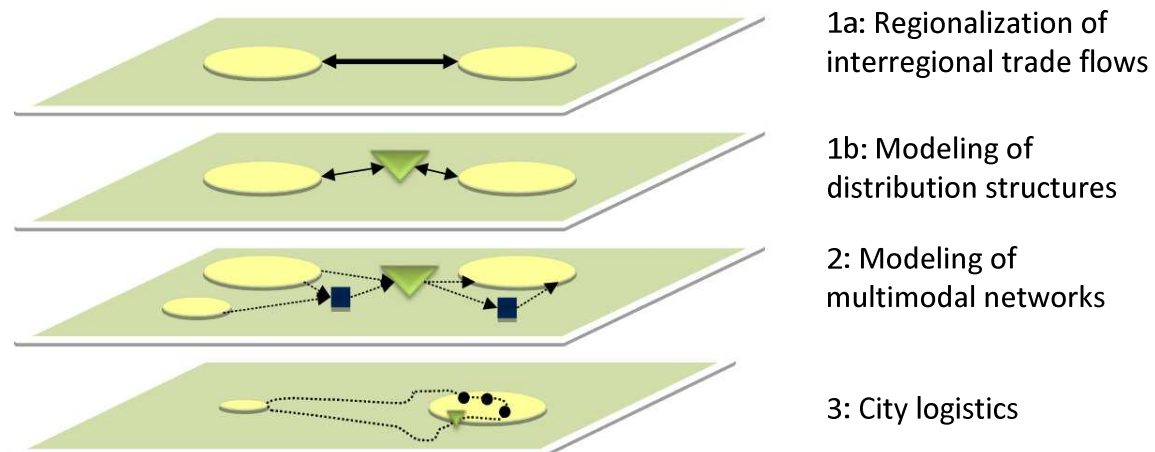


FIGURE 1.5. Structure of DBR freight modeling effort Tavasszy (2014)

Put together, the four parts of the DBR effort on freight modeling, represent a comprehensive freight model, covering all steps from production and trade (see for instance Lankhuizen (2012) on distance decay in trade) through the logistics modeling of this thesis to the question on multimodal networks (see Zhang, 2013) and the questions of city logistics (see Anand

(2012) on problems of city logistics modeling). An industrial implementation of the research results of the project would provide a powerful tool for the policymakers, broadening the policy assessment areas explained in section 1.1.

1.5. Concise guide to this thesis

This book is organized as follows. Chapter 1 has provided introduction to the problem considered in this research project, thus motivating the research effort of this thesis. Chapter 1 looks at the research applicability for the policy-related questions, identifying the parties that might be interested in this research as well as providing an indication on what problems can be tackled by the main research products described in this book.

Chapter 2 provides a detailed insight in the most relevant research contributions and draws the conclusions on the need and novelty of the research presented in this thesis. The chapter outlines the main logistics tradeoffs, especially those related to the distribution, which are relevant at both company and regional levels. It briefly looks into the body of knowledge on modeling at the company level in the domain of company-level logistics optimization. The literature sources on transport modelling, especially the sources at the macro level, i.e. the level of interregional trade and transport flows, are presented and discussed. Furthermore, Chapter 2 acknowledges that if one aims at constructing logistics models at the regional level, company-level optimization can only be used as a part of micro models. The chapter concludes with a positioning of the research of this thesis in respect to the most relevant research undertakings, identifying a novel research field.

Chapter 3 provides a mathematical definition and explanations for the quantitative models developed in the research project. First, main requirements and constraints for the logistics chain model are discussed together with the overall model design. The logistics chain model has to be extended with a gravity model for estimation of the chain model input in the form of trade flows. Additionally, the gravity model has its own interesting standalone applications and can be estimated and calibrated on transport flows. Therefore, Chapter 3 explains how the gravity model is formulated together with the estimation and calibration procedures. Chapter 3 takes a similar approach with respect to the logistics chain model. It provides a mathematical formulation for the model, as well as a description of the model's calibration procedure. Due to the emphasis on empirical validity of the modeling efforts, the calibration procedure plays an important part in this modeling effort, thus linking the modeling efforts directly with the data analysis of Chapter 4 and model calibration results of Chapter 5.

Chapter 4 gives an account of the data used in this research. Three datasets have been made available for the modeling: (1) Dutch road transport flows statistics extended with information on loading and unloading location types; (2) European trade and transport flows (road mode only) for the industrial and agricultural commodities at the NUTS2 spatial resolution level; and (3) trade and transport flows attributed to the food retail sector in Germany at the NUTS2 level. Chapter 4 devotes considerable attention to the analysis of the quality of the Dutch road transport statistics: time series (year-on-year changes) are analyzed and the data quality is assessed by the use of an unrelated dataset, sectorial employment statistics. This later analysis

has not only confirmed usability of the Dutch road transport dataset, but provided useful insights on its own, establishing a quantitative relationship between regional employment in the wholesale sector and the warehousing and distribution throughput. This quantitative link can be used in cases where there is no information on loading and unloading location types (e.g. other West-European countries).

Chapter 5 provides results of the separately calibrated gravity model estimation and the calibration of the logistics chain model. The separately calibrated gravity model has been applied to different flow types (e.g. production-consumption and production-distribution), thus showing that these flows have different properties, such as distance decay and transport price elasticity. The analysis provides an empirically supported argument for more detailed modelling in the logistics chain model. Chapter 5 also explains how the combined gravity-logistics chain model is implemented and estimated, presenting the quality of fit for the estimated transport flows, and reports on the estimated model variables. The logistics chain model estimation results are presented for all three classes of data (Dutch, European and German food retail sector). The materials of this chapter can be seen as a part of the proof of the empirical validity of the modelling efforts of this thesis.

Chapter 6 describes application areas for the logistics chain model and provides concrete examples of applications carried out in the research project. First, a discussion on the role of model variables is provided: it is essential to understand how these variables influence the modeling outcomes and what role the variables play in the choices that the model reproduces. This discussion explains what can be influenced in the model for the purpose of scenario analysis and what useful indicators can be constructed based on the model output, such as ton-kilometer shipped and vehicle-kilometer driven measures. Chapter 6 subsequently considers the case of logistics sprawl for the Randstad region, explaining what can be done to facilitate, or conversely break the trend of spreading of the logistics facilities from centralized clusters to the overall presence. The logistics facility sprawl case has allowed an estimation of the transport and warehousing price elasticities, thus enriching the literature with the estimation of the price elasticities in the logistics and distribution environments.

Chapter 7 reflects on an evaluation of the extent to which the research questions posed in Chapter 1 have been answered in this thesis. The chapter concludes the thesis with conclusions and reflections on the knowledge and experiences outlined in the dissertation. It further presents ideas for the future research efforts in the field of aggregate logistics modeling, providing for continuity in the research area.

2. Positioning of Aggregate Logistics Modeling

The aim of this chapter is to position research presented in this PhD thesis in the freight transport modeling research field. It outlines the most important aspects that play a role in the choices made related to the establishment of the warehouses and distribution facilities, as well as presenting and discussing the most relevant literature contributions. The chapter starts with an analysis of the main functions of the warehouses and distribution facilities, discussing the basic tradeoffs involved in the decisions on the design of supply chains and decisions on the number and location of the distribution facilities. Special consideration is given to the broader field of freight transport modeling, where review papers are discussed, as well as aggregate and disaggregate choice models are discussed. This chapter does not aim at comprehensiveness in its review effort of the broad field of freight modeling; however, the aim is to link the relevant pieces of modelling undertakings if they present a value for the modeling of logistics at the regional level.

Specifically, this chapter looks in detail into the most relevant studies that deal directly with the question of the regional logistics modeling. These studies present their own intrinsic value, pioneering the field of logistics modeling. Their analysis allows a positioning of this thesis's research in respect to knowledge foundation and identification of the knowledge gaps. Finally, based on the identified knowledge gaps, this chapter outlines the requirements for the new logistics model, which is discussed in this thesis.

2.1. Function of distribution, tradeoffs and optimization

Distribution and warehousing facilities play an important part in contemporary supply chains. The distribution and logistics process determine how products are retrieved and transported from the warehouses to the retailers (Beamon 1998). This process includes the management of inventory, transportation, and final product delivery. These processes interact with one another to produce an integrated supply chain. The design and management of these processes determine the extent to which the supply chain works as a unit to meet the required performance objectives.

Warehouses and distribution facilities provide for the functions of temporary storage of physical goods, order processing activities and allow splitting large inbound batches to the smaller batches (or shipment sizes) delivered to the retail facilities and individual clients. Despite the growing interest in the development of intermodal logistics centers by scholars in academia and public and private sector actors, a consensus on the definitions of distribution centers does not yet exist (Higgins et al 2012). A distribution center can also be called a warehouse, a DC, a fulfillment center, a cross-dock facility, a bulk break center, and a package handling center.

An efficient supply chain would involve Full Truck Load (FTL) shipments to a distribution facility, and shipments in smaller sizes (Less than Truck Load, LTL) such as individual pallets, roll cages or even parcels from the distribution facility to the customers. Therefore, distribution facilities provide a possibility to keep inventory (physical goods) close to the place where these goods are needed, and provide a way to minimize transport costs, as large shipment costs per ton-kilometer transported are generally smaller than small shipment transport cost per ton-kilometer transported. The degree of centralization will be influenced by future changes in the relative costs of logistics inputs and the evolution of management practices in this field (McKinnon, 2009). Trends in increasing international trade and transport, requirements for high quality of logistics, as well as enabling technologies such as IT technology led companies to continuously optimize their distribution networks (Ruijgrok and Tavasszy, 2007).

2.1.1. Trade-offs involved in distribution-related choices

The tradeoffs related to balancing business needs on the one hand, and transport, inventory and facility costs on the other hand are very well studied and understood by the industry. The function of inventory is essentially twofold: first it is used to satisfy customer demand directly from the storage, thus realizing the concept of Customer Order Decoupling Point, CODP (e.g. Rudberg and Wikner, 2004). The CODP separates decisions made under uncertainty from decisions made under certainty concerning customer demand in the production- and distribution- related activities. The stocks essentially ensure that the customer demand can be realized immediately from the inventory without the need to backorder. In many business environments, demand need to be satisfied from the stocks immediately (e.g. supermarket shelves: if a product is absent, the client will turn to another shop; online sales: it is becoming a custom for next day deliveries, if not in stock, the product would probably be ordered from a competing online store). It should be noted that the stocks are not free to keep, see section 2.1.2 for more information about stock costs. FIGURE 2.1 illustrates the first tradeoff between stock keeping costs and the cost of lost sales. Note that the issue of lost sales and demand accommodation from the stocks has an impact on spatial stocks allocation: if it is expected in the industry that the product is available on demand, the stocks will be allocated in warehouses close to the customers, even though it might not be justified by other cost tradeoffs.

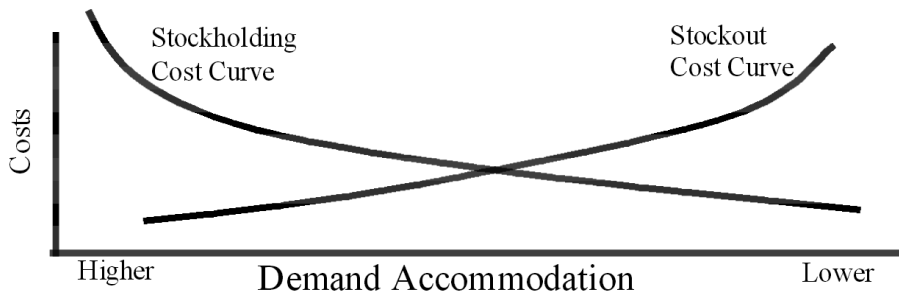


FIGURE 2.1. Inventory holding and stock out costs trade off, adopted from Tailor (1998)

The second function of inventory is in the reduction of transport costs. Suppose a demand of 10 items per day and transport costs of \$100 per delivery. If delivered daily, it would cost \$700 per week; if delivered once per week under assumption that 70 items fit into the vehicle, the cost would be \$100 per week (the real price can be higher as the service provider may charge an extra for a more voluminous delivery). However, the reduction in transport cost is compensated by the need to keep more products on hold. FIGURE2.2 provides a quantitative example of the tradeoff between transport costs and inventory holding costs under the assumption that per item transport costs for larger shipment sizes are smaller than the per item transport costs for the smaller shipment sizes.

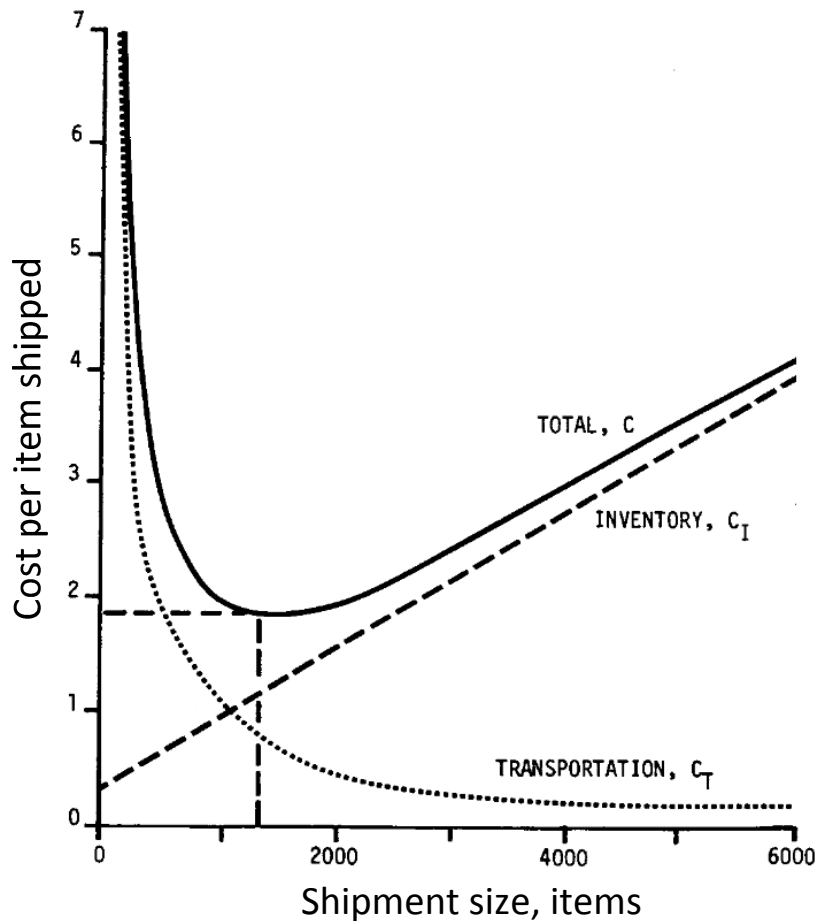


FIGURE 2.2. Total cost per item as a function of shipment size, adopted from Blumenfeld (1985)

The third basic tradeoff is related to the number of warehouses and the cost of inventory. In this context the inventory costs include not only the costs associated with the capital frozen in the stored goods, but also the cost of storage and handling, i.e. the costs associated with the functioning of the warehouse or distribution center. FIGURE 2.3 presents an example of a centralized stock location in the Netherlands (right) and a decentralized distribution structure, characterized by a number of regional distribution facilities (left). Clearly, the inventory costs will be larger in case of multiple stock locations, but the costs of shipping, i.e. the transport costs from the distribution facility to the customers, will be smaller under condition that the shipment sizes are smaller than a Full Truck Load (FTL).

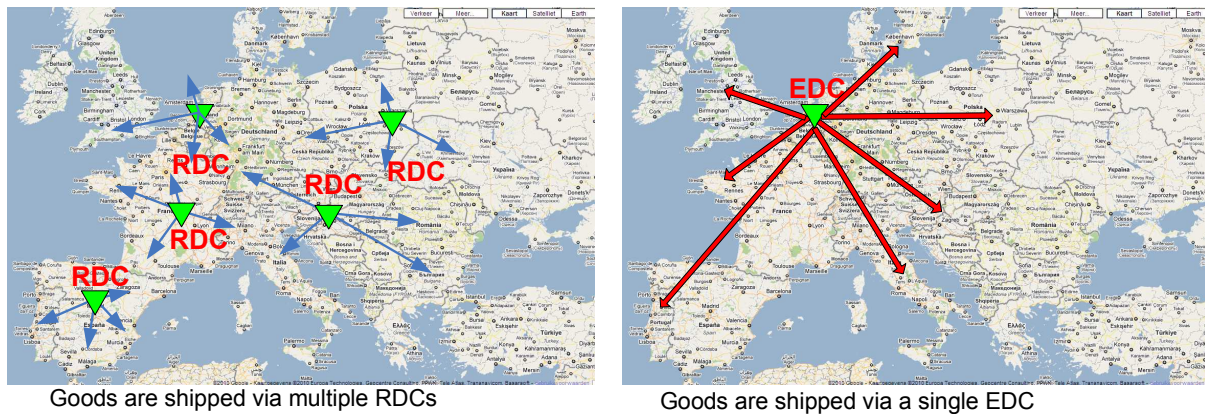


FIGURE 2.3. De-central and central inventory locations

The cost tradeoffs related to the centrality of inventory locations are very well known. For instance, Feldman et al. (1966) present heuristics on minimizing the total logistics cost composed of warehouse (inventory) and shipping costs, see FIGURE 2.4 for an example. The original drawing from 1966 has been illustrated to show that this tradeoff has been already studied some 50 years ago. The author of this thesis has also performed an industrial supply chain optimization for a large multinational company (see Davydenko, 2006), where the best tradeoffs, among other between the inventory costs and the transport costs, had to be found. A further accessible reading on the basic tradeoffs involved in the choice of location and number of warehouses can be found in US DoT (2013).

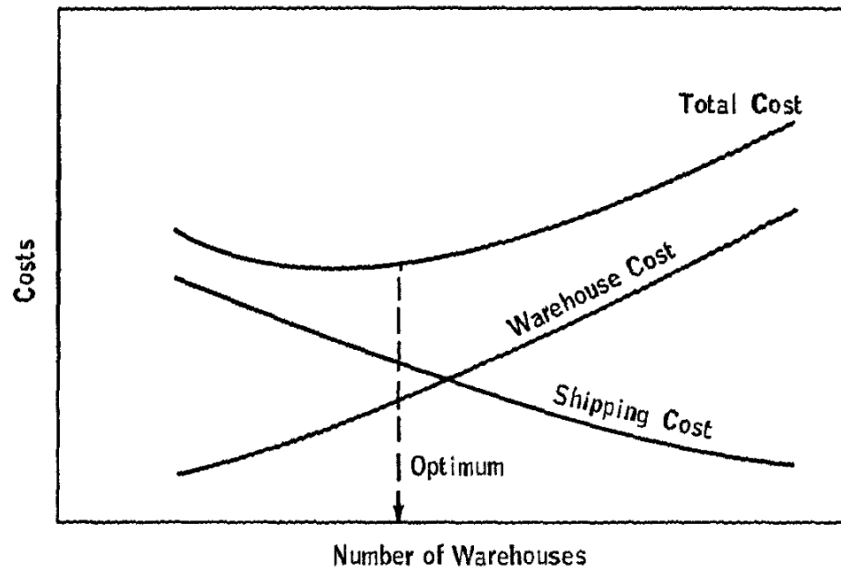


FIGURE 2.4. Tradeoff between shipping costs and warehouse (inventory) costs, Feldman et al. (1966)

Therefore, it is possible to conclude that, apart from niche applications for inventory (e.g. value appreciation for commodities like gold), the inventory has three main functions: satisfaction of customer expectations on immediate availability and hence reduction of a lost sales chance and costs; reduction of transport costs by allowing larger shipment sizes to be transported; reduction of transport costs by shortening of the distance between the customer and the product, and thus reducing the costs of expensive customer deliveries. The following section looks into more advanced company-level logistics optimization problems.

2.1.2. Logistics optimization at the company level

The essence of the company-level logistics optimization is aimed at the satisfaction of two diverging goals. The first goal is related to the nature of the business: provision of the clients with the desired service level. The second goal is to minimize expenses and costs associated with the logistics operations. Thus, any business that involves trading in physical goods balances the degree of customer requirements (or customer expectations) satisfaction with the costs of underlying operations.

The realization of this balancing act depends on the nature of the business. In almost all cases it comes down to availability of the physical goods at the right moment and at the right place, or in the definition of business logistics "having the right item in the right quantity at the right time at the right place for the right price in the right condition to the right customer" (Malik 2010).

In the real supply chains the place of production is often spatially separated from the place of consumption. The burden of moving the goods from production to consumption location is often carried out by the seller, which is the producer of goods, or a merchant wholesaler. The receiving party (consumer) expects availability of goods at the points of consumption or at the points of sale. This expectation is especially strong for competitive markets, where products

can be substituted by those of the competitors. Therefore, the producing party aims for a certain customer service level, namely availability rate of the products, which is critical to all members of the supply chain (Coyle et al. 2008). The availability rate is often measured in percentage of the time that the product is physically available.

In order to realize the desired service level, a company must ensure availability of the products by keeping them in stock at some points in the supply chain. However, keeping products in stock is costly. The cost of inventory is more than just its purchase price, it includes a variety indirect costs, such as interest cost, physical storage cost, the cost of obsolescence, and many other cost components. From the literature on modal split, it is known that time valuation of commodities is much higher than it could be attributed to the simple inventory holding cost (see Rich et al 2009). Therefore, the companies determine the desired availability rate as a balancing act between inventory holding costs and the costs associated with lost sales due to stock-out. These questions are dealt with in the Inventory Management body of knowledge, see for instance Pride et al. (2008).

Once a desired service level in the form of product availability rate is determined, a company can start with the optimization of its supply chain. In the most generic case, a company can start from scratch, only taking into account known (or expected) locations of customer demand. Subsequently the company identifies potential production locations (if applicable: an importer or wholesaler does not need production locations; a company with existing production facilities may also consider them as exogenous / given). The supply chain optimization problem is, therefore, reduced to the decision on which locations to use for production and distribution, such that the customer demand is satisfied according to the expectations, while the total costs are at the minimum, see FIGURE 2.5.

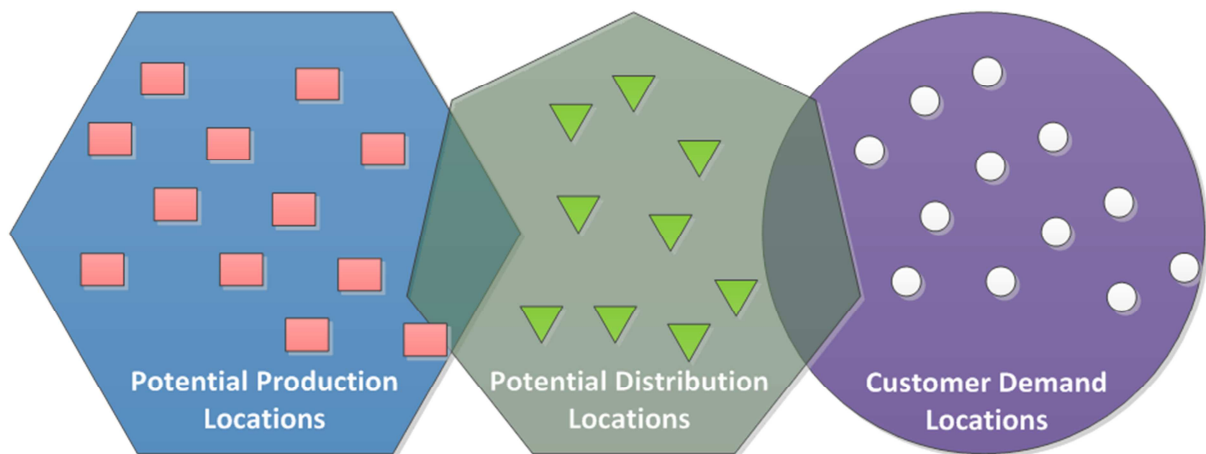


FIGURE 2.5. Conceptualization of scope of the distribution channel design problem

The supply chain optimization problem in general and the more relevant part of the problem on location of distribution facilities in particular, are very well studied in the literature. This is essentially a cost minimization problem with constraints related to the service level and solution feasibility. The literature extensively studies all possible facets of the problem. For instance, a widely cited review of the literature on facility location and supply chain

management (Melo et al., 2009), contains 139 references to the peer-reviewed works on this problem. As indicated in Chapter 1, the research effort described in this book is not aimed at supply chain optimization at the micro (company) level; thus the main principles of facility location problem are only touched upon here for the sake of completeness of the argument, as well as for those readers who are interested in micro normative approaches, where determination of the optimal facility locations may be one of the core elements of the research.

A general facility location problem involves an exogenous set of spatially distributed customers and a set of potential locations to serve customer demand (Drezner and Hamacher, 2004). The solution to the facility location problem answers the questions on which potential locations for facilities should be used (opened); which customers should be serviced from which facility (or facilities) so as to minimize the total costs (customer assignment problem). FIGURE 2.6 shows “a solution” to the facility location problem of FIGURE 2.5.

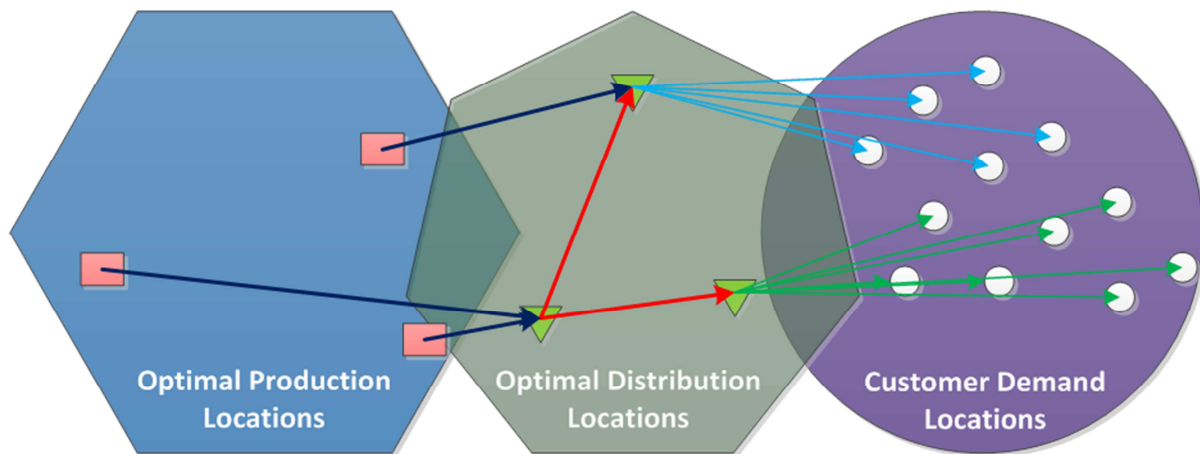


FIGURE 2.6. A solution for the conceptual distribution channel design problem of FIGURE 2.5.

A practical formulation of the facility location problem is presented by Melachrinoudis and Min (2005). The authors formulate a mixed-integer programming model to solve the warehouse redesign problem, proving its usefulness on a real-world application case in the United States, showing sensitivity of the solution. The model takes into account such model parameters as facility relocation cost, capacity of the warehouses and manufacturing facilities. The problem formulation is a mixed-integer programming, where some of the decision variables are either 0 or 1 (for potential facility locations there are only two outcomes: either a warehouse is opened, or there is no warehouse at the location). This type of mixed integer programming belongs to the class of NP-complete problems, which are the problems that cannot be solved in polynomial time: the computation effort growth exponentially with the number of possible locations to assess.

For many real world applications NP-completeness of the optimization is not a real problem due to the problem size: they can be solved to optimality because of the limited number of potential facility locations. However, in other cases, the solution space is large and heuristics must be applied. For instance, Bard and Nananukul (2009) look at the optimization problem

of a manufacturing company with a set of diverse clients. The optimization tradeoff is between delivery frequency and stock levels, under the condition of no back orders. The stocks may be located at the customers in the Vendor Managed Inventory (VMI) setting. The authors acknowledge computational difficulty of solving the problem and consider a number of heuristics that solve it with a good result. Another example of a heuristics application is the TNO RESPONSE (2009) model, which uses a genetic algorithm to solve computationally complex problems.

It is interesting to look at the review papers on facility location problem as the part of a broader supply chain optimization context. Maixell and Gargeya (2005) argue that the supply chain optimization research should address the cases when production is only partly controlled by the firm-decision maker, as some of the products can be produced within the firm, while other products may be sourced from the external suppliers. With respect to the richness of the Operations Research (OR) research done about the facility location problem, Melo et al (2009) provide a long list of references (139 references), splitting them by the class of solver used to find optimal (or good in case of heuristics) locations for the facilities.

This section looked at the function of the distribution and warehousing, as well as at the optimization methods for the location choice for these facilities. These questions have been considered from the point of view of individual business: a company, or a number of collaborating companies in a supply chain. The primary concern of this (micro) level is the satisfaction of company's customers at the minimum cost. However, if one aims at a construction of a model capable of determining of the warehouse and distribution locations in the interregional flows (i.e. the flows generated by countless businesses operating at the level of regional economy), the approach for company-level facility location would not suffice, as the systems at the regional level are not governed by a single decision maker but by many businesses active there, representing different strategies and optimization goals. The questions on the modeling of the aggregate flows are considered in the next section.

2.2. Review of relevant transport modeling literature

This section provides a broad review of the modeling efforts that are relevant in the context of incorporation of logistics choices into the freight modeling. The section distinguishes modeling efforts related to modeling of aggregate and disaggregate choices. Aggregate choices are related to the modeling that does not account for individual decision makers involved in generation of the flows; disaggregate choices look into the behavior of individual companies, who, we assume, attempt to optimize their operations. Some attention is given to the literature that does not fall into either category, such as review papers and broad approaches.

2.2.1. Review literature

De Jong et al (2004) and (2012) provide a review of the European literature on freight transport models that operate at the national or international level. The authors point out that the introduction of logistics decisions into the models has been a recurring theme in the

development of freight transport modeling since 1998. The logistics models are split into disaggregate choice and aggregate choice models, with a broad definition of a logistics model. For example, the World Container Model (Tavasszy et al 2011) is considered to be an aggregate choice logistics model (the model computes port choice in international maritime container transport). De Jong et al (2012) outline three broad areas for further research efforts on logistics in freight modeling: (1) incorporation of production networks' configuration to represent changes in logistic demands of products in the supply chain; (2) study on change in handling factors, length of transport haul and shipment size through the modeling of spatial distribution structures including location and use of warehouses; (3) acquisition, scheduling, routing and repositioning of vehicles, to represent changes in vehicle stocks, use of light duty vehicles (LDVs) and vehicle occupancy rates. The authors argue that various logistics decisions need to be considered together; and the core activity of firms—production—needs to be modelled. Hence the 5-step modeling framework (see FIGURE 1.2) can be reconsidered by modeling, for instance, the first three layers on production, trade and logistics in one model.

Homblad (2004) notices that the production-consumption flow (PC flow) is not equal to the transport flow (OD flow). He introduces a notion of “two worlds”: PC world that represents direct shipments from production to consumption and DC world, which represents flows via distribution structures. He further proposes a linear formulation for the routing problem: large-size shipments profit from the consolidation provided by the distribution structures, and small-size shipments have an advantage of direct shipments.

Chow et al (2010) review freight modeling practices looking for a suitable approach for the State of California. The authors review existing freight modelling approaches, emphasizing logistics, and corresponding data needs of the models. The authors conclude that logistics modes were applied in the Netherlands, Germany and Japan, but not in the US, because the private firm supply chain costs and operating behavior needed by these models are unavailable. The authors state that logistics models would require a costly shippers survey to obtain company level micro data, though without a provision of a reasoning for the statement. The paper provides an overview of freight models, especially taken from the US data availability point of view.

Yang et al (2010) reports on freight forecasting models, classifying them into 7 categories, whereas supply chain / logistics models get a dedicated model class. The authors draw the conclusion that freight modeling is not always limited to a single project at one point in time or to a single model class, thus the logistics models tend to overlap with other model classes depending on the development needs behind the construction of these models.

Tavasszy et al (2012) provide a review of the state-of-the-art on the Incorporation of logistics in freight transport demand models. The authors outline the most promising approaches to the modeling of logistics structures in the freight demand transport models as the nodes in chains of transport activity legs that connect production and consumption locations. The inventory choice, supernetwork choice, and hypernetwork modeling approaches are considered. The paper proposes 3 main avenues for research on logistics in freight transport modeling: (1) supply chain type choice (function, number and location of inventories) and vehicle type

choice (in particular, light vehicles vs. heavy duty trucks); (2) models linking supply and demand at different levels (within the classical multi-stage framework or in a hypernetwork model); (3) extension of the spatial and dynamic reach of models to allow studying the evolution of global logistics networks and their interaction with the systems at national and regional levels.

Tavasszy and De Jong (2013) recently published a textbook on various aspects of transport modeling, from the SCGE modeling down to the vehicle trips and urban freight models. Special attention is given to the modeling of distribution structures (Friedrich et al 2013) and inventories (Combes 2013). It is argued that transport costs, storage costs, ordering and handling costs and the costs of risk are the primary cost drivers influencing the design of the logistics organization. Friedrich et al 2013 distinguish micro and macro level logistics modeling, with the intrinsic challenges at each level.

Huber et al (2014) argue that while logistics hubs play a decisive role in transport process, there is still little knowledge about those hubs and they are insufficiently considered in most freight demand transport models. The authors state that in respect to logistics hubs, the modeling approaches and data availability vary significantly in Europe, with little chance for a transfer of models between the countries.

2.2.2. Aggregate models

The aggregate choice models are a class of models where the choices are made not at the level of individual decision makers (micro level), but at the level of populations of decision makers (macro level). These choice models operate at the level of aggregate agents or average costs. Specifically in the realm of freight transport models, practically every aggregate international, national or regional freight transport model in the world is lacking a modelled logistics system. Exceptions are the SMILE and SMILE+ models for the Netherlands (Tavasszy et al., 1998; Bovenkerk, 2005), the SLAM model for Europe (SCENES Consortium, 2000) and the EUNET 2.0 model for the Pennine Region in the UK (Yin et al., 2005). The SMILE+ and EUNET 2.0 models are considered in greater detail later in Chapter 2, as these models are closely linked to the logistics chain model described in this thesis.

The TRANS-TOOLS model deserves a special note, as this model is the main policy tool of the European Commission. The TRANS-TOOLS did not develop its own logistics module; the implementation of the logistics module is based on SLAM (Tavasszy et al., 2001). This module makes it possible to evaluate the impacts of changes in the logistics and transport systems within Europe on the spatial patterns of freight transport flows, through changes in the number and location of warehouses for the distribution of goods. The logistics model is embedded in TRANS-TOOLS and does not function as a standalone model: it is not validated and cannot be used outside of the TRANS-TOOLS environment.

Pattanamekar et al. (2009) introduces the notion of Production-Consumption flows into the aggregate mode choice problem. Distinguishing between transport flows and production-consumption flows allowed the authors to look beyond the classical modal split problem and incorporate multimodal transport into the modal split model. This effectively created a

transport chain model in the context of modal split. The authors present the quality of the estimated fit, which is based on a survey data.

Kim et al. (2010) proposed a logistics model called ‘physical distribution channel choice’ to estimate shippers’ choice of the logistics chains linking them to the consumption points. The authors distinguish between direct and indirect distribution channels, thus modeling the choice between direct shipments from production to consumption and shipments via distribution facilities. There are two important aspects to this study: first is that the model output has been validated empirically on the Korean distribution channel survey; second is that the model does not determine spatial locations of the distribution (i.e. there is no spatial component in the model and the authors only determine the type of distribution channel). The main peculiarity of the study is that the authors acknowledge that the shippers choose the distribution channels which minimize the overall logistics costs including inventory cost, transportation cost, etc. However, the modeling is done using variables not related to the logistics costs, such as market characteristics, product characteristics, company size, etc. The model proposed by Kim et al. (2010) is in many aspects related to the logistics model presented in this book, such as distinguishing between direct and indirect shipments and within the class of indirect shipments determination of the distribution channel (chain) using the multinomial logit. It is also similar in its empirical validation, however the use of variables unrelated to logistics costs makes it different. This paper will be further discussed later in Chapter 2 in more detail.

Based on a survey of German logistics facilities, the Huber et al (2014) developed a number of regression models, taking an approach similar to Davydenko et al (2011), and estimate transport volumes as a function of known variables such as size of the facility area and size of transshipment area. The authors subsequently propose a utility based gravity model to estimate the strength of customer-hub relations. Huber et al (2014) provide directions for the freight generation estimation by the German logistics hubs; however the model lacks the logistics chain decision functionality. For instance, only the hub-consumption leg is considered, while the choice on whether the hubs will be used at all (direct shipment) and how the upstream transport leg is organized, remain out of scope.

2.2.3. Disaggregate models

The disaggregate choice models are a class of models where the choices are modelled at the level of individual decision makers, for instance, routing of individual trucks, or in the context of logistics modeling, the choice of a chain for a particular supply-demand relation. This level allows applications of optimization techniques, using deterministic choices (i.e. the best choice gets all the volumes) or application of discrete choice with random utility, where the best choice gets a fraction of flow under the assumption that the utility or disutility (costs) of the choice are not fully known (e.g. logit). The drawback of the disaggregate approach is that, in principle, the full set of decision makers has to be represented in the model, which is mostly impossible in a direct way. Therefore this class models relies on a generation of a population of the agents (decision makers). The quality of the modeling depends to a large extent on the quality of the generated population that represents the agents.

An interesting school of thought has been developed at the Karlsruhe University. It is based on the observation that business relationships in the transport sector lead to an establishment of collaborative relationships between companies, especially those of middle-sized forwarding companies (Liedtke 2004). These “meso” structures are an intermediate step between micro-level of the individual company world, and the macro world of the economy. These relationships have to be taken into account when modeling behavior of parties (actors) that control supply chains: not only individual companies, but the collaborative meso structures are proposed to be taken into account.

These ideas have further been elaborated in (Liedtke 2006), proposing a conceptual framework for the organization of freight data patterns on a microscopic basis. The demand side of the transport market is expressed by microscopic commodity flows, while the supply side is explicitly distinguished by homogenous transport market segments. An important conclusion of this contribution is that the inter-sectorial commodity flows and the overall quantity in the transport markets fit together, which allows the construction of a microscopic freight model disaggregated according to sectorial commodity exchange processes and transport markets. This idea paved the way for construction of the SYNTRADE model (Friedrich 2010) for modeling of the food retail sector in Germany and the INTERLOG model prototypes (Liedtke 2009a). The later provides an overview of a bottom-up approach to commodity transport modeling in which logistics structures are not deduced or broken down from aggregate statistics, but are created by simulation. A number of logistics-related assumptions are made, such as locations decisions are strategic, while transport decisions are operational or tactic; the driving force behind transport is the shippers; shippers conclude service contracts with forwarders on a local market. To capture this logic, the INTERLOG uses a 3-step approach. The first step generates location patterns of heterogeneous companies in space based on available statistics (generation module): this is done using information over statistical distribution of company size and location and application of the Monte Carlo algorithm. In the second step supplier–recipient relationships are built up where “companies” determine the demand for commodities supplied. This procedure is done based on Input / Output tables as well as attraction between production and consumption. Finally in the third step, in an interactive market simulation, flows of goods are assigned into shipment cases, transportation contracts are allocated to forwarders and truck tours are constructed.

Interestingly, the steps 1 and 2 of the INTERLOG model lay the ground for the logistics simulation of the step 3. It allows using different modeling techniques (e.g. agent based modeling) to let the model decide upon shipment size, contract awards (shippers award contracts to forwarders) and some degree of cargo bundling. To our knowledge, the INTERLOG model has not been calibrated on real flow data; however, the share of empty runs can be used as an empirical indicator of the model validity. The ideas of the INTERLOG model have been further developed and validated in the SYNTRADE model. The SYNTRADE model presented in the PhD dissertation (Friedrich 2010) achieves empirical validity of the simulation of transport and stock allocations for the German retail sector (see section 2.3.4, where the model is analyzed in more detail, as it has been an important inspiration source for the logistics choice modeling and the source of data for one of the model applications).

Nonetheless, already in 2009 some limits of micro level simulation became apparent. Liedtke 2009b acknowledges that simulation models perform impressively, however most models still have considerable difficulty in incorporating simultaneously the existence of transport network, logistics and collaboration structures. Thus, the construction of comprehensive, stepwise disaggregate choice models is problematic. In other words, the authors acknowledge that there are practical boundaries to the application of micro simulation models. The practice shows that the application boundaries of micro simulation models do not cross the borders of homogenous economy sectors such as the food retail sector.

The example of micro-level modeling of actors' decisions is presented in PhD dissertation of Maurer (2008). Commercial supply chain optimization software CAST has been used to determine logistics chains for the Drinks commodity in Great Britain. The precise functionality of the CAST software is not known to the author, however, it incorporates the main elements of the company-level logistics optimization described in section 2.1.2 (see Groenewout 2005). The work of Maurer 2008 is analyzed in more detail later in Chapter 2. However in short, the main limitations of this work are related to the broader empirical validity of the modeling outcome and a modeling application to a very narrow commodity flow of Drinks.

The GOODTRIP model (Boerkamps 1999) is a simulation model that calculates the structures of supply chains. Based on consumer demand, the GOODTRIP model calculates the volume per commodity type in volume terms for every zone. The commodity flow matrix is based on goods attraction by consumers on the one hand and the producers or at the city borders on the other hand. Next, the goods flows of each goods type are combined by using groupage probabilities. Every combination of goods types is regarded as a different flow. The combined goods flows are assigned to vehicle tours. The conversion is done per origin-destination pair. The tours per mode are assigned to their infrastructure networks, resulting in mode-specific network loads. The network loads are then used to determine vehicle mileage per mode. The modelling process is sequential; there are no feedbacks to previous phases in the process. The paper, though, does not provide mathematical definitions on how the assignment is done: be it at the level of establishments or a more aggregate approach at the level of interregional flows. Also, there is no empirical validation.

The work of Wang and Holguín-Veras (2009) and Holguín-Veras (2008) has considered enhanced formulations to model commercial tours and vehicle empty trips at the aggregate level. The aggregate tour model starts from a micro level formulation, enumerating promising tours and calculates the most likely aggregate flow patterns using a maximum entropy formulation. Wang and Holguin-Veras (2009) describe two variants of entropy maximization formulations that are aimed to estimate the tour flows of commercial vehicles given the number of trips produced by or attracted to each node, and the impedance to travel. The paper shows that the entropy maximization formulations are an efficient way to accommodate the commercial vehicle tours into the aggregate-level urban freight demand modeling, thus using micro-level OD data it is possible to estimate macro flows. For the Denver area, the mean absolute percentage error of the estimations is around 7%. Wang and Holguin-Veras (2008) also show applicability of micro simulation for construction of goods-related vehicle tours that satisfy known transport OD flows.

De Jong and Ben-Akiva (2007) propose a logistics model for Norway and Sweden at the level of individual sender-receiver relationships, simulating shipment size and transport chain used for all relationships using a deterministic (EOQ based) cost function. A random utility function is used to estimate probability that a particular choice within a relation will be made. The authors acknowledge that three classes of data are necessary: infrastructure data, especially locations of logistics facilities, logistics and transport costs data and data on individual shipments. Thus, implementation of the model depends on availability of a comprehensive shipper and transport survey. The modeling is also complicated by the need to generate adequate populations of shippers and receivers, as no survey can provide a complete set of these companies and consumption points. In addition, only in-transit (pipeline) inventories are modelled and the function of intermediate inventories for formation of transport chains is disregarded. Thus, DC's serve only cross-docking function in this model.

Combes (2010) argues that freight transport is discrete in its nature: goods are bundled together into shipments, which are transported from origin to the destination. There is no notion of shipment in the classical 4-step freight modeling framework. Using the French shipper survey ECHO, Combes (2010) tested relationship between shipment size and order frequency using the classical Economic Order Quantity (EOQ) approach developed by Harris (1913) on the ECHO micro data on shipments. The EOQ formulation balances transport cost and inventory costs: a more expensive transport for smaller shipment sizes is compensated by smaller stock keeping costs. The estimation of the EOQ model on the ECHO data gives satisfying results confirming a general validity of the model on a large population of firms. Combes (2010) shows that EOQ provides a useful cost function to explain mode choice.

Samimi et al. (2009) point out that the gap in the behavioral freight data is a fundamental barrier in disaggregate freight modeling. Supply chain specifications, shipments characteristics, and transport network properties need to be considered in order to provide a realistic estimate of shipping behaviors. The authors argue that an activity-based framework can take into account the complexities of contemporary sophisticated supply chains. Indeed, at the level of micro simulation, the peculiarities of the individual supply chains can be taken into account: the micro level allows introduction of any levels of complexity and sophistication in the models. They develop a simulation framework that synthetically generates micro data.

There are also attempts to bridge the gap between micro- and macro- worlds. While focusing on micro simulation, Sahraoui and Jayakrishnan (2005) discuss macroscopic issues to the extent of a possibility to combine models into the hybrid framework. The approach is deemed to be applicable in advanced traffic management and information systems.

2.3. Detailed review of closely related approaches

This section reviews the five most relevant research and modelling undertakings that inspired and influenced the research objectives and the design of the Logistics Chain Model, which is the main product of this PhD thesis. All five models considered in this section treat logistics choices explicitly. Two of the considered models are normative models, the remaining three

are descriptive ones. The models have different levels of complexity and empirical validation. Although calibration is the weakest part in all these efforts, a detailed look is necessary to understand which specific techniques were used. This section draws conclusions on each model; section 2.4 provides a concise analysis of the research field and positions research results of the thesis in this broad research field.

2.3.1. SMILE / SMILE+ model

The SMILE model (Tavasszy, 1998) and its industrial successor, SMILE+ (SMILE+ Handboek, 2004), used by the Dutch Ministry of Transport, Public Works and Water Management, was the first explicit logistics module / model extension of the classical 4-step modeling framework. The SMILE model has been the most important inspiration source for the research activities described in this thesis. The SMILE model follows the classical freight modelling framework, presented in Chapter 1 (see FIGURE 1.2), with an addition of the explicit logistics module and an explicit treatment of the economy and transport volumes for the assessment of scenarios related to a specific future year, see FIGURE 2.7.

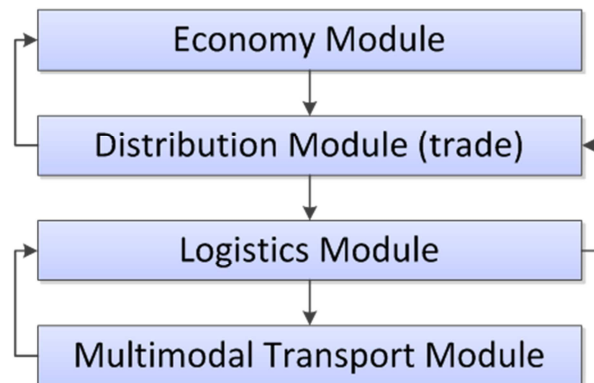


FIGURE 2.7. General Structure of the SMILE+ model

The SMILE+ economy module determines the locations and volumes of production and consumption; the distribution module estimates trade links that match production and consumption; the logistics module determines how trade flows are realized in practice, namely it determines transport chains that link the trading regions; the transport module determines how transport flows, estimated in the logistics module, are realized on the infrastructure networks, including modal split. The transport module also uses and provides network resistances, coupled with the projected growth factors through a recursive dynamic setup. The SMILE+ model ensures propagation of the expected growth and costs in the reverse direction (upstream) towards the economy module, i.e. the transport and logistics costs influence economic developments and distribution of trade.

The logistics module (FIGURE 2.8) presents the biggest interest in the context of the study. It translates trade flows into transport OD flows, determining alternative logistics chains (FIGURE 2.10).

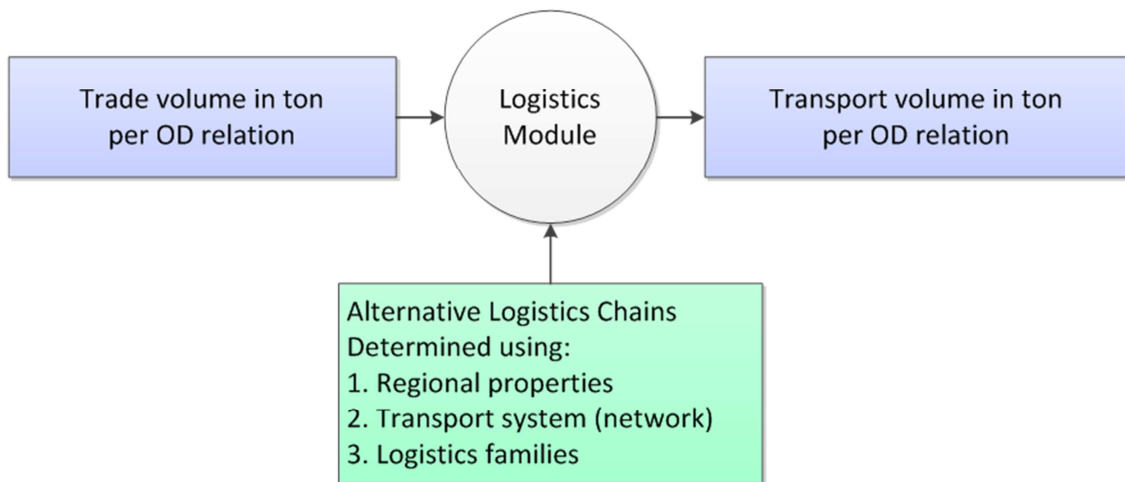


FIGURE 2.8. Conceptual representation of the SMILE+ logistics module

The logistics module of SMILE explicitly distinguishes 5 types of logistics chains (FIGURE 2.10): (1) direct shipments, meaning that the goods are physically loaded at the production facility and unloaded at the consumption location; (2) a chain that involves intermediate storage of goods or transshipment at the Regional Distribution Center (RDC): the goods are loaded at production, then unloaded at RDC, and subsequently loaded at the RDC and unloaded at the consumption location; (3) similar to (2), but the facility is the European Distribution Center; and (4) / (5) involve a multi-echelon distribution structure, where physical transport takes place between the distribution facilities, see FIGURE 2.9.

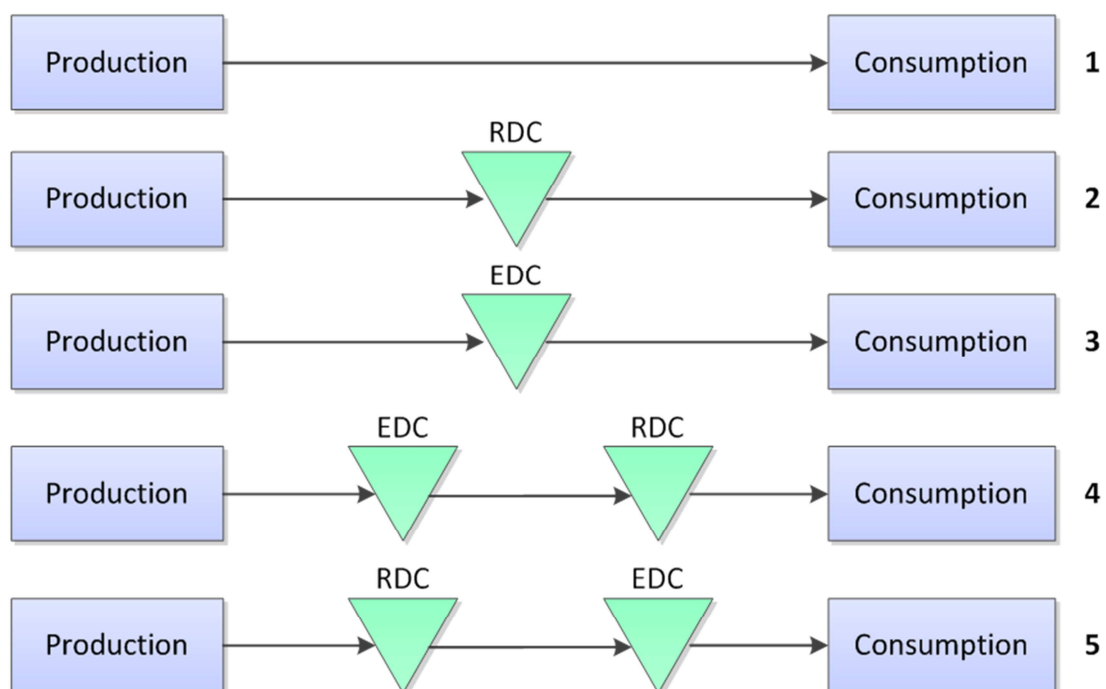


FIGURE 2.9. Conceptual representation of alternative logistics chains of SMILE+

The logistics module determines not only the logistics chain type, but also for those chains involving distribution facilities, it determines locations of the RDCs and EDCs. The chain type and locations of the distribution facilities are determined in the multinomial logit model,

which uses the disutility (costs) of the choices considered. The SMILE+ logistics module uses detailed datasets on the following classes of data:

1. **Logistics families.** The logistics families group together different commodity types that use similar forms of logistics and distribution processes. The physical properties of goods and typical logistics organizations around these groups play an important role in determining the logistics families. The total of 504 commodity types are grouped into 50 families using the following categories:
 - a. Delivery frequency
 - b. Stock replenishment strategy indicators
 - c. Levels of Safety Stocks and Seasonal Stocks
 - d. Weight-volume density (ton / m³) and packaging density (units / m³)
 - e. Lead time
 - f. Monetary value density
 - g. Shipment size
2. **Regional properties.** The regional attractiveness for the distribution activity plays an important role in SMILE. The attractiveness is partly determined by the location-related costs, while other factors have relation to the non-monetary terms such as regional accessibility and the centrality factor of the region. The model takes into account the following region-specific factors:
 - a. Intensity of regional economic activity
 - b. Accessibility index: the model determines the index endogenously
 - c. Regional flexibility: the model determines it assessing availability of different transport modes and the degree of regional connectedness
3. **Logistics costs.** The logistics costs determine the utility (disutility) of the choices with respect to the logistics chain form and the location of distribution facility. The logistics costs consist of the following components:
 - a. Transport costs (transport networks and transport systems), which depend on the network distances and transport mode used, as well as availability of the services
 - b. Stock keeping costs. These costs are determined by many factors, including the share of the goods kept in stock, monetary value density, replenishment strategy, and other factors, that incorporate the knowledge on how each logistics family is stored and replenished
 - c. Warehousing and distribution costs. These costs do not depend on the logistics family, but related to the throughput and storage of goods at the facilities. The costs are computed based on the following parameters: costs of storage per volume unit per time unit; costs of handling in and handling out per unit.

There can be made three observations as *the concluding remarks* on the SMILE / SMILE+ model. First is that the model is a comprehensive industrial scale model, covering all modeling steps from determining production volume per region per commodity, up to the realization of the transport flows on the infrastructure networks. The logistics module of the model gets the necessary input in the form of trade flow from the distribution module. The second observation includes the fact that the logistics module is very elaborate: it distinguishes 5 types of logistics chains, including two multi-echelon designs; and it works

with sophisticated cost data elaboration, which depends on many factors described above. The third observation is related to the fact that input data related to the logistics costs are not based on observations, but on the expert opinion. Furthermore, the model has been calibrated only at the level of total system-wide indicators. It has not been calibrated with respect to the endogenously estimated transport flows. Therefore, the main gap in the very detailed and sophisticated SMILE model is *the uncertainty of the empirical validity of the model with respect to the flows generated by the DC's and the locations of DC's*. The successor of the SMILE model, the SMILE+ model has not been reported in scientific literature.

2.3.2. Physical Distribution Channel Choice Model (PDCCM)

Kim et al. (2010) proposed a 'physical distribution channel choice model', which models in an empirically proven way the logistics path of goods travelling from the production facilities to the points of consumption or further rework. The model distinguishes 4 classes of logistics chains (or distribution channels in the authors' terminology). These 4 classes of logistics chains are as follows:

1. Direct shipment
2. Shipment with an intermediate stop at the distribution center
3. Shipment with an intermediate stop at the wholesale center
4. Shipment with an intermediate stop at an agency

The authors also include multi-echelon distribution structures, up to the depth of 2 echelons, namely logistics chain classes 2, 3 and 4 may involve one shipment leg between agency, wholesale center or distribution center. This detailed distribution chain classification is made possible by detailed transport data available to the researchers. The data distinguishes a number of location types, which are made compatible to the location types used in this research project (see Chapter 4 for more details on location types in the Dutch road transport survey data). The production, consumption and distribution location types are matched in the following list by the location classification used by Kim et al.(2010).

1. Production location
 - a. Factory
 - b. Ready mixed concrete factory
 - c. Quarry
2. Consumption
 - a. Factory
 - b. Industrial consumer
 - c. Retailer (retail store)
 - d. Construction site
 - e. Quarry
3. Distribution
 - a. Distribution complex, distribution center
 - b. Professional logistics company, sales company, wholesaler
 - c. Agency

Essentially, the authors estimated two logit models. The first model is a binary logit model, which determines whether the chain is direct or involves intermediate step(s) at the distribution location (i.e. chain class 1 or one of the 2-4 classes). The second model is a multinomial logit model that determines which class of the supply chain is used (1-4). The authors estimated both the binary and multinomial logit models on the following data:

1. Market characteristics
 - a. Population density
 - b. Density of firms
2. Product characteristics (there 340 samples of products and related businesses), distinguishing the families of products based on the following overlapping classes
 - a. High value product (12% of records)
 - b. Consumer goods (50% of records)
 - c. Producer goods (45% of records)
 - d. Food and drinks (53% of records)
 - e. Industrial input commodities (44% of records)
3. Business size distribution (based on the same sample of 340 records)
4. Type of distribution channel (logistics chain class)
 - a. Direct (47% of records)
 - b. Via distribution center (18% of records)
 - c. Via wholesale store (21% of records)
 - d. Via outsourced logistics (agency) (14% of records)

The calibration of the model yields positive results in respect to determination of the logistics chain class. The authors report that under the significance level of 90%, the model reaches R^2 of around 0,45 for the binary logit model (choice between direct and indirect shipment). For the multinomial logit model, the goodness of fit R^2 is around 0,28 under the same significance level.

The analysis of the work of Kim et al. (2010) draw the following conclusions. The authors presented an empirically validated case of a logit model application for the determination of the physical distribution channel, namely whether the products are shipped directly, or via logistics chains involving distribution structures. The model is validated on a Korean shipper survey containing 340 recorded shipments with information about product, company and logistics path. The logit model uses product and shipper characteristics in order to determine distribution channel. The model does not attempt to use (generalized or total) logistics costs as the basis for logit utility. As there is no logistics cost function involved, the authors did not attempt to determine the spatial organization of logistics chains, namely *the model does not provide any indication on where the distribution structures are*. The absence of the spatial component in the model does not make it suitable for translation of trade flows into transport flows in the logistics layer of the freight modelling framework. However on the positive side, the authors demonstrated applicability of the nested logit models for the determination of logistics typology.

2.3.3. Integrated Model for Estimating Emissions from Freight Transport (IMEEFT)

Maurer (2008) designed a transport model for assessment of the environmental policies in light with the environmental goals of the Kyoto Protocol (for more details see, for example, UNFCCC, 2007). The Kyoto Protocol sets the limits of Green House Gasses (GHG) emissions, such as CO₂ emissions, for the developed nations and allows the results of carbon mitigation projects in developing nations to offset CO₂ emissions of the developed nations. With transport and logistics currently contributing to an estimated 20 to 25% of overall global CO₂ emissions (Davydenko, 2014b and 2014d), the governments are interested in improved efficiency of supply chains. Transport models, at both micro and macro levels, are indeed very suitable for the purpose of ex-ante and ex-post estimations for the GHG emissions resulting from transport activities.

The conceptual modeling framework used by Maurer (2008) resembles the classical freight transport modelling framework with an extra functionality designed for GHG quantity computation as the result of transport activity and a scenario analysis module, see FIGURE 2.10. The model begins with an application of the LEFT, the LEeds Freight Transport Model (Fowkes et al., 2007), which determines the demand for transport services per distance class per transport mode and measured in ton-kilometers and vehicle-kilometers. In the second step, a trade flow matrix (PC matrix) is constructed using the gravity model. In the third step, the gravity model's output (PC matrix) is used as the input for the commercial supply chain optimization software CAST (Groenewout, 2005), which determines logistics organization and computes vehicle kilometer driven indicator. In the fourth step, the emissions module determines the physical amounts of the GHG emitted as the result of transport activities and assigns a monetary value to the emissions. The fifth step, applicable for the scenario estimation for the policy testing purposes, involves application of a scenario module (the module for policy testing), which determines the changes in model parameters as the result of an introduction of the policy under consideration. After this step, a new model interaction can be executed.

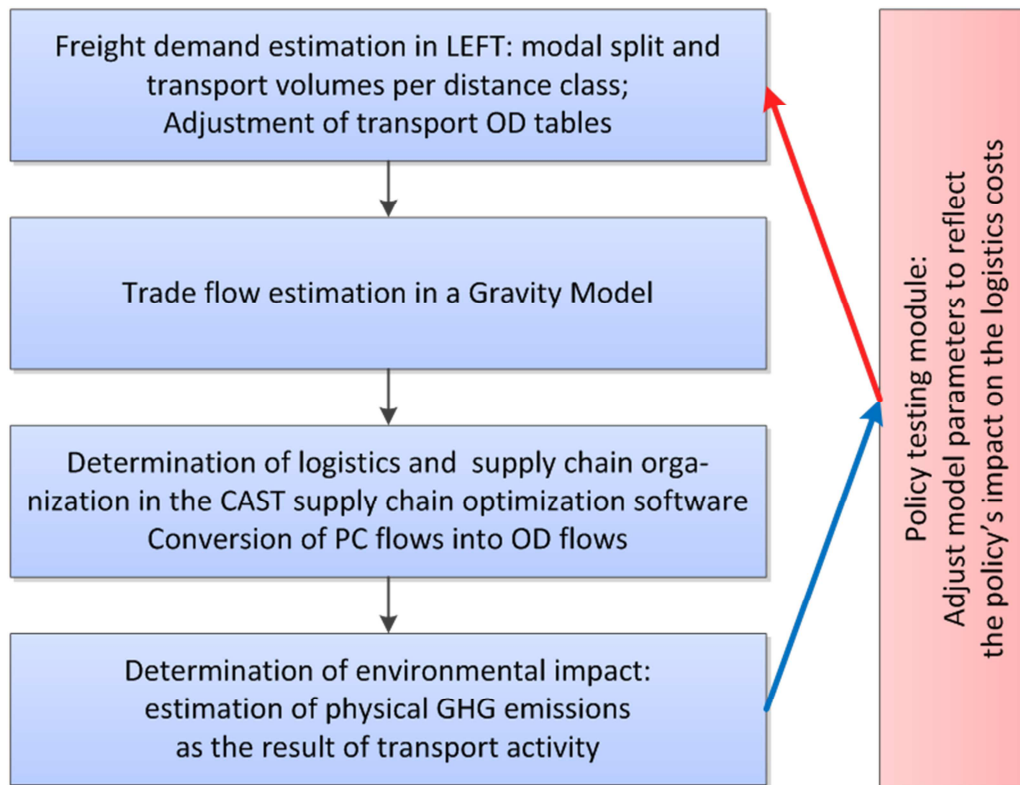


FIGURE 2.10. Conceptual representation of the Integrated Model framework (Maurer, 2008)

The LEFT model is instrumental for the environmental policy testing purposes: it computes the modal split and the total demand for transport services per distance class based on the policy-specific changes in the logistics costs, but the model does not provide a spatial distribution of the transport services, i.e. the LEFT model does not generate transport OD matrices. The transport OD flows are taken from the government statistics at the resolution level of 11x11 for the Great Britain. Subsequently, 119x119 sender-receiver relationships have been chosen at the firm level and the interregional flows assigned to these relations. The LEFT's output in the form of demand change is accommodated in the 11x11 transport OD matrix by shifting transport volumes between relations of specific distance classes. After that, the changes are propagated to the level of sender-receiver relations in the 119x119 matrix. Thus, the OD matrix is obtained by changing transport statistics flows according to the LEFT model output. Note that the transport data is limited to one commodity, Food, Drinks and Agricultural Products (FDA).

Maurer (2008) applies a gravity model to obtain a PC trade flow matrix using the regional employment in the relevant industries for the production of the FDA commodity as the generator and the regional population as the attractor. The resulting interregional flow between regions i and j was set proportional to the employment size in corresponding industry in region i and population size in region j . The proportionality factor was used to harmonize the estimations with the transport data (a fixed Handling Factor was used in order to account for distribution and warehousing). To introduce resistance, the flow between i and j were made reversely proportional to the distance. Therefore, the PC trade flow data for the FDA commodity are estimated.

The CAST software was used to determine how PC flows are shipped via logistics chains, which include direct shipments and shipments via distribution centers. The essence of the CAST software is that it is supply chain optimization software, realizing functionalities discussed in the section 1.2.2 on Logistics optimization at the company level, namely, determining the cheapest supply chain design that satisfies customer requirements. However, the CAST model needs cost data to be used in optimization. Maurer (2008) provides assumptions on the warehouse and transport costs, thus allowing CAST to optimize the supply chains and determine the resulting interregional goods flows, and importantly in the context of environmental policies, estimations on the vehicle-kilometers driven. It is not clear, however, whether these chains match with observed flows of distribution structures.

The environmental module is instrumental to the environmental policy assessment purpose of the model. In the context of freight transport modeling, this module, together with the module on translation of policies into the logistics costs are not particularly interesting and will not be discussed here. However, the demonstrated case of coupling of an environmental module to a freight model can be of interest for policy assessment functionality of other models, such as the SMILE or TRANS-TOOLS models.

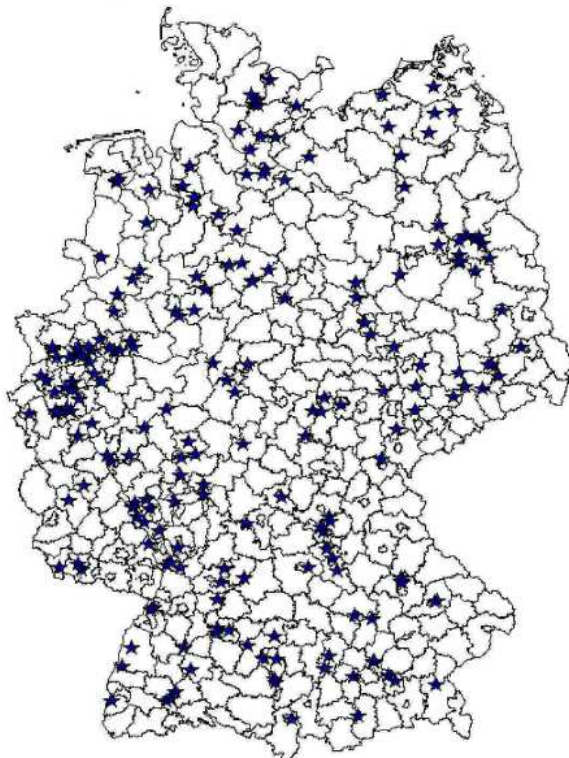
It can be *concluded on the integrated model for estimating emissions from freight transport* that the model realizes two layers of functionality of the transport modeling framework: distribution model, which is implemented in the gravity model to estimate production-consumption flows; and logistics model, which determines how production-consumption flows are realized in logistics chains and, by implication allows the determination of transport OD flows. The logistics model of Maurer (2008) is a micro normative model, which uses commercial supply chain optimization software to determine the optimal logistics configurations given known PC flows and logistics costs. This fact leads to the three relevant (in the light of this thesis) characteristics of the model. The first characteristic relates to the fact that the cost data is exogenous for the model; any estimations of the costs are just that, estimations, and cannot be considered to be true for all relations and regions. The second characteristic is related to the fact that optimal supply chain configurations (in the form of CAST output) are not always realized in practice, as the real companies operate suboptimal configurations due to practical issues, such expense to react on the constantly changing market conditions and fixed investments into existing facilities and long term labor contracts. It should be acknowledged that this characteristic does not invalidate the model's applicability for the policy studies, as it reflects on how supply chains would react to the changing cost structures as the result of policy measures, without accounting for the businesses' reaction speed. The third characteristic of the model is that it was not calibrated / validated on the observed transport flows. Maurer (2008) did not possess the data on distribution structures and had to use handling factors to determine the share of distribution-related flows. Thus, there is no proof that the model output matches flows observed in reality.

2.3.4. SYNTRADE model

The SYNTRADE model is a logistics simulation model of the food retail sector in Germany (Friedrich, 2010). The model was developed to artificially reproduce existing warehouse and distribution structures serving the food retail outlets. The objective of the model was to show

that this is possible on a large scale (for the whole economy sector), opening new possibilities for the freight demand modelling (Ben-Akiva et al, 2013). The model uses simulation by repeatedly executing the logistics decisions of the different actors in order to reach a stable state equilibrium. The logistics decisions included in the simulation are on delivery frequency, sourcing locations (thus estimating PC flows), distribution journeys (DC vehicle trips visiting multiple locations), and warehouse location choice at the spatial NUTS3 level for Germany. The model reproduces locations of the distribution facilities serving the food retail sector (see FIGURE 2.11), and provides estimations for the stock levels, reproduces trade flows, including exports and imports of the relevant commodities, and reproduces transport OD flows.

Existing warehouse locations



Simulated warehouse locations

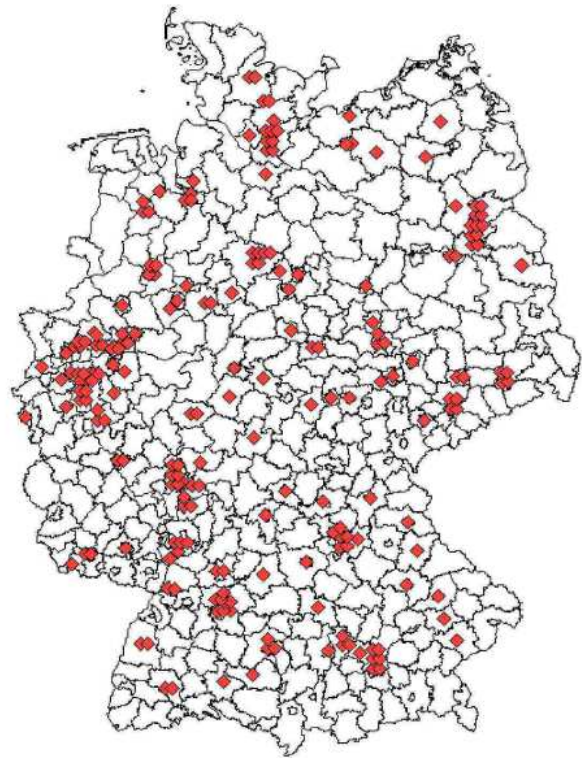


FIGURE 2.11. Real and estimated in SYNTRADE warehouse locations serving food retail sector in Germany, Friedrich (2010)

The German data related to SYNTRADE output has been used for the applications of the Logistics Chain Model (LCM) developed in this book. Section 4.4 on German Data describes the procedure used to obtain trade (PC) and transport (OD) flows for the German food retail market, however, it does not go deep into the procedures of the SYNTRADE model on conversion of the PC flow into the OD flow (the LCM application on the German data is presented in section 5.4.). Here the attention is concentrated on the part of SYNTRADE model that simulates logistics environment, determines locations of the warehouses and, by implication, determines the transport flows servicing the sector.

The SYNTRADE model considers four alternative supply paths, see FIGURE 2.12. The model design distinguishes between 3 types of distribution facilities, namely those of logistics

service providers (LSP), wholesalers (W) and retailers' own warehouses (FRC). The supply path may include a 2-echelon distribution, as it is the case if an LSP is involved in the process.

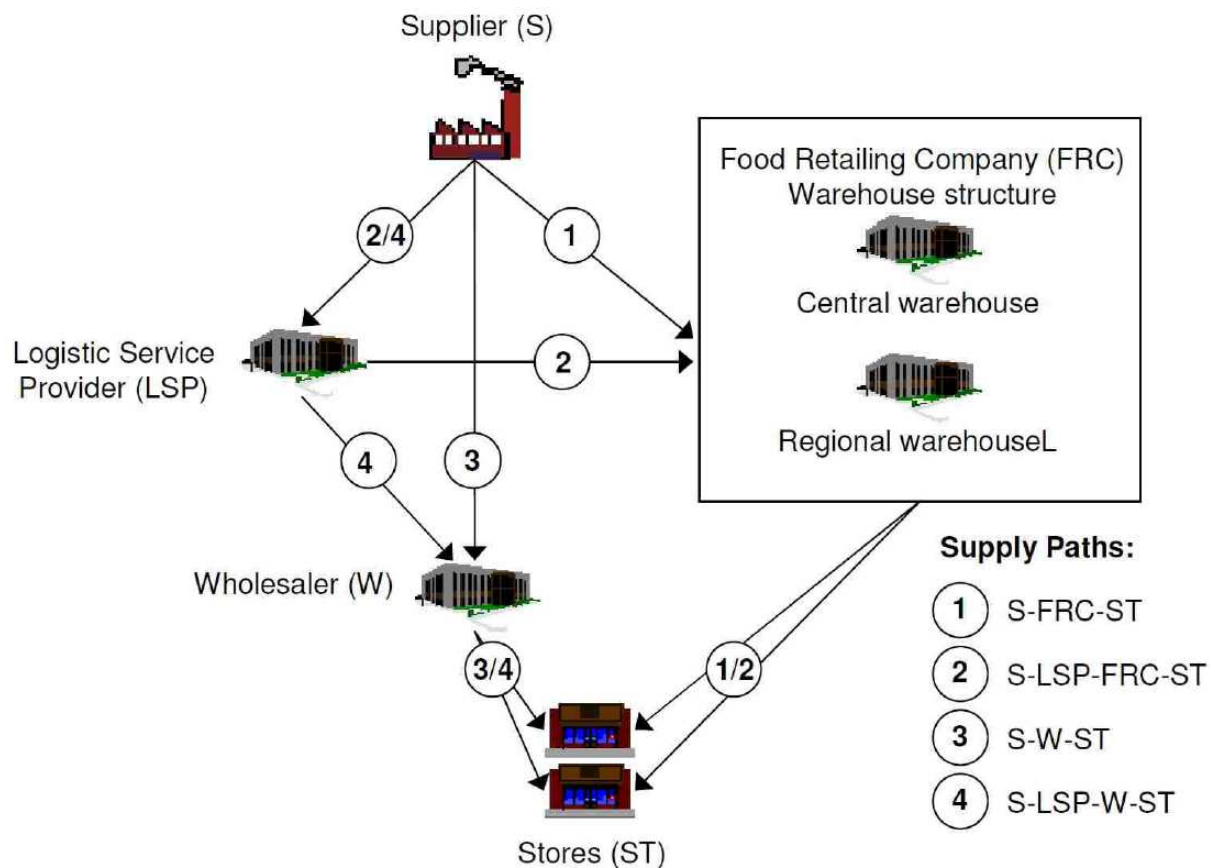


FIGURE 2.12. SYNTRADE supply paths alternatives, Friedrich (2010)

The decision on supply paths is taken based on logistic costs. For each supply path the total cost is determined, which is the sum of transport costs for each relevant transport leg within the path, storage (includes costs for the storage space as well as capital costs) and handling costs at each facility visited, cross docking costs and costs of distribution transport journeys to the retail stores. The transport and storage costs determine the shipment lot size; handling costs are throughput-based and do not depend on the shipment sizes.

The shipment lot size is determined in SYNTRADE in a cost minimization procedure, minimizing the total cost consisting of transport and storage costs, similarly to the inventory-transport costs tradeoff curve presented in FIGURE 2.2. The model uses real cost data for FTL transport between the regions; it is also assumed that the maximum lot size is an FTL vehicle. Transport cost increases per unit (pallet) transported if the vehicle is not fully loaded (LTL). The stock keeping costs depend on the time the products spend in stock, which is the lead time plus $\frac{1}{2}$ of the cycle stock related time.

Friedrich (2010) uses very detailed cost formulations for the costs arising within LSPs, such as assigning economically correct costs to the pallets shipped in bundled flows, thus modelling vital business components of the LSP's business model of bundling flows of

different clients together to reduce the costs per client. A similar problem is also considered on the side of the retailer companies: vehicles departing retailers' warehouses visit a number of stops (i.e. retail facilities such as shops and supermarkets), thus the costs are not assumed to be the costs of direct warehouse-shop, but spread over the vehicle tour visiting a number of shops (homogenous stores and daily deliveries to the stores are assumed). The number of tour stops is calculated considering the capacity of the truck and an assumed maximum number of tour stops.

The model outcome is determined in *a simulation procedure*. It consists of the four steps. First, the system is initiated (seeded) with an initial state, allocating supply paths based on default parameter values. In the second step, the supply paths of all flows not passing the warehouse structure of the food retailing sector are determined, leading to the third step, where the supply paths for each commodity flow are calculated. The third step is based on the cost minimization procedure for the flows. In the step 4, the outcome of the step 3 is compared to the supply paths states before step 2: if it did not change in steps 2 and 3, then the model reached a stable state and the simulation has determined the final outcome. The simulation outcome is most probably a "local minimum", which depends on the initial state of the model in step 1. A heuristic is used within step 3 to determine locations of warehouses and solve the allocation problem (what warehouse supplies what store). An implicit assumption here is that the inbound transport costs for the warehouses are not changing, while warehouse locations and shop assignment are based on cost minimization.

The SYNTRADE model has been *calibrated and validated* on the data from 5 companies, which includes data on warehouse locations and supply paths; the remaining data of all other German food retailing companies was used for model validation. The main model output parameter for calibration and validation is the number of warehouses per region. The essence of calibration procedure is in the choice of the initial model state (corresponding to the step 1 of the simulation procedure), such that the model output has a better fit than in the previous calibration iteration. Since each calibration iteration involves a simulation procedure, taking some 30 minutes to run, the emphasis has been given to the seeding of the initial state of the model with the data on real world supply chain configurations of the food retailers of different classes, such as a national discounter chain, a national full service chain.

Concluding on the SYNTRADE model (Friedrich, 2010), four observations can be made. First, the model is a micro descriptive model. At the core of the model is the knowledge on the populations of businesses; the model subsequently optimizes operations of these populations of businesses with respect to determination of supply paths (or logistics chains in the terminology of the LCM). The essence of the determination of logistics paths is that they lead to minimum logistics costs, under upheld service conditions such as shop replenishment frequency and the product perishability. Second, the model has been calibrated and validated, *thus proving* that a micro model can be used to describe organization of logistics for the whole economy sector. Third, the model requires extensive data for adequate representation of the population of the businesses working in the sector. It also requires understanding (tacit knowledge) on how business is done in the sector (e.g. assumptions on the daily deliveries to the shops). And finally, on fourth observation, the SYNTRADE model cannot be scaled up for the whole logistics flows of a country or Europe. It would require information and

knowledge on operations in each industry in the country under consideration, and in case of a European ambition, the exercise will have to be repeated for each country. Although theoretically possible, this amount of labor does not seem plausible for a logistics model. Thus, the SYNTRADE model presents an inspiring case of the German food retail sector; it is also a source of deep knowledge on the sector's operations in Germany, but the model lacks scalability for applications at the level of total logistics systems.

2.3.5. EUNET 2.0

EUNET 2.0 (Williams et al., 2005) is an aggregate model at the national level, which estimates (future) freight traffic flows on the infrastructure. The model includes functionality for modeling of economic activity and includes explicit treatment of logistics choices. The model starts with a regionalization of the national IO tables: the zonal demographic data, regional value added data, census on business activities and other data sources are used to perform the regionalization. The Spatial IO data is used to produce spatial trade flows per commodity group (PC flow data). The trade flows expressed in monetary terms are converted into the ton volumes using commodity value density tables (the model distinguishes 31 economy sectors and 22 commodity types; conversion is based on the European trade and transport database COMEXT). The logistics module (considered the most innovative part of the model in 2005) splits the trade flows between logistics chains, thus effectively translating trade flows into the transport flows. The model also accounts for empty vehicle trips and modal split is performed. The EUNET model ensures consistency with respect to interregional flow resistance and feeds the transport costs back into the trade regionalization module. Mode choice and choice of vehicle size are done by a hierarchical logit model. The transport flows are then assigned to the infrastructure. The model has a high degree of spatial resolution, accounting for 230 British regions, though it was applied only for the case of the Trans-Pennine Corridor in the Northern England.

Yin et al. (2005) provide a mathematical formulation of the logistics model. The logistics model uses fixed locations for the warehouses and distribution centers. The data on locations of warehouses comes from a 2001 District Census for the Great Britain. The authors use data on handling factor for each of the 22 commodity groups to determine how many warehouses and distribution centers are visited within commodity group on a trip from production to consumption. The authors refer to the results of the REDEFINE EU project (Campbell and McKinnon, 1997) as the source of the data on handling factors. Subsequently, a multinomial logit model is used to determine probabilities of possible logistics chains linking production-consumption relations. The disutility values used in multinomial logit come from the MEPLAN utility program called DERFR. The logistics chains in the logit model are commodity-specific and constrained by the data on the number of warehouse visits (handling factor).

The EUNET 2.0 has been calibrated for 2001 as the basis year. Yin et al. (2005) show good results in respect to empirical validity for three classes of data. The first class is the total system-wide ton lifted and ton-kilometer transported per 4 classes of road vehicle tonnage categories and 1 rail class, thus comparing 10 model-generated values to 10 empirically observed values, which come from various sources in the United Kingdom. The second class

of data is the data on modal split, where the model output is compared to the empirical data for 5 commodity groups, thus comparing 5 model estimated values to 5 empirically obtained ones. The third class of data is the data on two day lorry counts at 7 motorway junctions, thus comparing 7 values estimated by the model to 7 values observed in practice. The model is not shown to be calibrated and validated in respect to the distribution locations and transport OD flows.

Concluding on the EUNET 2.0 model, three observations can be made. First, the model is very similar to the SMILE+ model in its scale. The model starts with the data on sectorial input and output, regionalizes them and produces trade flows, which subsequently dealt with in the logistics module. The model includes the modal split and traffic assignment functionality. The logistics module is not detailed; the module works with 22 commodities and relies on the handling factor data for determination of the logistics chains. The second observation is that the model is calibrated on the global values only, namely on the total number of ton lifted and transported and a few traffic counts at the highways junction points. The calibration results do not show whether the model performs well at the transport flow Origin-Destination level; and it has not been shown if the model estimates reasonably well locations and throughput of the warehouses. The third observation is that although the model uses disutility values for determination of the resistance in the logit function, these values are exogenous for the model and come from an unpublished functionality of the MEPLAN software.

2.4. Literature based logistics model requirements

Section 2.2 of this chapter has outlined the broad context of the freight transport modelling field, and section 2.3 has considered five most relevant modeling undertakings, which dealt with logistics modeling of freight flows explicitly. This section analyzes the five considered models in respect to gaps, which opened up an inspiring research opportunity leading to the formulation of the Logistics Chain Model (LCM) of this book.

These five models can be assessed along the following dimensions: empirical validity, logistics costs as the choice means, model nature (normative vs descriptive), model level (micro vs macro), breadth of scope and level of detail. TABLE 2.1 summarizes assessment of the models in respect to these dimensions, and looking ahead, puts the Logistics Chain Model (LCM) developed in this thesis within the comparison framework for the illustration purposes.

TABLE 2.1. Positioning of reviewed models

Criterion/Model	SMILE+	PDCCM	IMEEFT	SYNTRADE	EUNET 2.0	This thesis
Empirical Validity of DC locations and flows	No	Yes on a limited dataset	No	Yes	No in respect to distribution and OD flows	Yes
Logistics costs as the choice means	Yes	No	Yes	Yes	Yes, but limited	Yes
Model nature	Descriptive	Descriptive	Normative	Descriptive	Descriptive	Descriptive
Model level	Macro	Macro	Micro	Micro / Meso	Macro	Macro
Breadth of scope	Very broad	Broad	Limited	Narrow	Broad	Very broad
Level of detail	High	Low	Low	High	High	Low

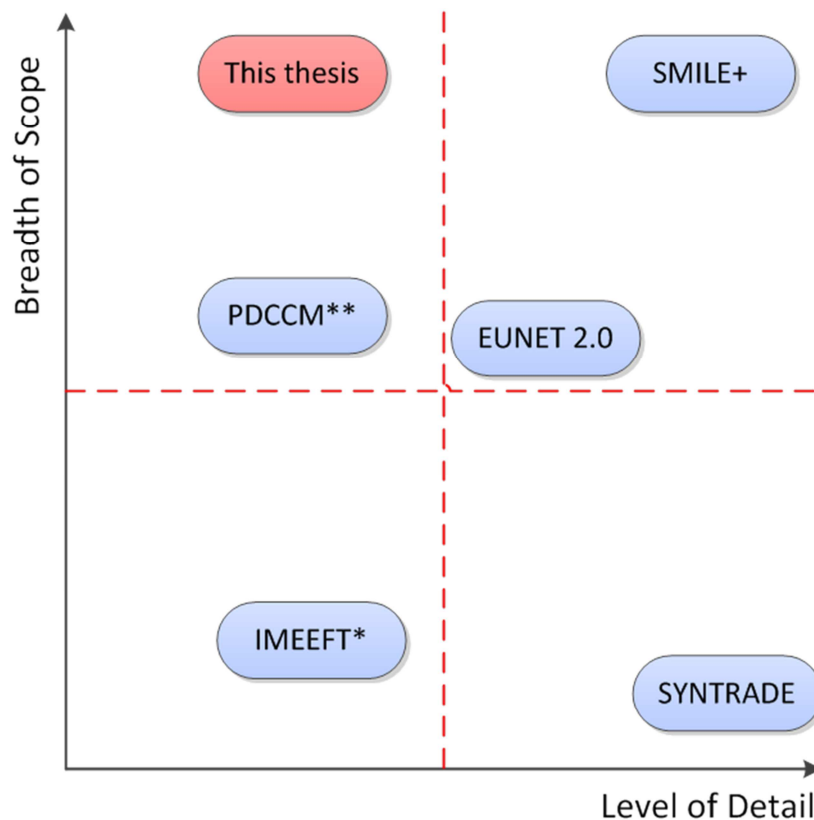
The **empirical validity** is a very important criterion for the logistics modeling at the regional level. At the micro level, when a company optimizes its logistics operations, the modeling done ex-ante and validity is ensured by the implementation of the modeling results. At the macro level, i.e. the level of interregional freight flows, empirical validity means that the model is capable of reproducing reality in the ex-post setting, reflecting the processes in the real world. The ex-post model validity creates trust that the model is capable of ex-ante assessment of policy-related scenarios, as without a proof on empirical model validity, one would have doubts in the modeling results. Amongst the five reviewed models, only two of the models have been calibrated on empirical data in respect to distribution. The limitations of other contributions are that the models have not been calibrated at OD level or the works have not been scientifically published. *Therefore, the first model requirement is that **empirical validity should be ensured.***

The use of **logistics costs** as the choice means in the logistics model has a two-fold purpose. First, the costs are the main driver for logistics choices in practice. Cost minimization is the main driver behind businesses' optimization efforts in respect to logistics choices. Second, the use of logistics costs as the choice means in the model makes it suitable for the policy analysis studies, which is the main application area of the freight transport models. Among the five reviewed models all but one use the logistics costs as the choice means, however, to a different degree of sophistication (i.e. empirical basis for the costs, distinguishing between different cost components, etc). *Therefore, the second model requirement is that the **model uses logistics costs as a choice means.***

The **model nature** in the context of freight modeling can be normative or descriptive. A normative model prescribes how the world in the model scope should look like; a descriptive model aims at capturing factors and (functional) relationships that lead to the world's state as it is. The normative models can also be used to describe the world's state, as the two of five models have successfully shown. However, the use of normative models for descriptive purposes faces a dilemma: *or* a large amount of data is needed, such as a good reflection on the population of business and the practices that the businesses use, *or* the modeling results would serve illustrative purposes only, without a well-established reality representation. A

search for the right balance between data needs and empirical validity of the results lead to *the requirement on the descriptive nature of the model*.

The **breadth of scope** is related to the model's ambition in respect to geographical coverage, number of commodities and economy sectors covered and the ability to cover all mode-specific goods flows in the region under consideration. The **level of detail** concerns the model's inclusion of various facets of logistics, such as, for instance, capturing of supplier-retailer relationships, different treatment of specific commodity groups, etc. The breadth of scope and level of detail can be put together, see FIGURE 2.13 for the illustration of the positioning of the reviewed models and the LCM of this thesis.



*Integrated Model for Estimating Emissions from Freight Transport

** Physical Distribution Channel Choice Model

FIGURE 2.13. Model positioning in respect to breadth of scope and level of detail

Considering model requirements, the scope of the logistics chain model applicability has been deemed to be more important than the detail level. Therefore, the *requirement in respect to the model breadth* is that the model is capable of modeling of the majority of road transport related freight flows in the region under consideration. The model should also be able to work in an international setting: applicable not only for the Netherlands, but demonstrated for other countries as well. The required effort and model detail can be narrowed for the incorporation of the broad modeling scope.

2.5. Concluding remarks

This chapter has analyzed scientific literature, which has relation to the field of logistics modeling. Among the broad body of knowledge on freight transport modeling, five research undertakings have been shortlisted as the most relevant research endeavors. These five models have been looked at in detail in order to find a scientifically new niche for the research of this thesis, given the practical requirements for the expected research output.

It can be concluded that the current state-of-the-art does not include a freight model with an explicit treatment of logistics, which would satisfy simultaneously the following three requirements: *Empirical validity* + *Logistics costs as the choice means* + *Broad scope*. These three requirements are not satisfied at the same time by any of the five studied in detail models. The choice of detail level and nature of the model is instrumental for the satisfaction of the three criteria. Thus, it is possible to uniquely position the logistics chain model of this thesis in the research field on the logistics modeling in freight transport models.

The following chapter further develops the modeling requirements, given the identified niche. Practical, data driven solutions have had to be found for the design of the logistics chain model. Chapter 3 outlines the modeling choices made for the combined gravity-logistics chain model and provides full mathematical specifications of the models.

3. Model Specifications

This chapter introduces and specifies the logistics chain model as a combination of a gravity model and a nested logit chain model. It provides a discussion on the model requirements, explains the choices made in the design of the model and provides a full formal mathematical models' specification.

3.1. Introduction

Logistics structures such as distribution centers and warehouses deserve special attention in the domain of freight transport modeling. Distribution centers and warehouses create substantial goods flows: in the Netherlands they account for at least 14% of all loaded Heavy Goods Vehicle (HGV) ton volumes and at least 12% of all HGV trips (based on the extended CBS road transport survey, Davydenko (2011)). The traditional 4-step freight models do not capture explicitly the logistics aspects of freight transport. The modelers often assume that transport flow is equal to trade flow multiplied by a certain factor to account for transshipments and distribution. Accurate modeling of logistics requires explicit modeling of warehouses and distribution centers in production-consumption relations.

This chapter proposes two models, which are necessary to perform aggregate logistics modeling for the Netherlands and can be applied for other parts of the world, such as Germany and Continental Europe. First, the gravity model is used for estimation of trade flows (or so-called production-consumption flows). The gravity model can also be used for the estimation of transport origin-destination flows. The second model is the logistics chain model, which determines how trade flows are translated into transport flows, taking into account that some of the goods are delivered directly from the production to consumption locations, while other goods follow a path via warehouses and distribution facilities.

3.2. Model requirements

At the core of the model design there are three major requirements with respect to model domain application and empirical validity. The following list summarizes these requirements.

1. The model should link trade flows and transport flows via logistics distribution structures
2. The model should determine spatial distribution of logistics facilities such as distribution centers and warehouses
3. The model should provide empirically tested output with respect to the volumes of regional distribution flows to provide good quality estimations of interregional transport flows at the origin-destination level.
4. The model should use total logistics costs as choice criterion
5. The model should be broad in scope, being applicable for trade and transport flows of a broad set of commodities and sectors, and should be tested in different geographical areas.

As the review of the modeling practices of Chapter 2 points out, there are models that aim at capturing logistics and distribution choices. However, none of the reviewed efforts is able to satisfy all five requirements. A special attention is given to empirical validity. A satisfaction of all five requirements in one modeling suite provides a unique opportunity to bring modeling of freight flows to a qualitatively new level. Below the modeling requirements are considered in more detail.

1. The linkage between trade and transport flows. Trade flows, or production-consumption (P/C) flows are not equal to transport origin-destination (OD) flows. For instance, if a product is produced in region A and consumed in region B, it does not necessarily mean that the product is loaded into a vehicle in region A and offloaded from the vehicle in region B. In many instances, there will be a vehicle movement with the product on board first from region A to region C, where a distribution center or warehouse is located, and then from region C to region B.

It is important to underline in this context the difference between production-consumption flow matrices and transport origin-destination ones. A P/C matrix shows interregional trade flow, namely goods are produced in the region of the first index, by convention vertical row number in the matrix, and consumed (or reworked) in the region of the second index, horizontal column number of the matrix. The transport OD matrix has the same indexing structure, however, shows the number of tons loaded onto vehicles in the origin and offloaded from the same vehicle at the destination.

In the example above, if only one product produced at A and consumed at B, the P/C matrix would be equal to the OD matrix in case of direct shipment from A to B; in the case of shipment via a warehouse in region C, the P/C and OD matrices will be different with the sum of OD flow two times bigger than the sum of P/C flow. Therefore, production-consumption and origin-destination flows are essentially different.

The traditional freight models that do not model logistics explicitly use a waterfall approach, going directly from trade distribution sub-models, which estimate P/C flows into the modal split sub-model, assuming that OD transport flows are equal trade P/C flows multiplied by a certain constant to account for distribution and other logistical applications. This is essentially incorrect, as distribution-related flows are not structurally equal to the trade flows, and

because distribution centers and warehouses are often located in other regions than production and consumption. Chapter 4 shows that the properties of transport flows to distribution and to customers are different: the flows show a difference in distance decay and transport demand price elasticity.

Therefore, we need a model or modeling technique that properly accounts for distribution-related flows, especially given the fact that distribution and warehouse related transport flows and vehicle movements represent a sizable share of all road transport. Such a model is very well fitted into a waterfall large-scale “industrial” models such as TRANS-TOOLS and SMILE. These models already have a logistics sub-model, which make incorporation of the model described in this chapter relatively easy. The practical added value of the new logistics model is in the achievement of empirical validity.

2. Determination of warehousing and distribution facility locations. There are a number of factors which make understanding of spatial distribution of logistics facilities important. They are mainly policy related. Distribution facilities and warehousing sectors do not have a clear positive or negative political image: these facilities are very good for the provision of work places, especially for low education level work, as they bid a substantial number of labor-intensive workplaces such as order pick up, loading and offloading operations, etc. The warehousing sector may also attract other economic activity, as it provides physical infrastructure for the movement of goods. On the negative side, distribution and warehousing sectors create logistics sprawl (Dablanc 2011), contribute towards freight and trip generation. These extra flows are worsening the local environmental conditions through emissions of hazardous particles and gasses, and worsen local traffic conditions by increasing the number of vehicle-kilometers driven.

A model that links trade flows and transport flows via distribution facilities is, therefore, very useful for understanding of the warehousing and distribution sector, as well as for elaboration of the policy measures through consideration of the scenarios. In practical terms, such a model should determine the flows that go through distribution facilities at the regional level and provide estimation mechanisms to consider what-if scenarios. The model should be able to work with changing transport costs, warehousing costs and determine / allow changing regional warehousing attractiveness.

3. Empirical validity. There are models of logistics chains and distribution; however, these are without a good empirical proof of the modeling efforts. For instance, the SMILE model (Tavasszy 1996) incorporates a detailed logistics reasoning, however, the model output has not been verified by empirical observations at the level of transport origin-destination flows. For these models the question on how good they represent reality remains open.

The empirical validity of a model assumes that the empirically observed data are matched by the model’s output with a certain accuracy level. This means that there should be empirical data available in the form of the model output, which is in the domain of transport modeling often a serious challenge. Therefore, an empirically tested model should provide output in a form that could be matched with the available data.

The match (fit) between model output and observational data can be improved through model calibration. The process of model calibration searches for model parameter values such that the difference between the observed data and computed model output is at a minimum. The result of the calibration is not only a good quality model output, but also model parameters, such as sensitivity factors and cost-related variables.

In the context of the logistics chain model, there are two levels of empirical validity used. The first, a simpler level, refers to the validity at the level of regional distribution throughput, or regional flow generation. For a model that covers n regions, empirical validity means that a set of n items of model output is matched with a set of n items in observational data. The second level of validity is at the level of OD flows, namely model output in the $n \times n$ matrix form is compared to the observational data matrix of the same dimensions. The empirical validity of the model is under condition of realistic values of the estimated parameters: model parameters such as logistics costs and choice sensitivities link the mathematical model to the reality of the logistics systems. With the same number of model variables, it is substantially harder to have a close match between observed and computed data at the flow level.

4. Logistics costs as the choice means. It has been demonstrated in Chapter 2 that logistics operations are mostly cost-driven, under the condition of satisfaction of the target customer service level. Therefore, the choices that the model represents, should be based on the concept of generalized (or total) logistics cost, see Tavasszy et al (2010). Transport costs and inventory-related costs should be included in the modeling scope.

5. Breadth of scope. The requirement on the breadth of scope concerns geographical applicability of the model, and its applicability to the freight flows. In respect to the geographical model coverage, the minimum requirement is that the logistics model should be able to capture logistics systems of the Netherlands. It can further prove usability of the model if other geographical regions are also taken into the modeling scope. With respect to the breadth of freight flows, the model should be applicable to a large number of commodities, as opposed to, for instance, a single commodity model (Maurer 2008) and single economy sector (Friedrich 2010). The requirement with respect to the commodity coverage is that the model takes at least all palletizable goods types into its scope, i.e. those types of goods, which are suitable for transport and handling on pallets. The requirement on breadth of scope is dictated by the policy assessment domain of the model application, as it concerns the functioning the whole transport system, as opposed to individual sectors.

3.3. Overall model design

There is a number of modeling design choices that has been considered, taking into account model requirements as described in section 3.2 and the practical aspects of model implementation. The main choices in respect to the model nature are *micro* or *macro* model and *normative* or *descriptive* model. In the context of logistics chain model, the micro model models decisions and choices made at the level of individual decision maker, which is normally at a company (firm) level. Indeed, the decisions on whether to use distribution or not are made by the company who owns or transports the goods. The same concerns the choice of distribution facility location, if it was decided to use one. Conversely, the macro model does

not look at the individual decision makers, but models the choices that are visible at the aggregate level. At the macro level, all individual decisions are aggregated (or hidden away) in the total regional flows.

It is tempting to base the logistics chain model at the micro level, the level where logistics decisions are taken. There are examples of micro-level models, which include simulation (see Chapter 2 on modeling practices review). However, with the aim of building a model at the regional level, such a model would need to be able to capture decision choices of all decision makers in the region under consideration, which is obviously not possible. Therefore, a micro model would need to deal with a 'fictive' set of decision makers. One possibility is to generate a representative set of companies in the region, divide them into realistic classes according to certain dimensions, for instance, industry, size, form of ownership, etc. Within each class, service requirements and sourcing / output decisions should be captured. Then the modeling would have to incorporate the choices that each class of companies makes and represent the logic that is used to determine the outcome of the choice, in a normative or descriptive way.

The micro level has its own difficulties. There is no database available that would consist of extensive information about companies, their activities and transport needs. This lack of information could be overcome by generating such a dataset, which is generally a possible approach, however, it would introduce substantial error margin, as the composition of companies and their transport needs, sourcing decisions and other parameters would need to be guessed or estimated. The synthesized company database would, therefore, introduce some extra dimensions of freedom, without contributing towards model's goals and requirements. The need to estimate the company population and companies' transport needs would also contribute towards model complexity, as many company-level structures and decisions have to be accounted for and properly implemented. This extra complexity would reduce implementability and usability of the model, especially in the domain of policy-making decision support tools. Therefore, the logistics model of this thesis is a macro level model.

A macro model does not need to have detailed information on individual decision makers in the domain of freight transport. It works with the aggregate goods flows, observed at the regional and inter-regional levels. The decisions of individual firms are thus hidden in the macro flows. It is also possible to work with the concept of total logistics cost, which are easily converted to utility (or disutility, as the modeling is cost-based) of the choices.

The model design choice between normative and descriptive model has also been considered. A normative model assumes optimality of the operations, which in reality translates into the (cost) optimization of the processes. For instance, a company may be willing to serve its customers with a certain service level (e.g. every day deliveries, deliveries from stock within agreed lead time) at the minimum cost. Descriptive models on the other hand, do not require full optimality of the operations, but intent to find underlying factors that explain the system.

Some experiments have been done to make the choice between normative and descriptive model designs within this PhD project. From the experiments it follows that descriptive model based on total logistics costs substantially outperforms simple normative approaches (in the

simplified normative model it has been decided to look at the Netherlands as an entity controlled by one decision maker and to solve the problem of production-consumption flow translation into transport flow via an optimal set of distribution and warehousing facilities). The conclusion of this exercise has been that with minimum data requirements, empirical validity of a descriptive model is substantially better than that of a normative one.

3.3.1. Data-driven choices

In the domain of transport modeling, data availability has been always perceived as a real challenge. The comprehensive 5-step modeling framework tries to overcome this problem, by the means of a waterfall approach in which the necessary data for transport models are generated in upstream sub-models, such as the economical and distribution models (see FIGURE 1.2). This approach works sufficiently well in comprehensive models such as TRANS-TOOLS and SMILE, however, does not provide a reasonable possibility to verify the model's output.

Coupling with the gravity model

This research project has also struggled with the data availability. With the help of Statistics Netherlands, a comprehensive road transport survey dataset has been obtained (see Chapter 4 on data for more details). However, this dataset is related to Origin-Destination (OD) transport movements, and not production-consumption (P/C) flows: in other words, the transport survey dataset is good for model calibration and validation, because it is equal to the logistics chain model output. But the transport survey dataset cannot be directly used as the logistics chain model input, as it contains transport flow data, as opposed to the necessary trade flow data. Also the data do not contain information on actual logistics chains.

The trade flow data can be estimated on the basis of the Input-Output tables generally observed by national statistics agencies (Davydenko 2013). Before P/C flows can be calculated, however, some fundamental assumptions and modeling are necessary to be made. These efforts concern the regionalization of the national Input-Output data and conversions from a sectorial classification, in which input and output of the sectors in economy are measured, to the goods type classification; the monetary units at which economic inputs and outputs are measured should be converted into units of freight (metric tons). If the problem (as it is the case of the logistics chain model of this thesis) relates to one mode of transport, the modal flows have to be isolated as well, filtering out flows related to rail and inland waterways transport modes. Once these operations are done, the flows from production regions can be matched with the flows into consumption regions (second layer of the framework in FIGURE 1.2).

Each data transformation step that converts national Input-Output tables into production-consumption flows entails data errors; some steps are technically challenging. The Input-Output tables are in the form of Euro input and output value per economy sector. The Euro-ton and sector-commodity conversion steps would introduce a substantial error. Furthermore, production and consumption must be regionalized, as normally IO tables do not have information on where (in which region) inputs and outputs are taking place. After the

regionalization, a match between production and consumption must be made (distribution model), which also takes into account the fact that the economic system of a country is not closed; there are substantial import-export activities, especially in such an open economy as the Netherlands. Also observations on trade within a country need to be available – they are not. Finally, as the logistics chain model is applicable in the domain of road transport, the other transport modalities must be filtered out from the trade flow data.

Considered together, all these transformations presented too large a challenge in this project (these issues were a part of a parallel project, see Lankhuizen et al. (2012)). Instead, the road transport survey data can also be used for the purpose of empirically valid logistics chain model. As it is discussed in Chapter 4 on data, the survey contains information not only on the region of loading and unloading, but location type as well. Thus, it is possible to estimate regional production and regional consumption based on the outgoing and incoming transport respectively. An additional advantage is that these production and consumption volumes are related to road transport, thus making the system closed.

The regional production volumes are the sum of volumes loaded into road vehicles in the region at production locations and destined to all other regions and location types. Similarly, regional consumption is the sum of volumes unloaded from the road vehicles in the region at consumption locations. Determination of regional consumption and production does not rely on the flows, but only on the sums of incoming and outgoing flows. This fact makes it possible to use the same dataset as the model input and for the purpose of model calibration: the empirical validity of the model is not affected by this fact.

The regional production and regional consumption volumes are not the necessary P/C flow input for the logistics chain model yet. The production and consumption vectors are 1-dimensional arrays, which show the sums consisting of n values each. The flow is a two-dimensional matrix $n \times n$, which still need to be obtained. A gravity model has been developed to translate production and consumption vectors into the flow matrix. The resulting model consists, therefore, of two closely linked and simultaneously estimated models, namely the gravity model to match production with consumption, and the logistics choice model, which further translates trade P/C flows into transport OD flows. As both models have model parameters to estimate, it has been implemented in a way that the two models are estimated at the same time. Besides chains, also a trade table is estimated in parallel, thus providing the necessary input for the effort on estimation of the transport chains. Therefore, the coupling of logistics chain model with the gravity model is a practical way to construct the necessary data, while adhering to the requirements on the empirical validity of the logistics model. The coupling is further discussed in the section 3.4 on model structure.

3.3.2. One echelon and multi-echelon distribution structures

There are many supply chain designs that can be observed in real world. Supply chains have evolved from a decentralized form, when production and consumption were mainly local affairs to a centralized form, which came along with globalization and decreasing transport costs, (Tavasszy et al. 2003). The pressure from fuel costs, pollution and congestion as well as demand for customized products and short lead times push the balance back towards

decentralized supply chain structures, or complex mixed ones (Davydenko 2010). However, the degree to which this phenomenon manifests itself in real logistics chains is not known. FIGURE 3.1 adopted from Davydenko (2010) presents some examples of the often used supply chain designs.

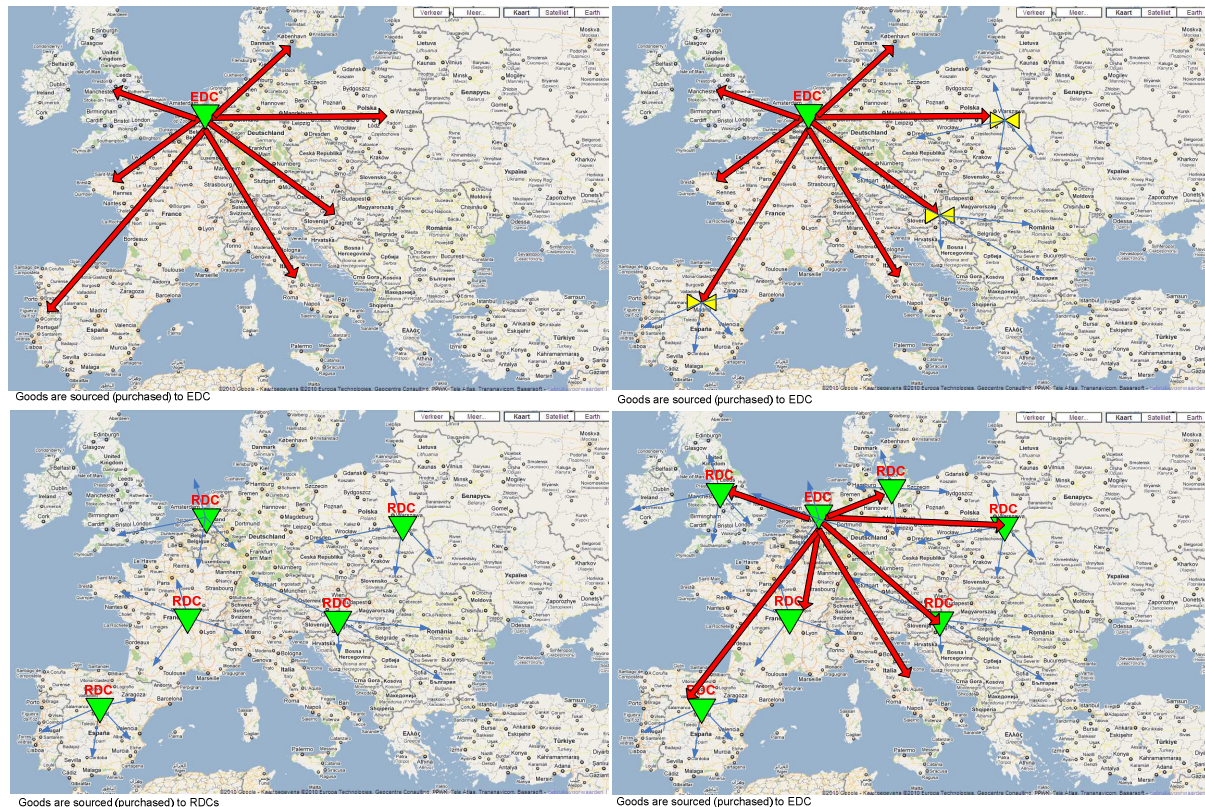


FIGURE 3.1. An example of dominant supply chain structures in Europe (Davydenko 2010)

Depending on the design of supply chain, there could be no intermediate stock points (direct deliveries); one stock point between production and consumption locations (upper left and lower left designs in FIGURE 3.1) and cases with more than 1 consecutive stock point. The road transport survey indicates a presence of the multi-echelon distribution structures in the aggregate goods flow. However, modeling of logistics chains with one echelon depth of the distribution structures is already sufficiently challenging. This thesis develops the logistics chain model that captures logistics-related flows of up to one echelon of depth, leaving modeling of the multi echelon logistics structures for the future research efforts on the subject.

3.3.3. Cost functions and location attractiveness

Tavasszy et al (2010) introduce the concept of generalized logistics costs. The generalized logistics costs consist of the three top-level cost components: the inventory costs, handling costs and transport costs. These three top-level cost components can be further detailed, for instance, the inventory costs consist of pipeline and safety stock keeping costs. These costs can be further detailed; functional relationships with the replenishment policies can be made. The advantage of the concept of generalized logistics cost is that it provides a clear conceptual structure for the cost components applicable in the realm of logistics. The cost components

considered below are the parts of the generalized logistics cost components. This thesis uses the terms “generalized logistics costs” and “total logistics costs” interchangeably.

Exact transport costs are generally not known. Shipping companies with regular transport needs often sign the so-called annual SLA (Service Level Agreements), which specify transport needs on the one hand, and transport parameters (speed, reliability, costs) on the other. These SLA are normally bilateral contracts and are not disclosed. Transport companies, on the other hand, estimate transport costs based on their internal cost estimations, such as per time unit costs, per distance unit costs and overhead. The prices paid by the shipper are related to service provider costs plus profit margin. The prices also depend on how well operations and transport needs of shipping companies are fit into transport network and structures or the service providing companies. The marketing considerations, availability of competition and the ability of the customers to pay, often play a decisive role in the formation of transport prices paid by the transport service users.

The problem of unknown transport costs can be overcome by letting the model estimate these costs during calibration, indeed within a plausible range. In addition to transport costs, the transport batch sizes are also not known (i.e. how full a vehicle is loaded with the useful cargo). Both transport costs and transport batch size can be estimated using a single estimation parameter. If vehicle-kilometer cost is fixed to a certain constant, for instance, 1,30 Euro / kilometer, then ton-kilometer cost is a function of the batch size. Thus, if vehicle-kilometer cost is fixed to a realistic number, transport batch size and transport costs can be estimated using only one parameter.

Additionally to transport costs, other costs components of the generalized logistics cost are not known. The sensitivities to the price signals (gravity model sensitivity parameter and (nested) logit sensitivity parameter) have to be estimated too. Research experiments with the model have shown that these model parameters can be estimated sufficiently well and that these estimations are stable with respect to variations in heuristics used for calibration. The research experiments have also shown that the model should contain regional distribution attractiveness parameters, which essentially represent extra costs of a location (region). These extra costs can be negative (i.e. it is necessary to decrease distribution costs in a region for the model to compute distribution flows that are matched by those observed in reality), or positive, if the model without regional attractiveness parameter overestimates the flows.

The model cost parameters are linked to the reality via one single model parameter: the cost of vehicle-kilometer driven, assumed in the model to be 1,3 Euro / vehicle-kilometer. Other cost parameters are functionally linked to it and depend on this assumption. More details on model parameters are provided in the sections 3.5 and 3.6 on mathematical formulations of the model.

3.4. Model structure

The model structure described in this section was first published in Davydenko (2013). This section borrows some materials from the paper. We use a two-step modeling approach to model regional warehouse and distribution systems. First, regional production volumes are

matched with regional consumption volumes using a gravity model. The gravity model estimates interregional goods flows in a matrix form, namely production-consumption flows or P/C flows, essentially representing physical trade flows between regions. Second, a logistics chain model is used to estimate how P/C flow is physically moved between production and consumption locations. The logistics chain model splits the P/C flow between direct shipments and shipments via warehouses, estimating throughput of regional warehouses. The model is calibrated on transport survey data in such a way that estimated annual warehouse throughput is close to the observed in real life quantities. FIGURE 3.2 shows a schematic representation of the model.

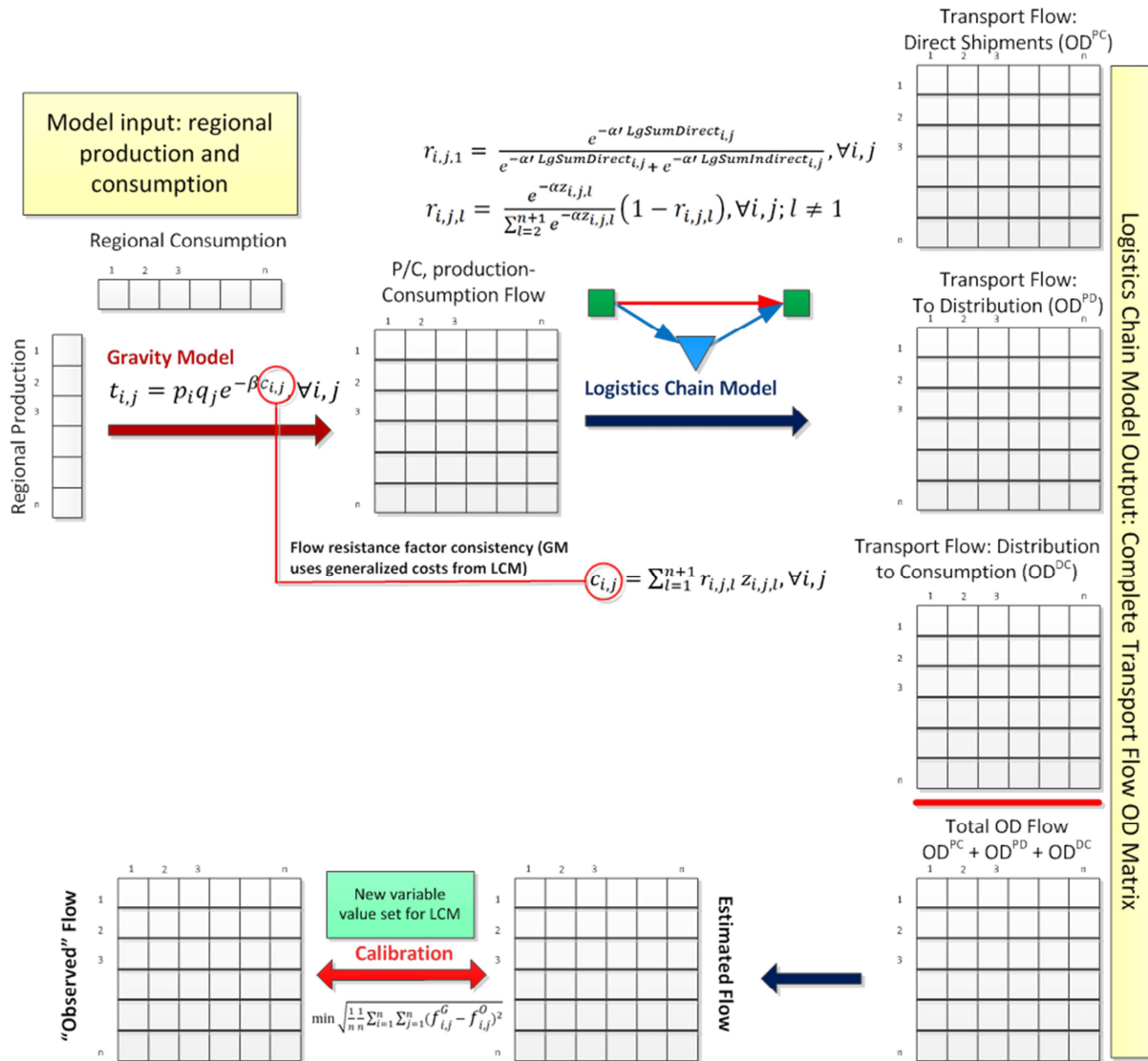


FIGURE 3.2. Conceptual representation of gravity and logistics chain combined model

FIGURE 3.2 presents a conceptual representation of the gravity and logistics chain combined model. Top left to right: the GM matches producing regions with the consuming regions in a P/C flow table; the LCM splits production-consumption flow into 3 distinct transport O/D flows. Bottom: model calibration for the interregional transport OD flows. Chapter 3 further provides a complete model formulation: sub-section 3.5 describes in detail the gravity model

in two forms: the first form is a standalone gravity model (separately calibrated) and the second form the one that is used in the combined gravity-logistics choice model. The subsection 3.6 describes the logistics choice model.

The emphasis of this research is on showing an empirically valid modeling approach, which translates trade flows (P/C matrix) into transport flow (OD matrices) by the means of determining logistics chains and locations of warehouses and distribution facilities. As there are no P/C flow data in a ready form, a combination of gravity and logistics choice model has been pursued (see section 3.3 on modeling choices and model design). The combination of gravity model and logistics choice model provides a way to overcome data insufficiency problem, while showing empirical validity of the approach.

The two models are estimated simultaneously. The combined model starts with the gravity model, which translates input regional production and regional consumption vectors into P/C flow. The P/C flow is the input for logistics choice model. The logistics choice model translates P/C flows into O/D flows of three types: PC (production to consumption), PD (production to distribution) and DC (distribution to consumption). For the Netherlands, the spatial resolution of the model is at the NUTS3 level (also called COROP in the Netherlands), for which the country is divided into 40 regions. Therefore, the input production and consumption vectors contain 40 cells each, the resulting matrices have 2-dimensional (origin-destination) structure $40 \times 40 = 1600$ cells. The resulting transport OD flow can be obtained as the sum of the three sub-flows, i.e. $OD = PC + PD + DC$. The following sections provide a formal mathematical formulation of the combined gravity and discrete choice chain models, as constituting parts of the logistics chain model.

3.5. Gravity Model

The first steps of this research project included development and application of a gravity model. The gravity model is not new; its variations have been used in various research, economical, policy and other applications. Anderson (1979) described GM as probably the most successful empirical trade device of the last twenty-five years. This statement has probably been true for more than 30 years since, and applications of the GM reach far beyond the field of trade. The gravity model has also proved to be very useful for this study on construction of logistics chains.

A first gravity model application in this PhD project was on estimation of interregional OD transport flows (see also Bergkvist and Westin (1998) and Bergkvist (2000)). This model showed that transport flows based on the CBS transport flow survey can be estimated by the gravity model with a reasonable accuracy (see also gravity model applications in the BASGOED model (de Jong, 2011)). This application also allowed determining distance decay factors, as well as regional production (push) and attractiveness values for different transport legs (or logistics chain segments). Subsequently, a modified gravity model has been used to estimate interregional production-consumption (P/C) flows, which are a direct input for the logistics chain model (see early applications such as Black (1972), and more recent Burger et al. (2009) and Disdier and Head (2008)). This section provides motivation for the usage of the

gravity model, explains underlying principles and modeling choices and provides the reader with the implementation details and modeling outcomes.

3.5.1. Essence of Gravity Models

Gravity Models (GM) use an analogy of Isaac Newton's law of gravity and apply it in various situations. The basic idea behind GM is similar to the Newtonian mechanics of celestial bodies, namely that attraction of two planets is proportional to the production of mass of the planets and reversely proportional to the squared distance between them. The idea to apply GM in the field of trade has given prominence to the GM. Typically, the log-linear equation specifies that a flow from origin i to destination j can be explained by economic forces at the flow's origin, economic forces at the flow's destination, and economic forces either aiding or resisting the flow's movement from origin to destination (Bergstrand, 1985). In other words, the flow is proportional to the "push force" of the origin region i , proportional to attraction force of the region j and reversely proportional to the size of the barriers that the goods need to overcome on their trip from i to j .

The gravity model is not only applicable for trade, but for transport modeling as well and can be seen as a particular case of gravity model application in trade. When it comes to transport systems, trade-related barriers such as import taxes, administrative processes, etc., are not taken into account, or deliberately corrected for. The basic idea behind application of the GM in transport modeling is that the resistance function is based on logistics costs, which include transport, distribution, warehousing and interest costs. Generally speaking, the logistics costs increase with distance, therefore it is possible to speak about a distance decay phenomenon.

Conceptually, the movement of goods between regions can be attributed to the price disparity between the trading regions (Erlander and Steward, 1990). If there is a difference in price for an arbitrary product, this product will be shipped from the region with a smaller price to the region with a higher price, if the costs of the product shipment are lower than the price differential. FIGURE 3.3 illustrates this concept. It assumes that the expenses of shipping of a product from location A to location B grow proportionally to distance, or other resistance measure. Profit, which can be realized as the result of a shipment, is the difference between prices in regions A and B, minus shipping expenses, see formulae 3.1.

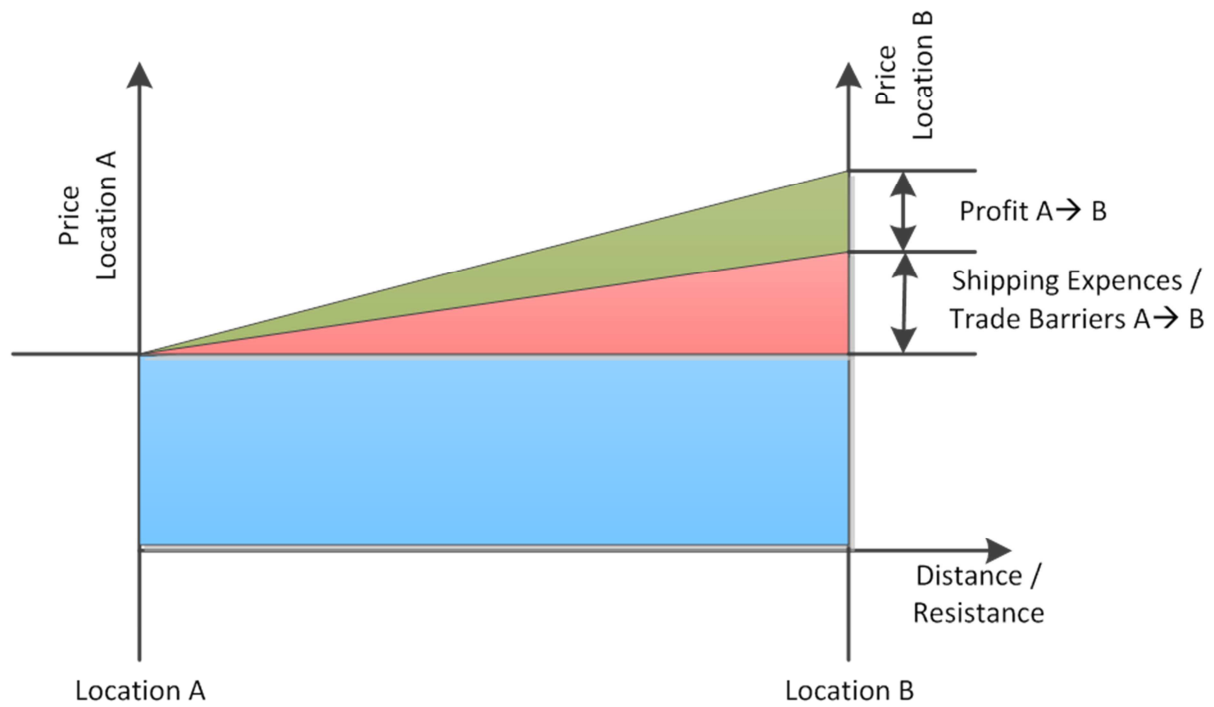


FIGURE 3.3. Conceptual representation of a price differential between regions

$$V_{A,B} = C_B - C_A - k_{AB} \quad (3.1)$$

Where $V_{A,B}$ is the profit of shipping a product from region A to region B, C_A is the price of the product in region A, C_B is the price of the product in region B and k_{AB} is the monetary equivalent of the shipping expenses of a product from region A to region B. k_{AB} can be not only direct out-of-pocket transport costs, but customs duties and trade tariffs, as well as a quantification of the non-monetary trade barriers.

This basic profit computation becomes more sophisticated if there are more than two trading regions. Supposing that there are n regions, the question is where will a product produced in region i end up? We use an exponential form in order to estimate the probability that a product produced in region i will be shipped to region j .

$$P_{i,j} = \frac{e^{V_{i,j}}}{\sum_{i,j=1}^n e^{V_{i,j}}} \quad (3.2)$$

The numerator in formulae (3.2) can be rewritten as follows

$$e^{V_{i,j}} = e^{(C_j - C_i - k_{i,j})} = e^{C_j} e^{-C_i} e^{-k_{i,j}} \quad (3.3)$$

Noting that the denominator of (3.2) is a constant, we subsequently use formulation of (3.3) to define the practical implementation of the gravity model defined in (3.4) and discussed in the section 3.5.2 on formalization of separately calibrated gravity model.

A calibrated gravity model applied to interregional goods flow is capable to estimate the “push” force of a particular country or a region on the one side of flow, and attractiveness of a country or a region on the other side of the flow. Such a model also contains estimations of

the flow resistance, which in a simple case, is a distance decay factor. Such a model describes the system in terms of the size of the regions, i.e. how strongly a region pushes products produced there in surplus, as well as attractiveness of the regions, i.e. how strong is the demand for products produced somewhere else. The resistance function is also of considerable interest: it translates costs associated with movements of goods into a resistance factor. If the cost function is based on the concept of the Generalized Logistics Costs (GLC) (Tavasszy 2010), then such a model possess *predictive power* to describe (future) goods flows based on adjusted cost components. Widely used examples of such a model application are: finding effects of fuel price change, road pricing, as well as impact of productivity growth or spatial and technological changes in logistics systems.

Two implementations of the essentially same gravity model have been developed: Separately Calibrated Gravity Model (SCGM) and Auxiliary Gravity Model (AGM). The first one, SCGM, estimates interregional transport flow O/D matrices. It takes as the input regional transport outflows and regional transport inflows in a vector form and estimates interregional O/D transport matrix. This model further compares estimated and observed O/D matrices, such that deterrence function can be estimated in the model in an optimization loop. This model shows that it is possible to estimate interregional O/D road transport flows for the Netherlands reasonably accurately (R^2 is in the range of 0,50-0,90 depending on the flow type). It also finds gravity model parameters, which are the proxies for the distance decay factor on the one hand, and lay a relationship between the distance travelled and transport costs on the other. The quality of this gravity model estimation presents the upper boundary for the quality of other freight transport and flow generation models, where a relationship is made between the costs and flow. The SCGM applications also showed that different logistics chain segments have different properties (e.g. distance decay, elasticity), confirming the assumption that these segments should not be treated uniformly in freight transport models and thus presenting the case for a logistics chain model.

The second implementation of the gravity model (AGM) serves an auxiliary function for the logistics chain model (LCM). In the absence of data on reliable and consistent production-consumption (P/C) flows, this implementation of the gravity model is used to construct these flows. The model takes regional production and regional consumption volumes from CBS road transport survey and spatially matches production and consumption. The model's deterrence function uses transport costs between pairs of regions from the LCM, where they are computed depending on the chain parameters. The output of this model implementation is P/C flow matrix, which is subsequently used as the logistics model input. This implementation does not have an ability to calibrate gravity model directly because there is no P/C flow observed. However, the output of logistics model in the form of transport O/D flows is compared to the data available in the CBS transport survey. Thus, the second gravity model implementation is not directly calibrated, but calibrated together with the logistics chain model. The combined gravity and logistics chain model is described in full detail in Section 3.6.1.

3.5.2. Specification of Separately Calibrated Gravity Model

The Separately Calibrated Gravity Model (SCGM) described in this section estimates interregional transport flows $t_{i,j}$, $\forall i, j$. Where $i = 1, \dots, n$, $j = 1, \dots, n$ and n is the number of

spatial regions in the area under consideration. For instance, the spatial resolution of the available CBS road transport survey is at the NUTS3, or so-called Dutch COROP level, thus, for the Netherlands, $n = 40$. The interregional estimated flow values $t_{i,j}$ are the elements of the O/D transport flow matrix with the dimensions 40x40. The flow is estimated according to formula (3.4).

$$t_{i,j} = p_i q_j e^{-\beta c_{i,j}}, \forall i, j \quad (3.4)$$

Where p_i and q_j are the estimated parameters of the gravity model representing regional Heavy Goods Vehicle (HGV) transport volume production and attraction respectively. β is the sensitivity parameter of the gravity model. $c_{i,j}$ is the cost friction factor in the form of road transport cost per ton shipped between origin region i and destination region j . For this model implementation, we use transport cost factor as a linear function of distance between regions i and j . The use of linear cost function showed good estimation performance of the model.

The model relies on observed transport flows, $o_{i,j}$, $\forall i, j$, which are of the same nature as $t_{i,j}$, but observed in the CBS road transport survey. Let us denote P_i to represent all road transport volumes originating in i (3.5), and C_j to represent all road transport volumes destined to j , $\forall i, j$ (3.6). The constraints (3.7 and 3.8) relate gravity model (3.4) to empirically observed flows.

$$P_i = \sum_{j=1}^n o_{i,j}, \forall i \quad (3.5)$$

$$C_j = \sum_{i=1}^n o_{i,j}, \forall j \quad (3.6)$$

$$\sum_{i=1}^n |\sum_{j=1}^n t_{i,j} - P_i| < \varepsilon, \quad (3.7)$$

$$\sum_{j=1}^n |\sum_{i=1}^n t_{i,j} - C_j| < \varepsilon, \quad (3.8)$$

The solution of the gravity model is in finding values of p_i and q_j such that constraints (3.7 and 3.8) are satisfied. In these constraints we introduced the ε term, which defines the accuracy of the gravity model. The smaller ε is, the more accurate the solution, however, the accuracy comes at the expense of computational complexity. Generally, we used ε value of 100 ton, which is accurate enough if it is taken into account that all road transport flows in the Netherlands consist of some 600 million ton lifted annually. These constraints make sure that the regional incoming transport volumes and the regional outgoing transport volumes in the estimated transport O/D matrix $t_{i,j}$ are equal within the error margin ε to the incoming and outgoing volumes observed in reality $o_{i,j}$.

Finding values of p_i and q_j .

The described algorithm searches iteratively for the values of p_i and q_j until the equalities (3.7 and 3.8) are satisfied. If the gravity model is solvable, each iteration step reduces the difference between estimated and observed regionally loaded volumes (3.9) and regionally unloaded volumes (3.10).

$$\min \sum_{i=1}^n \left| \sum_{j=1}^n t_{i,j} - P_i \right| \quad (3.9)$$

$$\min \sum_{j=1}^n \left| \sum_{i=1}^n t_{i,j} - C_j \right| \quad (3.10)$$

Step 1. Initiation. Initiate vectors p_i and q_j with arbitrary values (in the described implementation, all initial values are set to 1). Chose a single p_i or q_j value for normalization, i.e. the value will remain 1 in all optimization iterations. In the model implementation, we chose $p_{30} = 1$.

Step 2. Proceed with the following computations until inequalities (3.7 and 3.8) are satisfied.

Step 2.1. Compute estimation of interregional flows according to (3.4).

Step 2.2. Check if (3.7 and 3.8) are satisfied. If satisfied, go to Step 3. If not, proceed to step 2.3.

Step 2.3. Set new values of p_i , except for the normalized cell p_{30} .

$$p_i^{new} = p_i^{prev} \left(\frac{\sum_{j=1}^n o_{i,j}}{\sum_{j=1}^n t_{i,j}} \right) \quad \forall i \quad (3.11)$$

Step 2.4. Set new values of q_j .

$$q_j^{new} = q_j^{prev} \left(\frac{\sum_{i=1}^n o_{i,j}}{\sum_{i=1}^n t_{i,j}} \right) \quad \forall j \quad (3.12)$$

Go to Step 2.1.

Step 3. The values of p_i and q_j are found (3.7 and 3.8 are satisfied). Stop.

Computational effort and solution feasibility

It takes some 50-100 iterations of the solving algorithm described below to find the values of p_i and q_j in order to satisfy the constraints for the specified accuracy ($\varepsilon = 100$). The number of iterations depends on the initial values of the vectors p_i and q_j , which are set to 1, and on the values of exponent in formulae (3.4). For large values of the sensitivity parameter β the number of iterations required is larger. This can be explained by the fact that for a large sensitivity parameter, the model becomes very sensitive to the price signals, in other words, it redistributes large portions of the flow to the origin-destination pares where the costs are small. The rest of O/D pairs get very small flows, in essence creating sparsely populated matrix $t_{i,j}$ (only a few O/D pairs have effectively non-zero flows in this case).

A sparsely populated matrix can lead to a situation where there is no feasible solution for the gravity model. A matrix in which logical condition (3.13) is satisfied, will lead most certainly to an infeasible gravity model solution. The condition (3.13) formally specifies the case when there exist a column or a row in the observed or estimated O/D flow matrix, for which the sum of the elements is zero.

$$\exists j \sum_{i=1}^n t_{i,j} = 0 \quad \cup \quad \exists i \sum_{j=1}^n t_{i,j} = 0 \quad \cup \quad \exists j \sum_{i=1}^n o_{i,j} = 0 \quad \cup \quad \exists i \sum_{j=1}^n o_{i,j} = 0 \quad (3.13)$$

If there is a zero sum column or row in the matrix, the iterative algorithm will not converge towards a solution, for which constraints (3.7 and 3.8) are satisfied: there are no p_i and q_j values for which the constraints can be satisfied.

The feasibility issue puts some practical limitations on the application of the gravity model. The model can provide practical results if the observed flow matrix $o_{i,j}$ does not contain a column or row with a zero sum of its elements. A zero sum column or row may occur if, for example, a limited number of commodities are selected for the flow. The estimated matrix cannot contain a column or a row with zero-only values in this model formulation, however near-zero flows can be generated by a large negative value of the exponent in (3.4). For practical considerations, the iterative algorithm described in this section is limited to a sufficiently large number of iterations (step 2), to avoid a possibility of an endless looping. If the number of iterations has exceeded a predefined maximum number of iterations, the algorithm is stopped with an error message.

If the gravity model has a feasible solution, the values of p_i and q_j can be estimated to any arbitrary small error in the resulting flows: the accuracy constraint value ε can be chosen as small as it suits for the goal of model estimation. If the accuracy value ε is reduced to 1, the number of iterations increases to around 150 from around 100 iterations for the value ε of 100. In most cases such small accuracy values are not necessary.

Calibration of the model

The gravity model estimation algorithm described in this section finds the values of production and attraction vectors p_i and q_j , such that the gravity model is solved and feasible transport O/D flows $t_{i,j}$ are found. However, this model estimation does not provide any indication on how realistic these flows are, namely, how well they reproduce the flows observed in reality. The model validation and calibration can be done with the available CBS road transport survey data. The data contains observed in reality annual road transport flows. The goal of the calibration is to minimize the difference between the observed and estimated flows, as specified in (3.14).

$$\min \sqrt{\sum_{i=1}^n \sum_{j=1}^n (t_{i,j} - o_{i,j})^2} \quad (3.14)$$

Formulae (3.4) estimates the flows $t_{i,j}$: the values of p_i and q_j are found in the iterative algorithm described above: the only remaining variable to be estimated is the value of the exponent in (3.4). The exponent value depends on the costs of shipping a unit of weight (one ton) from the origin to the destination $c_{i,j}$ and the value of the sensitivity parameter β . In the calibration process, we searched for the best values of β in order to minimize (3.14). Transport costs have been considered as exogenous in the model (see subsection on various cost functions below).

A simple grid search method is used to find the value of the sensitivity parameter β satisfying (3.14). The simple grid procedure has been chosen because it was not known whether the function to minimize in (3.14) has one minimum within a plausible range of β values; the

plausible range of the β values was also not known. The nature of the problem and the speed of basic gravity problem solution permitted the use of exhaustive search in the form of grid method for the specified range of values and specified search resolution. The following subsection describes this method.

Grid Search Procedure

The purpose of the grid search procedure is to find the value of a model parameter for which the objective function is at minimum (or at maximum, depending on the nature of the problem). First, the boundaries for the values of the parameter are defined. These boundaries define the space in which the best value of the parameter is considered to be. In practice, the boundaries define a range of plausible values of the parameter, outside of which the parameter would not have a reasonable meaning. Second, the grid resolution is defined. The resolution parameter specifies the number of parameter values for which the model will be computed. For instance, if the lower parameter boundary is 0 and the upper is 10, and the grid resolution is 0,5, the procedure would assess the objective function of the model for parameter values 0, 0,5, 1, 1,5, ..., 19,5, 20. For the SCGM model, we formally define the grid search procedure as follows.

Step 1. Initiation. Set *LoV* and *HiV* search boundaries values, set grid resolution increment *inc*. Set arbitrary big number *M*, set initial value of β to *LoV*. Set β^* (best found value) to β .

Step 2. Solve gravity model, as defined above under the header *Finding values of p_i and q_j* for the chosen β value.

Step 3. Find quality of the solution from step 2. Solution quality *Q* is defined as

$$Q = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (t_{i,j} - o_{i,j})^2} \quad (3.15)$$

Step 4. Compare *Q* and *M*. if $Q < M$ (current solution is better than previous one), then set $M = Q$, set $\beta^* = \beta$.

Step 5. Check if all values have been assessed. If $\beta < HiV$, then increment $\beta = \beta + inc$ and **go to step 2**. Otherwise, **go to Step 6**.

Step 6. Grid search completed. Set $\beta = \beta^*$ (best found parameter value), solve gravity model, as defined in *Finding values of p_i and q_j* for the best found β value. The separately calibrated gravity model is solved and calibrated in respect to sensitivity parameter β .

Various cost functions

The gravity model definition (3.4) uses the cost friction factor $c_{i,j}$ in the form of road transport cost per ton shipped between origin region *i* and destination region *j*. Transport costs are generally dependent on the distance between the loading and unloading regions, reflecting reality of transport systems. In the realm of road transport, the costs are proportional to the distance between the regions. This is due to the fact that cost components related to the truck

movement are proportional to the distance: fuel used is distance-related, vehicle wear and tear is also distance-related, labor costs are proportional to the time worked. However, we studied the possibility to use more sophisticated cost function definitions, which account for possible non-linearity in transport cost. The simplest definition of transport cost between regions i and j is given in formulae 3.16.

$$c_{i,j} = d_{i,j} \frac{c_{vkm}}{L} \quad (3.16)$$

where c_{vkm} is the vehicle cost per kilometer travelled, Euro;
 L is the vehicle load, expressed in ton;
 $d_{i,j}$ is the distance between loading region and unloading region.

It is obvious that the ratio $\frac{c_{vkm}}{L}$ is the ton-kilometer cost, i.e. the cost of moving one ton of freight over the distance of one kilometer. This cost definition is the most robust cost definition. Nonetheless, an introduction of other cost components has also been studied, leading to non-linear costs with respect to the distance.

It is generally assumed that costs of shipments over relatively short distances are higher than the cost of shipments over longer distances, expressed in ton-kilometer values. In real world operations, a shipment incurs not only direct travel-related costs, but also other operations that create costs. For instance, if a vehicle has to ship goods from region i to region j , the vehicle has to be available in region i for loading. It may entail an (empty) trip to region i . Further, the vehicle has to be loaded, meaning idle time at the loading location. The same considerations may apply for the unloading region j .

The richness of the Dutch road transport survey data allowed identification of not only transport-related costs, but commodity and warehouse related costs as well. The following list presents the cost components, which were taken into account during model implementation (see section 5.1 on gravity model application to the Dutch data).

1. Vehicle cost per kilometer travelled, EUR / km
2. Trip setup cost, EUR / trip. This is a fixed cost per trip
3. Average vehicle load conform formulae (3.16)
4. Stock keeping costs in the form of generalized annual costs, % of the value of goods per annum. This component includes all Value Of Time (VOT) related costs, such as interest rate (opportunity cost), obsolescence cost, market change cost, and other.
5. Distribution center handling cost, handling in and handling out combined, EUR / ton
6. Distribution center storage cost, EUR / ton*day
7. Average DC dwell time, days
8. Commodity value density, EUR / ton
9. Average vehicle speed, km / h

In the context of SCGM, the cost factors presented in the list can be reduced to a generalization of the formula (3.17), which computes per ton cost. Indeed, if the model is applied to a transport flow of a known type, of a known commodity (or a combination

thereof), the various costs can be split between the distance-dependent part and trip-dependent part.

$$c_{i,j} = a + d_{i,j} \frac{c_{vkm}}{L} \quad (3.17)$$

The cost components represented by the items 2, 4, 5, 6, 7 and 8 can be aggregated into a single cost factor a per trip. The other cost items are distance dependent and result in the cost of ton-kilometer component. The implementation of the model allows using the detailed split up of the costs, as well as the simpler, condensed form as stipulated in 3.17.

3.6. Logistics Chain Model

The logistics chain model determines how the P/C flow is physically transported between producing region i and consuming region j . For each i, j pair we determine the fraction of flow that is loaded into HGV vehicles in region i and unloaded in region j : these are direct shipments. A share of the flow between regions i and j is not shipped directly, but via warehouses in other regions. Therefore, we determine the share of goods that is shipped via warehouses in region k , thus creating logistics chains. For the Dutch case with 40 regions, there are 41 possible ways to ship goods between two arbitrary regions i and j , namely directly from i to j and via warehouse in region k , $k = 1, \dots, 40$. We model explicitly the case when warehouse is located in region i or region j (i.e. $k = i$ or $k = j$) because it still has an impact on the flow via the chain $i \rightarrow k \rightarrow j$. The warehouse in the producing or consuming region entails extra costs (handling in and put, storage), and is realized through two transport segments (PD and DC) as opposed to a single transport segment direct shipment PC. Therefore, chains containing a warehouse in the producing or consuming region should not be treated as direct chains: they are regular PD-DC chains. Consult FIGURE 3.4 for graphical representation of the choices.

Let $r_{i,j,l}$ denote the probability that products of region i are shipped to region j via chain l , $l = 1, \dots, n+1$. Index l takes the values in the range $1, \dots, n+1$ value due to the fact that the warehouse can be located in any of the n regions in addition to direct shipments between i and j . A flow conservation constraint is introduced in order to guarantee that the flow from i to j is carried out:

$$\sum_{l=1}^{n+1} r_{i,j,l} = 1, \forall i, j \quad (3.18)$$

The probability of a direct shipment between i and j , $r_{i,j,1}$ is computed according to (3.19)

$$r_{i,j,1} = \frac{e^{-\alpha' LgSumDirect_{i,j}}}{e^{-\alpha' LgSumDirect_{i,j}} + e^{-\alpha' LgSumIndirect_{i,j}}}, \forall i, j \quad (3.19)$$

Where $LgSumDirect_{i,j}$ and $LgSumIndirect_{i,j}$ represent utility of the direct and indirect choices in the top-level logit discrete choice. These utilities are computed as logsum of the underlying nested alternatives. Alternatively, $LgSumDirect_{i,j}$ can be substituted with the cost of direct shipment and $LgSumIndirect_{i,j}$ can be substituted with the cost of indirect

shipments. In this case the cost of indirect shipments can be computed as the weighted average of the costs of each n alternatives for the location of distribution.

$$LgSumDirect_{i,j} = \ln e^{-\alpha z_{i,j,1}}, \forall i, j \quad (3.20)$$

$$LgSumIndirect_{i,j} = \ln \sum_{k=2}^{n+1} e^{-\alpha z_{i,j,k}}, \forall i, j \quad (3.21)$$

Therefore, equation (3.19) computes the probability of direct shipments $r_{i,j,1}$ and equation (3.22) computes probability for indirect shipments ($l \neq 1$).

$$r_{i,j,l} = \frac{e^{-\alpha z_{i,j,l}}}{\sum_{l=2}^{n+1} e^{-\alpha z_{i,j,l}}} (1 - r_{i,j,1}), \forall i, j; l \neq 1 \quad (3.22)$$

where $z_{i,j,l}$ is the total logistics cost (TLC) of shipment from region i to region j via chain l , per ton. α' is the logit cost sensitivity parameter for the top level choice (direct or via distribution) and α is the logit sensitivity parameter for the nested choice. Smaller values of this parameter make the system less sensitive to the cost differences, higher values of the parameter make the system react more strongly to cost or price signals.

The total logistics cost consists of two main components, transport costs and stock-related costs. In case of indirect shipment via a warehouse, the transport costs include the costs of shipment from producing region i to warehouse in region k and the cost of shipment from region k to the consumption region j . In case of direct shipment, transport cost consists only of the transport cost from i to j . The stock related costs include the costs of warehouse-related handling, such as offloading of the inbound HGV vehicle, movement of the goods to storage (in case they are physically stored at a distribution center or warehouse), the costs of storage itself (interest rate for the capital frozen in the goods, cost of warehouse storage facilities, costs of depreciation and obsolescence and other). The formal definition of TLC is given in (3.23) and (3.24)

$$z_{i,j,l} = \frac{d_{i,j} c^{vkm}}{L^{PC}} \quad \text{if chain } l \text{ is direct} \quad (3.23)$$

$$z_{i,j,l} = \frac{d_{i,k} c^{vkm}}{L^{PD}} + \frac{d_{k,j} c^{vkm}}{L^{DC}} + c^w + A_k \quad \text{if chain } l \text{ includes warehouse in region } k \quad (3.24)$$

where:

$d_{i,j}$ is the distance between centroids of the regions i and j ,

c^{vkm} is the cost of transport per vehicle-kilometer. It is a constant in the model.

L^{PC} L^{PD} L^{DC} HGV loads in ton for production to consumption leg (direct shipment), production to distribution leg, distribution to consumption leg respectively. HGV load variables are model calibration parameters

c^w is the cost per ton of warehouse or distribution center ton throughput. The stock-related cost c^w is the same for all regions; it is a model calibration parameter

A_k is the attractiveness of region k for distribution or warehousing activities. The attractiveness parameter is similar to the stock-related cost c^w , but takes different values for different regions. Research experiments with the logistics chain model have shown that the regional distribution attractiveness parameter is very important for empirical validity of the model. Without this parameter, all regions have the same attractiveness and the spatial distribution of transport flows is solely determined based on transport costs in this case. As some regions have intrinsic affinity with the distribution, transport costs only are not sufficient to accurately model the flows. Therefore, the model has been extended with the regional distribution attractiveness parameter. It is also a model calibration parameter.

The gravity model described in section 3.5 uses the cost friction factor $c_{i,j}$ in the form of generalized logistics cost per ton shipped between production region i and consumption region j . There are $n + 1$ ways to ship goods from i to j in the described logistics chain model. The friction factor $c_{i,j}$ is computed as the sum of total logistics cost $z_{i,j,l}$ of the chain l multiplied by the probability that this chain is used $r_{i,j,l}$. Equation (3.25) makes the gravity and logistics chain models consistent in the terms of costs used.

$$c_{i,j} = \sum_{l=1}^{n+1} r_{i,j,l} z_{i,j,l}, \forall i, j \quad (3.25)$$

The logistics chain model allows estimation of transport Origin-Destination OD table from the trade flow P/C table. Let $f_{i,j}^G$ denote physical transport flow between regions i and j estimated by the chain model, measured in ton volumes. We distinguish between 3 types of transport flow, namely, $f_{i,j}^{G,PC}$ production-consumption flow: goods are loaded at production and delivered directly to consumption without intermediary stops at warehouses; $f_{i,j}^{G,PD}$ production-distribution flow: goods are loaded at production and delivered to intermediate stock or distribution locations; $f_{i,j}^{G,DC}$ distribution-consumption flow: goods are loaded at warehouses or distribution locations and delivered to consumption. Note that we use index G to indicate that the flow is estimated (generated) by the model; index O is used to show that the data is observed (based on transport survey).

$$f_{i,j}^{G,PC} = t_{i,j} r_{i,j,l \in \text{direct shipment}}, \forall i, j \quad (3.26)$$

$$f_{i,j}^{G,PD} = \sum_{k=1}^n (t_{i,k} r_{i,k,l \in DC \text{ in } j}), \forall i, j \quad (3.27)$$

$$f_{i,j}^{G,DC} = \sum_{k=1}^n (t_{k,j} r_{i,k,l \in DC \text{ in } i}), \forall i, j \quad (3.28)$$

$$f_{i,j}^G = f_{i,j}^{G,PC} + f_{i,j}^{G,PD} + f_{i,j}^{G,DC}, \forall i, j \quad (3.29)$$

FIGURE 3.4 Illustrates how a P/C flow is split into an OD flow and how matrix summation occurs. Cell (2, 3) in P/C flow matrix $t_{i,j}$ is split between direct shipment (2, 3) and shipments via warehouses located in the regions 1, 3 and n . Shipments via warehouses generate two

transport legs: from production region to the region of warehouse and from warehouse to the consumption region.

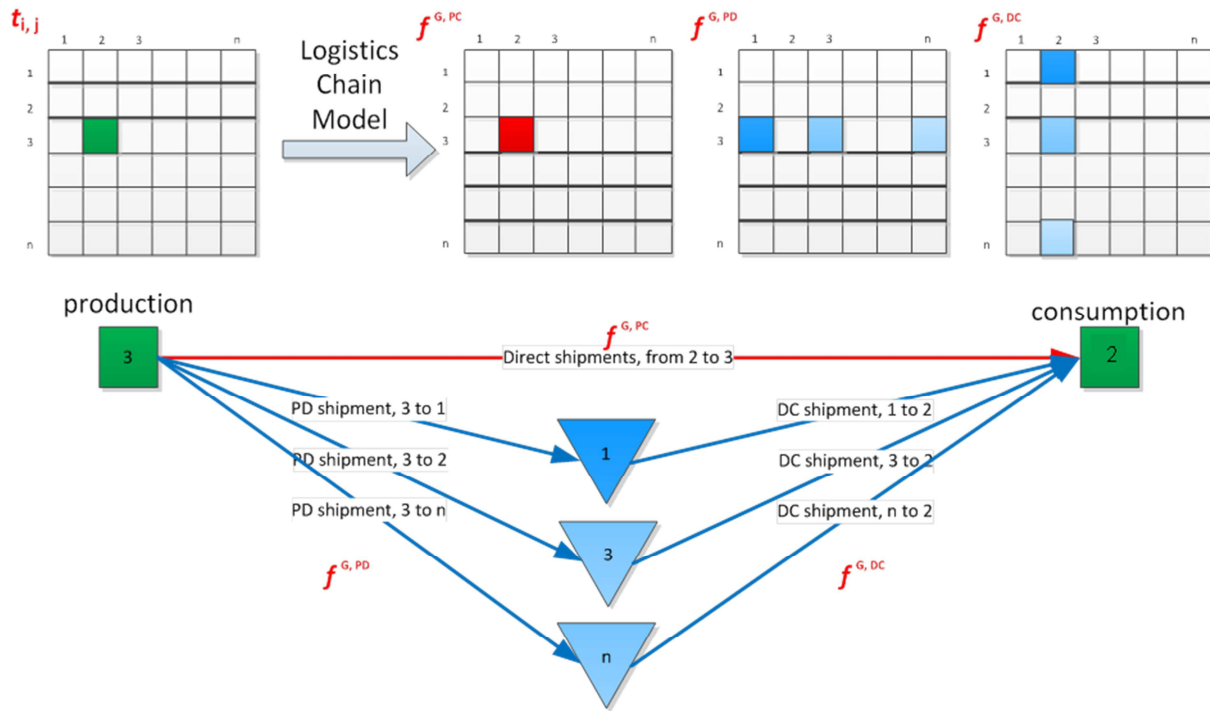


FIGURE 3.4. A graphical example of P/C flow conversion into transport O/D flow

Note that FIGURE 3.4 illustrates the flow for only 1 cell of the $t_{i,j}$ matrix. Estimation of a complete transport flow matrix requires summation of the flows generated on all production-consumption relations, i.e. for all $t_{i,j}$ cells.

3.6.1. Combined Model Calibration

The model can be calibrated in respect to regional warehouse throughput and interregional OD transport flows. In the case of throughput calibration, the model parameters are searched such that equation 3.30 is at the minimum. In other words, the Root Mean Square Error (RMSE) between estimated throughput volumes and observed in reality volumes is at the minimum.

$$\min \sqrt{\frac{1}{n} \sum_{i=1}^n (w_i^G - w_i^O)^2} \quad (3.30)$$

The model can also be calibrated in respect to the transport O/D flows, as shown in formulae 3.31. Calibration at the level of the transport O/D flows is more difficult than the warehouse throughput calibration, because the number of estimated values is 1600 against only 40 in the case of warehouse and distribution throughput.

$$\min \sqrt{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (f_{i,j}^G - f_{i,j}^O)^2} \quad (3.31)$$

The following variables were used as model calibration parameters: c^w (cost per ton of warehouse or distribution center ton throughput); the vehicle loads for the three transport stages (L^{PC} L^{PD} L^{DC}); the regional factor A_k ; the logit sensitivity parameter α in the chain model and the gravity model sensitivity parameter β . A_k is an alternative specific constant accounting for unobserved, location specific factors that hold for all P/C flows such as the existence of historical (legacy) industries and the availability of labor and infrastructure.

We applied a single variable, iterative optimization procedure. In each iteration step, the best value for each calibration variable is found. In the next iteration step, the variables are initiated with the best values from the previous step, while the search for the best value continues. The values stabilize very quickly, after 4-6 iterations, and are insensitive to (reasonable) variations in starting (seeding) values. Chapter 5 provides full details on the model application for the Dutch case, as well as model application for the German and European cases. Chapter 5 also presents and discusses model estimation results and the quality of output.

3.7. Alternative definition of the logistics chain model

Formulae 3.23 and 3.24 provide a definition of the Total Logistics Cost (TLC), which is used as (dis)utility value for the discrete choices realized through formulae 3.19 and 3.22. The TLC in specifications 3.23 and 3.24 consist of the following cost components

1. Transport costs of direct transport (PC) from production to consumption per ton-kilometer
2. Transport costs of shipments from production to the distribution (PD) per ton-kilometer
3. Transport costs of shipments from the distribution (PD) to consumption per ton-kilometer
4. Warehousing costs per ton of throughput
5. Regional attractiveness for warehousing (expressed in positive or negative costs per ton of warehouse throughput).

These components of the TLC are generally unknown, however, it is possible to assume a known range of plausible values for these cost components. These components of the TLC are estimated in the model calibration run (see section 3.6.1 on Model Calibration). Yet, the model can be estimated in a different way, where the TLC is a function of the (exogenous in respect to the model) shipment size.

Consider a shipment characterized by its size S expressed in ton. The costs of shipping it via chain l ($S * z_{i,j,l}$) can be estimated by formulae 3.23 or 3.24, depending on whether l is direct or involves a warehousing. However, 3.23 and 3.24 can be re-written using the fact that the shipment size is known, by substituting vehicle loads with the known shipment size S .

$$z_{i,j,l} = \frac{d_{i,j}c^{vkm}}{S} \quad \text{if chain } l \text{ is direct (3.32)}$$

$$z_{i,j,l} = \frac{d_{i,k}c^{vkm}}{S} + \frac{d_{k,j}c^{vkm}}{S} + c^w + A_k \quad \text{if chain } l \text{ includes warehouse in region } k \text{ (3.33)}$$

In this formulation, direct shipments will always be more attractive than shipments via a warehouse, as the sum of distances from production to distribution and from distribution to consumption are never smaller than the direct distance; and there are additional warehousing costs. However, if one of the primary functions of the warehousing and distribution is taken into account, namely consolidation of incoming flow, then it is reasonable to assume that the shipment size to the distribution center is equal or bigger than the shipment size S to the consumption location (i.e. transport movements between production and distribution contain multiple shipments), and thus formula 3.33 should be rewritten as follows

$$z_{i,j,l} = \frac{d_{i,k}c^{vkm}}{FTL} + \frac{d_{k,j}c^{vkm}}{S} + c^w + A_k \quad \text{if chain } l \text{ includes warehouse in region } k \text{ (3.34)}$$

where FTL represents the average Full Truck Load weight.

Essentially, formulation 3.34 says that goods are delivered to the distribution centers in full vehicles, and shipments from distribution to consumption carry individual shipments. Indeed, this is a simplification of the reality, as shipments to the distribution may not always be fully loaded, and transport movements from distribution may contain multiple shipments. Nonetheless, this alternative formulation makes the warehousing function more prominent and clear, than it is the case in the main model formulation.

The alternative formulation conceals the fact that the vehicle loads are unknown by the introduction of the shipment size S . However, this formulation does not introduce new information into the model, thus it needs to be estimated too, but in a somewhat different way. The shipment size S is unknown: estimating it in the same way as vehicle loads are estimated would result in essentially the same model. Because there is no empirical information on the shipment size, it is reasonable to assume that it is normally distributed with the mean S_m and standard deviation S_d .

For smaller shipment sizes S it is more attractive to send them via distribution centers such that they can travel in a vehicle together (combined) with other shipments from production to distribution with relatively small ton-kilometer costs, and then shipped from distribution to consumption individually over relatively short distances. A direct shipment for smaller shipment sizes is less attractive, as the ton-kilometer costs would be relatively high. Conversely, larger shipment sizes would prefer direct shipments, as there are no extra distribution costs involved and the distances travelled are generally shorter. The ton-kilometer costs for larger shipments are already relatively small.

The model calibration for this alternative design is conforming to the specifications in section 3.6.1 on Combined Model Calibration. The difference is that instead of the vehicle loads parameters for the three transport stages (L^{PC} L^{PD} L^{DC}), the normal distribution parameters for the shipment sizes need to be estimated (S_m and S_d). The shipment (load) size for the PD flow L^{PD} is fixed to the average load of an FTL shipment, meaning that if shipment size S under consideration is smaller than the FTL shipment size, transport costs associated with an FTL shipment are used (consolidation case); if shipment size S is larger than FTL shipment, then still FTL shipment costs are used (large shipment is split into a number of vehicle movements). The implementation and estimation of the alternative logistics chain model design is presented in section 5.2.6

3.8. Conclusions on model specification

This chapter provided a detailed specification on a combined gravity and logistics choice (chain) model. It has argued that the 4-step freight modeling framework can be extended with a logistics chain model, and can be estimated on real world observations of distribution centers as trip ends or transport O/D flows. The chapter also presented background information on modeling requirements and modeling choices that had been made at the stage of model design.

Five main requirements for the logistics chain model have been defined. First, the model should provide a functional link between trade flow and transport flow; second the model should show spatial distribution of warehouses and logistics distribution facilities; third the model should be empirically valid, such that the output of the model matches reality closely; fourth logistics costs should be used to determine size of the transport flows; and fifth the model should be sufficiently broad in flow and geographical scopes. The presented model formulation adheres to all of the five requirements.

The proposed model is a macro model, which does not look at the operations of individual companies, but works at the level of interregional flows. The model is also a descriptive model (as opposed to the normative nature of the model), and describes the interregional flows, finding the ways how the flows can be described. These choices allowed the usage of disutilities as the main model variable set. Cost functions such as the transport ton-kilometer and warehouse throughput ton cost, as well as regional warehousing attractiveness are used to determine utility of the choices that the model reproduces. The logistics design of the model is limited to one echelon logistics distribution structures to limit the problem size in order to maintain computational feasibility of the model estimation and calibration.

The macro model takes into implicitly the basic logistics tradeoffs of FIGURES 2.1, 2.2 and 2.4. The logistics requirement on demand accommodation (FIGURE 2.1) stipulates availability of goods at the locations where they are needed. The logistics systems do it in order to avoid the expensive stock-out situations. The goods can be made available by the means of frequent deliveries (expensive transport) or by the means of inventories (expensive stocks). The macro model finds a balance between these two cost components of the total

logistics cost, thus finding a balance for the costs of FIGURE 2.2. The effect of the economies of scale for the warehouses is counterbalanced by the transport costs: centrally located warehouses are cheaper in respect to warehouse-related costs (the system needs fewer), but more expensive in terms of outbound transport costs, as shipments will cover longer distances (FIGURE 2.4). The model finds a balance between these divergent forces by determining patterns (locations) for the distribution.

The gravity model itself presents an independent model of interest, as sections 3.5 and 5.1 of this thesis show. The gravity model can also be estimated independently with the good estimation fit of the model output, thus showing the validity of transport costs as the predictive parameter for interregional transport flow volume. The gravity model sensitivity parameter can serve a useful function of road transport price sensitivity.

The model parameters, such as transport batch sizes for P->C, P->D and D->C transport flows, regional attractiveness for warehousing and model cost sensitivity parameters, were estimated on empirical transport flow data. As Chapter 5 on model application shows, the estimated costs and batch sizes have realistic values.

Experimentation with the base logistics model design has led to a realization that an alternative model formulation is possible, which relies not on the estimated vehicle loads as the proxy for the transport costs, but on the shipment sizes, which determine the routing of the shipments. In the alternative model design, the consolidation function of the warehouses and distribution centers is more prominent. The model explicitly captures the logic that ton-kilometer costs for inbound shipments to the warehouses are smaller than the ton-kilometer costs of outbound shipments to consumption or the costs of direct shipments. The alternative formulation explicitly captures the logic of how the extra costs associated with warehousing still make the total chain-level costs smaller for the shipments of certain sizes. Section 5.2.6 provides details on the alternative model formulation estimation results and also provides considerations on comparability of it to the main model formulation.

Future efforts in respect to the logistics chain model development can be split into three lines. The first line is the full integration into large scale industrial models such as the TRANS-TOOLS or SMILE model lineage. The logistics chain model will contribute towards accuracy and empirical validity of these models on the one hand, and the logistics chain model will take advantage of the all upstream steps that these large models take in order to estimate trade flows on the other hand.

The second line is the enrichment of the logistics chain model with multi-echelon logistics structures and usage of a more detailed transport survey information, which promises to distinguish between different logistics facilities such as warehouses, distribution centers and cross-docks.

The third line of research is related to the enrichment of the model with respect to details of logistics operations. First, this includes explicit treatment of distribution networks, especially the fact that vehicles leaving distribution centers do not normally go to a single consumption location, but make a round trip visiting a number of locations where the goods are offloaded.

Second, the economies of scale in logistics networks should also be researched in this context: larger distribution facilities and a more dense spatial customer distribution should lead to smaller costs per unit shipped, which can be treated explicitly in the model. Third, vehicle technology innovations and societal acceptance of large vehicles, so called Long and Heavy Vehicles (LHVs), are changing the landscape of both direct deliveries and distribution-related transport flows. The effects of these vehicles can be included in the model formulations and tested on the real world data. Fourth, the effects of collaborative use of distribution facilities may be included into the model.

4. Data

This chapter describes data used in the calibration and modeling applications. The data described here form input for the Separately Calibrated Gravity Model (SCGM) and for the estimation and applications of the Logistics Chain Model (LCM).

4.1. Introduction

This chapter provides information on the following three classes of data.

1. *Data for the Netherlands.* The Dutch model applications have been the main purpose of the modeling efforts, therefore, substantial attention is given to the description and analysis of these data. The Dutch data are unique, as they are based on unpublished detailed statistics provided by the Statistics Netherlands. The uniqueness of the Dutch data is in the fact that road transport statistics are extended with the information on loading and unloading location types. It is also directly observed through survey efforts of the Statistics Netherlands, as opposed to the other datasets discussed in this chapter. The structure of the Dutch data has had a large impact on the design of both SCGM and LCM models. The Dutch dataset is analyzed with respect to its structure, quality and is validated by other types of data, such as employment statistics. Section 4.2. provides complete information on the Dutch data, and section 4.3 confirms the Dutch data quality using an unrelated dataset.
2. *Data for the continental part of the European Union.* The work of the Netherlands Environmental Assessment Agency on reconciliation of trade and transport flow data has made it possible to apply the LCM at the much more detailed spatial resolution level compared to the Netherlands. The datasets contain both trade and transport flows and have been obtained through modeling involving only one parameter (handling factor), reconciling trade and transport data. Section 4.4. provides an explanation of the data and method used to obtain them.
3. *Data for the food retail sector in Germany.* The work at the Karlsruhe Institute of Technology on freight transport analysis in food retail logistics for Germany (see Friedrich 2010), has led to a creation of trade and transport data related to flows

generated by the German food retail sector. These data are smaller in size, compared to the Dutch and European data, as it concerns one country geographically and are related to the performance of one economy sector. Section 4.5 provides a conceptual overview of the procedure that has been used to compile the German Dataset.

Following this description, Chapter 5 will focus on the estimation of the models described in Chapter 3.

4.2. Data for the Netherlands

The data described in this section have been provided by the Dutch Statistics Bureau, Statistics Netherlands (CBS). Some aspects of the road transport data, such as the quality of the location type variable, are not directly monitored by the Statistics Netherlands, therefore, some special attention is provided to the analysis of the data quality. Furthermore, this section provides information on the sampling sizes and year-on-year stability of the data.

4.2.1. Background information

Statistics Netherlands publishes so-called harmonized annual transport statistics in which all transport flows within the Netherlands are described at the NUTS3 spatial resolution level (40 regions within the Netherlands). The main relevant properties of these data sets are the following:

1. Annual flows, including flows to the Netherlands, within the Netherlands and from the Netherlands. The main constraint is that at least the point of loading or point of offloading is in the Netherlands
2. Origin and destination described at NUTS3 level, wherever possible
3. Mode of transport is specified
4. For each relation, there is information about net weight (weight without the weight of loading units such as containers) and total weight, which includes the weight of loading units
5. For each relation, a number of journey is provided
6. Good type according to NSTR 2-digit classification
7. Information on whether the goods are dangerous and the form of appearance

These datasets belong to the public domain and can be used for any purposes. They provide a complete picture of what is going on in the transport sector in the Netherlands. While CBS collected statistical information on loading and unloading location types, these public data do not contain location-specific data. The location data is important: it allows defining which part of flow on a specific relation is attributed to distribution and to other activities. The data without location specific information are much less suited for the purpose of research of this thesis.

As a public statistics bureau, CBS functions under strict confidentiality constraints. In practical terms it means that it should not be possible to trace back operations of any company from the publicly accessible statistics. CBS employs Eurostat's privacy protection rules that

state that each record in a dataset should cover operations of at least 10 companies (observations) and that any specific company should not be accountable for more than 70% of the volume on a published relation. CBS complies with the rules on indiscernibility of individual businesses in the reports it publishes. A three party Non-Disclosure Agreement (NDA) was signed between TNO, CBS and the author of this thesis to get access to the so-called CBS micro data. This agreement has allowed the researcher to work with a more detailed road transport survey dataset, which included information on the loading and unloading location types. No individual company can be recognized in the dataset; the input data, data analysis and modeling results presented in this thesis are sufficiently aggregated. This implies that the future estimations of this model can be done with public, transparent datasets.

4.2.2. Data structure and properties of location-related information

The datasets that include information on the location type are similar to the harmonized transport statistics, and contain less information on data elements non-essential for distribution modeling. For instance, there is only road transport data available. The CBS road transport data includes only heavy duty vehicles (HGV) and light duty vehicles (LGV) movements, which have the full total weight of more than 3,5 ton. The operation of vans is out of scope in the survey. The data contains the following fields:

1. Year
2. Type of flow (domestic, from abroad, to abroad)
3. NUTS3 identifier of the loading region. In the case of the Netherlands, the country is split into 40 NUTS3 regions. The data contains information on flows from and to other countries, but strict confidentiality rules of CBS prevented the use of these data in the modeling. If goods are loaded in a foreign country, the loading region is aggregated to one of the three values: Belgium, Germany and all other countries.
4. Type of loading location, in a form of a choice among 9 location types. This is the key field, which is not present in publicly available data. The possible values of this field are Production, Consumption, Sea Port, Inland Waterways Port, Rail Terminal, Airport, Entrepot, Distribution, and other or unknown.
5. NUTS3 identifier for offloading point, which has the same properties as the identifier for loading point
6. Type of offloading location, which has the same properties as the identifier for loading location
7. NSTR 2-digit code for commodity group transported between loading and unloading location
8. Weight of transported goods in ton, for this specific relation
9. Number of truck trips for the specific relation

FIGURES 4.1 – 4.3 and TABLE 4.1 present some relevant properties of the road transport flows described in the CBS dataset. FIGURE 4.1 visualizes the size of distribution activities per Dutch NUTS3 region. Intensity of the color is proportional to the ton volumes loaded at distribution location in each of the 40 Dutch regions. The goods are unloaded at any location within the Netherlands. The map on the left shows absolute distribution volumes and map on

the right side of FIGURE 4.1 shows distribution volumes per person living in the region. FIGURE 4.1. shows a high degree of distribution concentration around Rotterdam, North Brabant province and central regions of the Netherlands (e.g. Utrecht). Distribution volumes per region inhabitant are more evenly spread over the country, with North Brabant and the north of the country showing some greater per inhabitant distribution volumes.

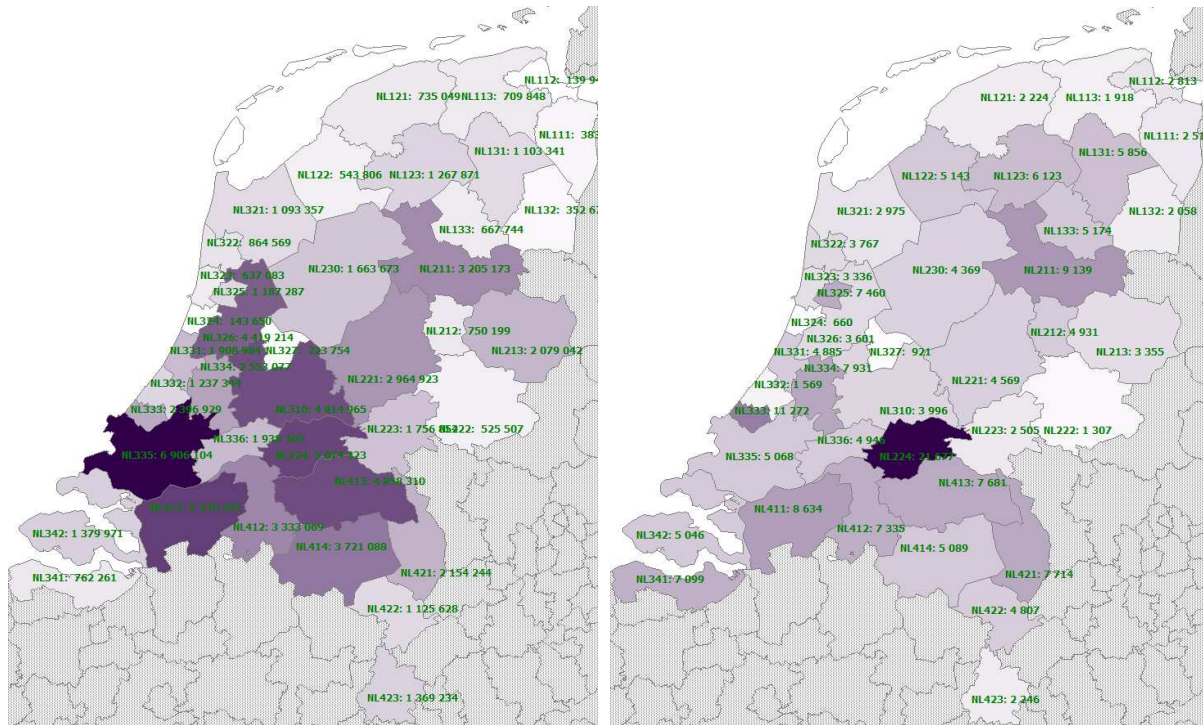


FIGURE 4.1. Annual ton volumes loaded at distribution per NUTS 3 region (left); annual volumes loaded at distribution, normalized by population size (right, kg / person) Color intensity is proportional to the volume

FIGURE 4.2 zooms into the distribution volumes generated by one region. For a selected NUTS3 region (left side is Flevoland and right side is Utrecht), the color shows the volumes originating at the distribution in the region and unloaded everywhere in the country. The figure illustrates that the two neighboring regions have completely different patterns of spatial flow distribution. Flows loaded at distribution locations in the region of Utrecht tend to stay (i.e. are unloaded) predominantly in Utrecht and, to a lesser extend in the nearby regions. Flows originating in Flevoland spread over the country, showing a less steep pattern of flow decline with the distance. It can be concluded from FIGURE 4.2 that distribution has region-specific properties.

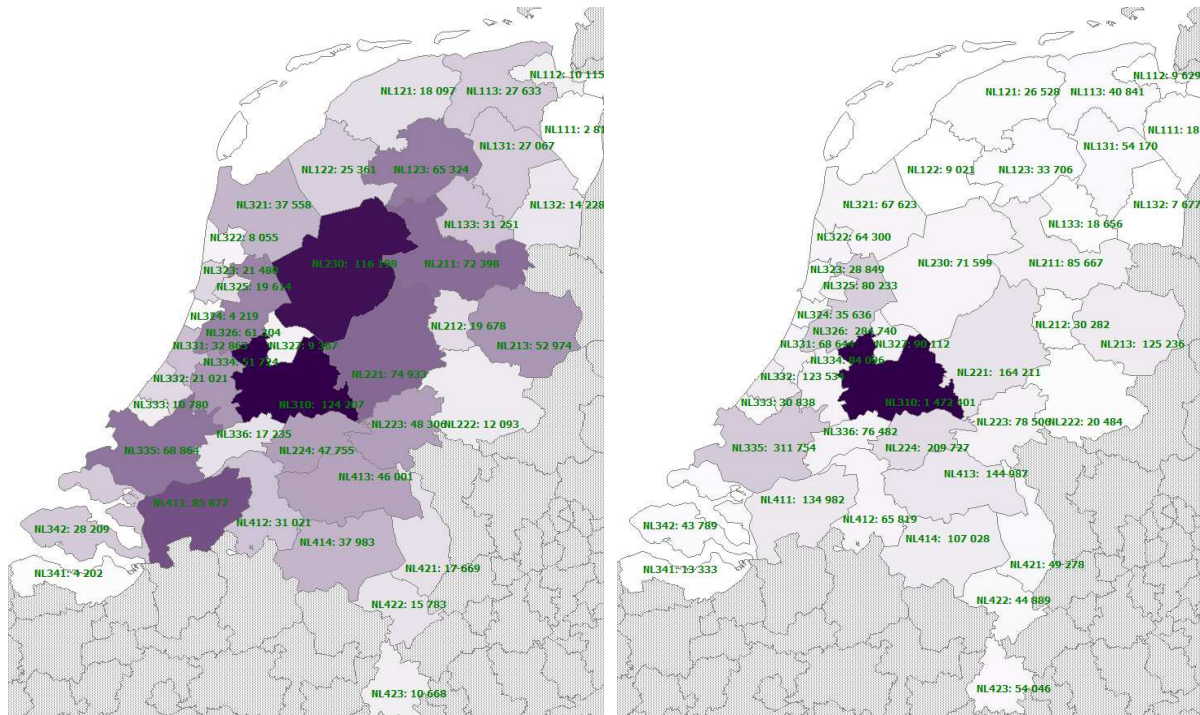


FIGURE 4.2. Annual ton volumes loaded in a region
(left – NL230 Flevoland, right NL310 Utrecht)
and unloaded in all other regions

TABLE 4.1 presents the aggregate flow statistics and flow split between PC, PD and DC subtypes of flow. For all commodities together, direct flow (PC) represents some 89% of the production-consumption flows. Some 66% of the special goods are delivered direct and in case of drinks, the chains via distribution take the majority of the flow (48% of drinks are delivered directly). For the majority of the goods, direct deliveries dominate the flow.

TABLE 4.1. Annual flow size per type of flow and commodity type
or collection of commodities

Commodity	Flow Type		
	PC	PD	DC
All commodities	196.112.085	24.112.554	27.120.893
99-Special goods	10.162.997	5.214.320	3.666.123
12-Drinks	827.755	901.899	463.716
02-potatos	3.669.585	210.642	54.811
All Chapter 9	27.245.721	9.119.529	7.182.742

FIGURE 4.3 shows that for all three flow types the size of the flow decreases with the distance. Data in the figure should be interpreted in the following way. On the horizontal axis, the distance classes are plotted. Distance class 30 relates to the flows where loading and unloading locations are in the range 0-30 km; distance class 50 relates to the flows where loading and unloading locations are in the range 31-50 km, etc. Note that the interregional distances are measured as the distance between region centroids; intraregional flows (“diagonal” flows) are not taken into account. On the vertical axis, the share of flow falling

into the corresponding distance class is shown. The share sum over all distance classes is 100%. From FIGURE 4.3 it can be concluded that direct flows (PC) have the steepest decline with the distance and flows to distribution (PD) have the less steep distance decay.

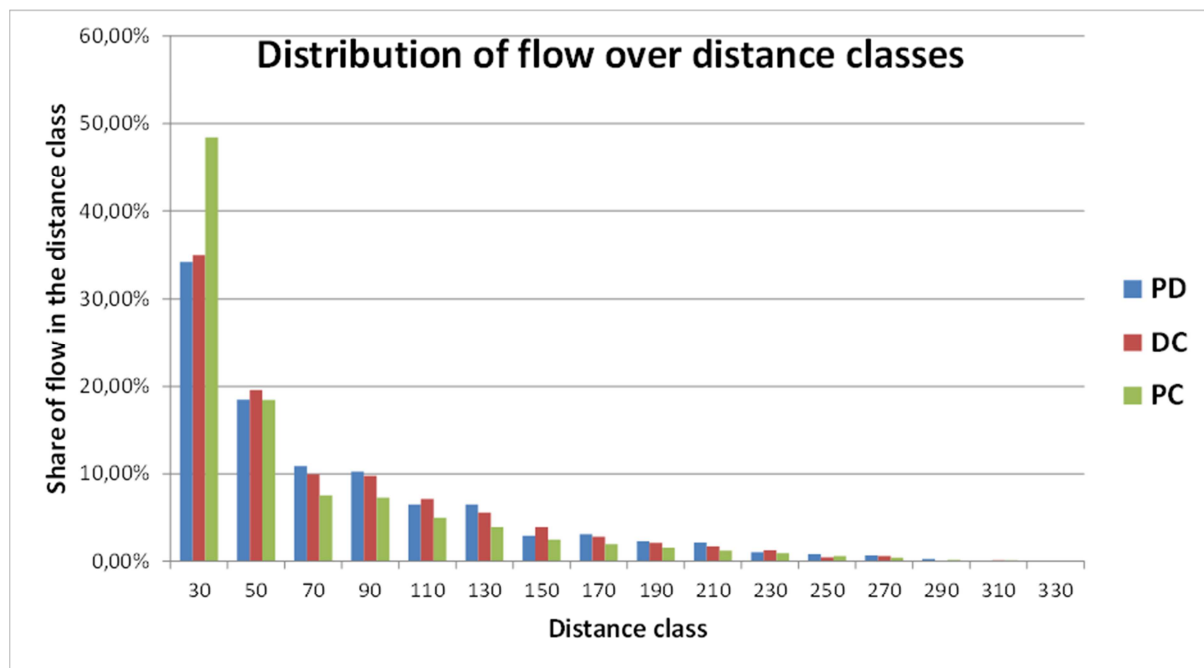


FIGURE 4.3. Distribution of flow over distance classes per flow type

The data on location type is routinely collected through statistics inquiry forms that companies fill in as a part of the government-regulated statistics reporting. Due to the fact that location type data are not published in the harmonized transport statistics, the location variable is not controlled by the CBS and thus of unknown quality. This fact has had to be taken into account as one of the factors influencing outcome of the modeling work. The location type data is of key value: there are no other reliable empirical data available on location type. Given its uniqueness, the quality of the data needed to be assessed.

The location type variable has been collected by CBS in the form of paper and online questionnaires. Professionals responsible for the provision of the statistical information to CBS specified this information manually, which justifies a cautious approach to the question on the data quality. CBS continually works on quality control and integration of transport data, see for example Linders et al. (2008). However, the location type data are not reported in primary CBS reports and are not submitted to the Eurostat statistics agency. There is also a degree of ambiguity and uncertainty with respect to the location type variable. First, the respondents can use the categories 'unknown' and 'other' as the location types. In our analysis, the 'unknown' and 'other' location types are neglected. Second, the ambiguity is related to the question on when a location is a distribution location. What exactly does it mean? The uncertainty may also relate to the fact that people providing statistical information may not have an unambiguous idea about what location the company's truck has visited. These questions provide motivation for the location variable quality analysis and for the confirmation of the data quality with an unrelated dataset, such as sectorial employment data.

Recently CBS has started using electronic data exchange with the reporting companies in the form of XML reports. This way of reporting provides substantial benefits for both CBS and reporting businesses, as it substantially reduces the costs of data preparation (for those companies who are capable of an automatic extraction of the necessary data from the company IT system) and for CBS, as the data come in a form that is understood by the processing software. However, this development will have an impact on the collection of the location type information, as location type in majority of the cases currently can only be specified by a manual input. The developments around XML threaten continuity of the current mode of provision of data on location type. However, there is an ongoing research on how the rich XML data can further be used for the modeling and policy advice purposes (see Davydenko et al., 2014c). Some algorithm-based approaches such as matching vehicle board computer GPS data with property databases might provide a reliable automatic way of generation of the location type data.

4.2.3. Quality of the location type variable

There are two issues related to the quality of the location type variable. The first issue is related to perception of people who fill in the form about location type. For instance, how to distinguish production location from distribution location, or production location from consumption location? A production location can also have a distribution facility or warehouse to store the produce; at the same time production is often related to the consumption of the input materials. The second issue is related to the knowledge of the people providing CBS with the transport data. These people may often not have specific information on locations visited by the company's trucks, or even may not bother to get such details as the questionnaire bids a choice of using 'unknown type'.

As location type variable plays the central role in the modeling efforts of this thesis, it has been very important to obtain an impression on the quality of the input data. We have used two ways to check it. The first one is to look at the stability of the location type related flows over the years. Volatility in the flows related to a specific location type would suggest some problems with it, as location types do not change very fast, as physical facilities cannot be moved quickly. This stability of location type related flows is a useful indicator, but probably not a complete one, as the random inaccuracies may cancel each other out in the aggregated data. The second way is to relate location-type specific flows to other data. If the link is stable, then this is a positive sign of reliability of the data and it is possible to draw conclusions on the quality of the data.

4.2.4. Year on year changes in location-type related flows

It is possible to look at the quality of the variable through developments over time: there are data for the years 2004, 2006, 2007, 2008, and 2009 available. First, an analysis of the variable quality has been conducted at the highly aggregated scale, namely aggregating all loading and offloading volumes per location type and analyzing their developments over time. TABLE 4.2 presents year-on-year change in incoming and outgoing transport. Due to the unavailability of data for the year of 2005, year 2006 is compared to 2004, thus the change is realized over the period of 2 years.

TABLE 4.2. Change of the ton volumes loaded and offloaded per location type on the year earlier

Type of flow	2006	2007	2008	2009
<i>Change in weight volumes in comparison to the previous year*</i>				
All location types (total change in road volumes)	0,4%	1,7%	-1,6%	-0,1%
Production out	1,2	2,3%	0,5%	14,1%
Consumption out	6,9%	8,8%	-10,5%	2,7%
Seaport out	8,0%	10,7%	-6,9%	-5,7%
Inland port out	15,0%	14,2%	14,5%	16,3%
Rail terminal out	1,9%	26,9%	-13,4%	-5,9%
Airport out	10,4%	-21,6%	10,8%	13,91%
Entrepot out	11,9%	13,4%	22,2%	1,0%
Distribution out	0,4%	2,1%	-2,4%	-7,7%
Other out	-3,8%	1,4%	-7,6%	-13,1%
Production in	-1,9%	12,1%	-7,7%	-2,9%
Consumption in	4,0%	4,5%	-4,3%	15,3%
Seaport in	4,0%	22,7%	-11,1%	-12,2%
Inland port in	13,5%	-14,6%	11,5%	4,4%
Rail terminal in	1,7%	11,5%	3,9%	-29,2%
Airport in	-1,3%	-10,2%	2,7%	-6,6%
Entrepot in	15,2%	1,1%	21,9%	-1,6%
Distribution in	-0,1%	-5,4%	5,9%	-7,0%
Other in	-2,4%	1,8%	-5,0%	-9,6%
*Note that 2006 is compared to 2004, presenting annualized percentage change over the period of two years				

There was a positive growth observed in the total road transport volumes for the years of 2006 and 2007, and a decrease of volumes in 2008 and 2009 (2009 showed almost the same volumes transported as in 2008). An interesting point is the discrepancy between annual changes of “in” and “out” volumes of the same location category. For instance, in inland ports in 2007 14% more goods was loaded than in 2006, while 14% fewer goods was taken from them. This might be explained by the sampling issue with the data, but can also be explained by the changes in the structure of the flow and relatively low volume of it: inland terminals account for only some 2,5% of the road transport volumes.

4.2.5. Sample sizes

Another (descriptive quality assessment) perspective is to look at the sampling numbers, namely how many times specific movements were recorded by the CBS. These give an understanding on the size of the sample on the one side, and stability of the number of observations on the other side. TABLE 4.3 presents the number of observations per location type and year, and provides the annual change in the number of observations.

TABLE 4.3. Number of observations per year per loading and offloading location type and change on a year earlier

Number of observations	Absolute number of observations				Change on a year earlier		
	2006	2007	2008	2009	2007	2008	2009
Production out	20 314	20 640	19 440	18 372	1,6%	-6,2%	-5,8%
Consumption out	12 974	13 517	12 189	12 070	4,0%	-10,9%	-1,0%
Seaport out	2 874	2 745	2 545	2 310	-4,7%	-7,9%	-10,2%
Inland port out	915	1 027	918	843	10,9%	-11,9%	-8,9%
Rail terminal out	272	286	190	212	4,9%	-50,5%	10,4%
Airport out	352	383	392	371	8,1%	2,3%	-5,7%
Entrepot out	4 145	4 524	5 455	5 686	8,4%	17,1%	4,1%
Distribution out	17 089	17 283	15 344	13 710	1,1%	-12,6%	-11,9%
Other out	25 038	25 351	22 647	21 011	1,2%	-11,9%	-7,8%
Production in	13 157	13 517	12 766	11 884	2,7%	-5,9%	-7,4%
Consumption in	22 999	23 326	21 424	20 863	1,4%	-8,9%	-2,7%
Seaport in	24 21	2 406	2 110	2 011	-0,6%	-14,0%	-4,9%
Inland port in	735	795	680	607	7,6%	-16,9%	-12,0%
Rail terminal in	308	336	212	247	8,3%	-58,5%	14,2%
Airport in	374	388	381	290	3,6%	-1,8%	-31,4%
Entrepot in	3 621	3 912	4 537	4 632	7,4%	13,8%	2,1%
Distribution in	15 325	15 528	13 562	12 095	1,3%	-14,5%	-12,1%
Other in	25 033	25 548	23 448	21 956	2,0%	-9,0%	-6,8%
Total	83 973	85 756	79 120	74 585	2,1%	-8,4%	-6,1%

Comparing data in TABLE 4.2 and TABLE 4.3 it can be noticed that the year-on-year change in the number of observations varies stronger than the changes in transport volumes. This might be caused by an increase in the average load factor of vehicles: an annual change in the number of observations between 2008 and 2009 is -6,1%, while the change in the number of vehicle movements is some -4% and the change in volume is only -0,12%. A special case is the situation with the number of observations to and from rail terminals, where a decrease of some 50% in loading and a decrease of some 58% in offloading number of observations is noted between 2007 and 2008, while volumes changed by -13,4% and 3,9% respectively. This might be attributed to the small sampling size of a few hundred records; moreover rail terminals do not represent a significant value for the research on distribution structures. The behavior of important location types (production, consumption, distribution and sea port) does not show unexpected patterns. The following sections will look further into the question of data quality by the means of gravity model experiments and an analysis of the relationship between distribution location type flow and regional sectorial employment.

4.2.6. OD flows and location attractiveness analysis

In addition to the descriptive analysis of the location type variable quality, it is useful to check the quality of the variable at the flow level. The analysis has been performed for the change

occurred annually between 2006 and 2007, in other words, the volumes of 2007 have been compared to the volumes of 2006 at the level of origin-destination relations. A selection of palletizable commodities has been made, i.e. those commodities that are suitable for transportation on pallets and suitable for handling at distribution facilities. FIGURE 4.4. presents the list of palletizable commodities (checked) in the NSTR-2 classification, in the form of a screenshot of a Delphi (object Pascal) implementation of the gravity model.

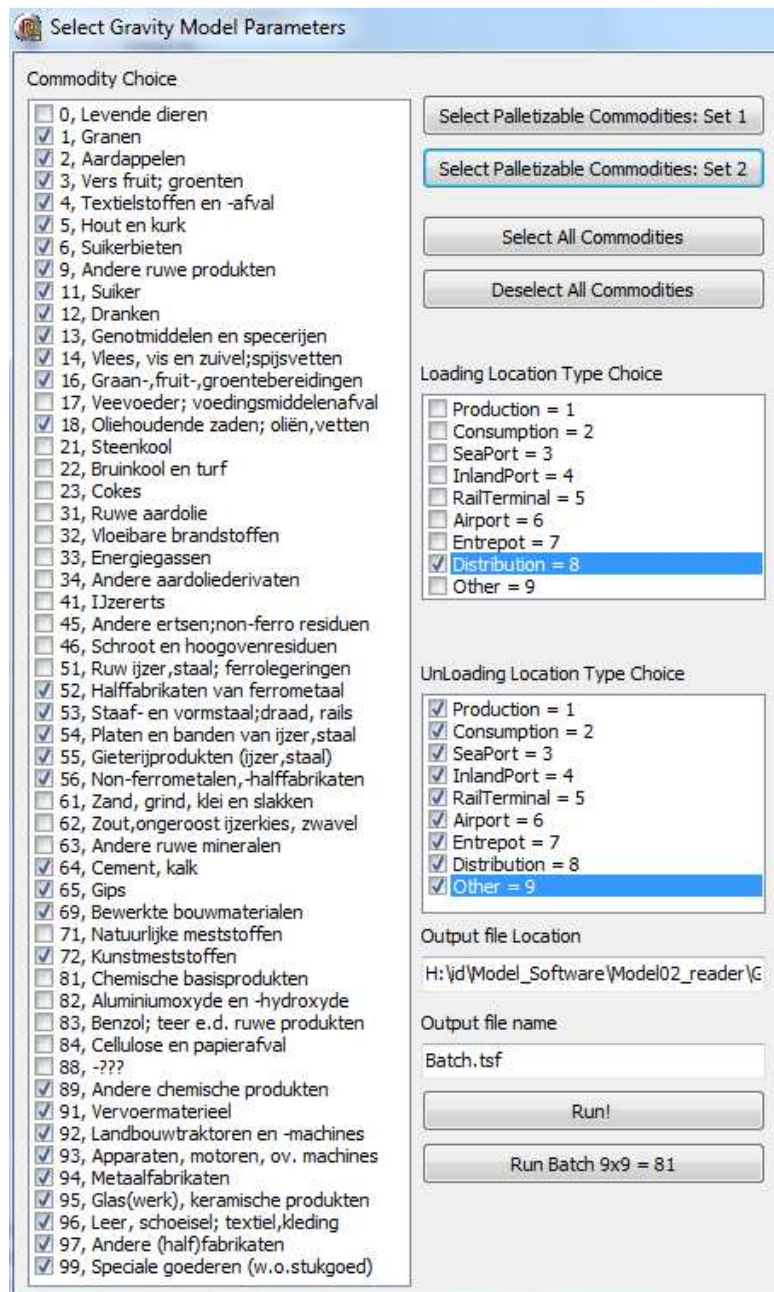


FIGURE 4.4. Selection of palletizable commodities (left list, checked items, Dutch commodity names)

At the NL NUTS3 (COROP) level, there are 1 600 origin-destination relations (40 loading regions x 40 offloading regions). The intra-region flow has not been taken into account. The two 40x40 matrices for 2006 and 2007 volumes have been converted into two single-dimension vectors, each cell representing volumes per relation.

In total there are 257 zero cells in 2006 and 245 in 2007, of which 143 cells have zero volumes for both 2006 and 2007. Zero cells stand out as an issue, as they sometimes present a problem for modeling, namely, influencing solvability of the gravity model (see section 3.5 on gravity model). It is not clear whether there was no real world flow observed on those zero-relations, or it is a sampling issue, as CBS asks companies that operate vehicles to report only one week of operations per year and then scales up the data to the annual level.

Those 16% of the zero cells in a sample do not present a problem for the modeling, however, if the selection criteria becomes more exclusive, for instance, for a flow loaded at distribution location and offloaded at consumption location, the number of zero cells in a sample is substantially larger. If the selection of commodities is smaller (e.g. less inclusive, individual commodities), the gravity model can become infeasible to solve (see section 3.5). A comparison of 2006 and 2007 volume vectors shows a correlation coefficient of 0,68, which can be interpreted as stable flow. The two vectors plotted on a graph (FIGURE 4.5), give a graphical representation on the relation between volumes of palletizable commodities loaded at the distribution location type and unloaded anywhere (all location types) in 2006 and 2007.

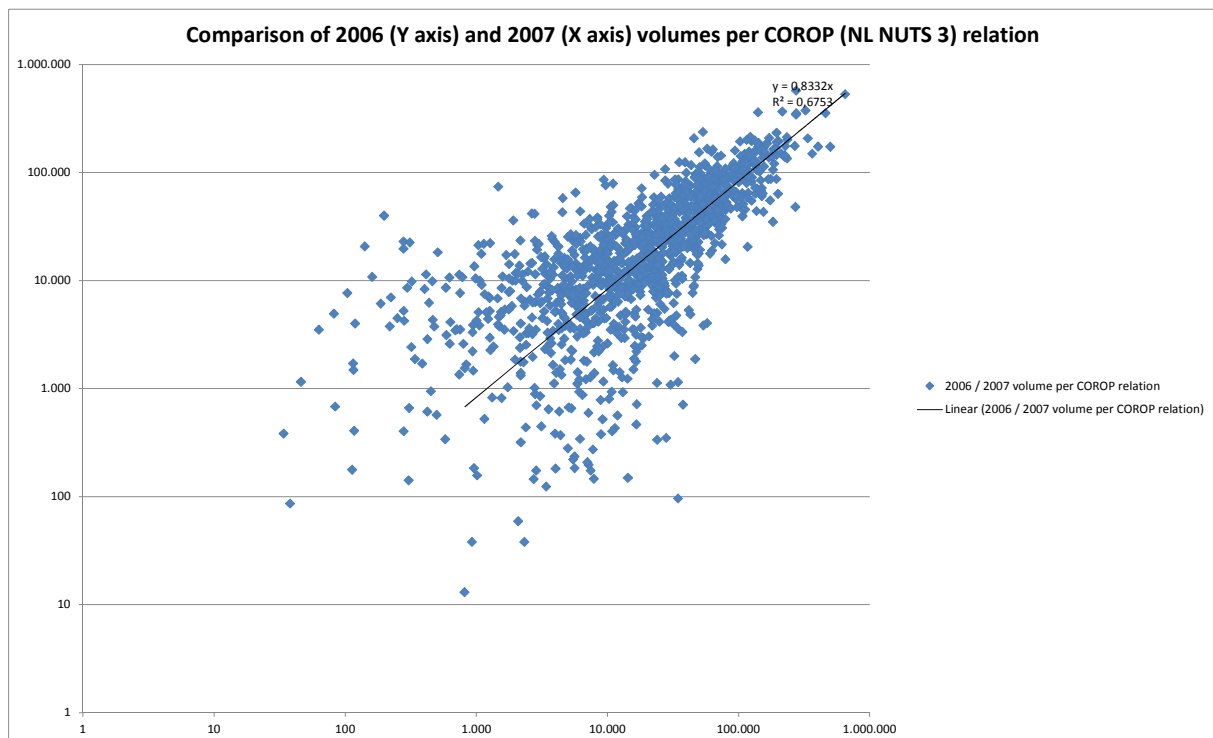


FIGURE 4.5. 2006 and 2007 annual volumes confirm selection in FIGURE 4.4

Gravity model application (attractiveness stability analysis)

The gravity model, formulated in section 3.5, has been applied to check the stability of location type variable. Transport flow data for 2007, 2008 and 2009 have been used for the experiments. From these data, 4 datasets have been made:

1. 2007 and 2008 combined (7-8)
2. 2007 and 2009 combined (7, 9)

3. 2008 and 2009 combined (8-9)
4. 2007, 2008 and 2009 combined (7-9)

The term “combined” is used in the sense of summation: road transport statistics have been summed up for several years, representing road goods flow observed in those years. For instance, if a relation between regions region NL111 and region NL112 for the commodity 92 showed 1000 ton in 2007 and 850 ton in 2008, then in the resulting dataset the flow between NL111 and NL112 for commodity 92 is 1850 ton.

The need to combine several years in one statistics dataset has been touched upon in the preceding section: for some relations the statistics become too thin, while it is not possible to draw a conclusion on whether there are some unobserved flows due to the sampling method or there is truly no flow. Combination of several year statistics in one dataset makes the relations thicker, however, at the expense of distinction of possible underlying trends.

The purpose of the gravity model experiments of this section is twofold. First is to look at how stable the distribution location variable is over the years of 2007-2009. Second is to check whether there are substantial underlying trends. In other words, if the variable is stable, then the quality of the statistics is sufficiently good; if the variable is not stable, then it could be due to a poor quality or the presence of a trend that shifts distribution activity pattern.

The model has been run for the 4 abovementioned datasets, for 2 types of flow: from all types of locations to the distribution (All2D) and from distribution to all types of locations (D2All). The estimated in gravity model production p_i and attraction q_j values (see formulations in section 3.5, formula 3.4) for these flows have been compared for the 4 datasets and 2 types of flows. The R^2 measures between pairs of the p_i and q_j vectors are within the range of 0,94-0,99. It shows a very stable behavior of the estimated transport volume production and attractiveness within the period 2007-2009 of the regions where distribution is located. This fact suggests that the distribution location type is of sufficient quality at least with respect to region-specific production and attraction of transport volumes.

4.2.7. Zero-volume relations

The issue of sparsely populated flow matrix is not directly related to the data quality, but may have some modeling implications. This may result in a problem when less inclusive flow criteria split the volumes thinner on the origin-destination relations leading to almost zero O/D matrices. Indeed, if one takes into account the possible size of the dataset, which includes 40x40 regions, 52 commodities and 9x9 combinations of location types (the total number of combinations is some 6,7 million), it is not difficult to observe that with the sample size of some 80 000 records per year in the survey, most of the relations would be empty, or zero-flow. The nature of the problem is fundamental because CBS cannot record all transport movements and scales the recorded movements up to the annual level. The underlying question is whether zero volumes on certain relations are a result of sampling error, or it is a real world phenomenon. Note that in this context the recent developments on electronic data accrual via XML data exchange might provide an alleviation of the issue.

To alleviate the problem of too thin statistics, combined datasets have been made for the period 2007-2009. An application of the gravity model shows a good performance of the distribution location type variable value for this period. This fact led us to the conclusion that the combined dataset can be used for the modeling; and that the quality of location type variable, especially the “distribution” instances of it, perform very stable for this period.

We expect the problem of zeros to be solved in the coming years by a technological shift in the statistical data collection methods. The data used in this thesis has been collected through manual input, using paper-based and web-based questionnaires. This technique makes it impractical to collect information about all vehicle movements and all shipments. However, new automatic data collection technics that are being currently deployed by Statistics Netherlands make it possible to automate this process and remove human labor completely, see for example Davydenko et al. (2014c). The automatic data collection allows obtaining all transport data: in this case a zero for a certain relation will mean indeed a zero flow.

The following section goes further in the analysis of the quality of the location type variable in the CBS road transport survey. The analysis of the relation between employment in certain economy sectors and throughput of the warehouse and distribution facilities was not intended for the purpose of quality check. It was meant to find an estimation method for distribution volumes in other (European) countries, for which CBS cannot provide data, but the needs for logistics models in no way is smaller there. The results of the exercise in matching employment and distribution throughput have been encouragingly good, as section 4.3 proves, thus providing further positive evidence on the quality of the location variable in the road transport statistics.

4.3. Link between employment and distribution volumes

This section borrows materials from the paper presented at the TRB annual meeting in 2012, (Davydenko 2012). Having analyzed the CBS road transport survey (section 4.2) in relation to the location type variable quality, the question arose on the applicability of these data and for distribution modeling for other (European) countries. To our knowledge, only Dutch CBS among the European statistics bureaus can provide data on location type variable or similar data, which distinguish flow according to its purpose. The location data is vital for the modeling, as explained in model description Chapter 3 of this thesis.

This section explores the idea of finding a proxy for the regional warehouse throughput data and distribution trip generation in other, publicly available datasets. Using the proxy data, it might be possible to overcome the absence of location type variable for other countries. A strong correlation has been observed between employment in certain economy sectors on the one hand, and freight flow via distribution centers and truck trips attracted to and generated by distribution facilities on the other hand. This observation provides evidence that the location type variable is of good quality for the logistics chain model and that the employment data can be used as a useful proxy to the regional distribution volumes in case if location type data is not available.

The data over employment comes from the two sources. The first data source is based on the Eurostat sectorial employment data (regionalized Structural Business Statistics (SBS)) at NUTS2 level for 89 economy sectors. At NUTS2 level, the Netherlands is divided into 12 regions, thus representing Dutch provinces. The second data source is CBS employment data at NUTS3 level for 37 economy sectors (CBS company classification based on SBS'93 classification). At NUTS3 level, the Netherlands is divided into 40 so-called COROP regions. NUTS3 level has sufficient spatial resolution to capture details at the urban level: per one NUTS3 region there are approximately 400 000 people, some 14 NUTS3 regions can be attributed to the Randstad urban area. This second dataset has more spatial detail, but contains a less detailed economy sector classification.

A strong correlation is observed between distribution facilities throughput measured as the number of ton going in and out from distribution facility per region, and the employment in certain sectors of the economy. This correlation suggests a linear relationship between employment and throughput. Let x_{ij} denote the number of Full Time Employees (FTE) in region i ($i = 1, \dots, n$) and the economy sector j ($j = 1, \dots, m$). Let us also denote y_{ik} to be ton throughput of distribution facilities (i.e. DCs freight generation volume) of the region i and commodity group k . The vectors X_j and Y_k have the same spatial dimension (index i denotes region's number). The DC throughput and DC freight generation can be estimated according to (4.1).

$$Y_k = \alpha_j + \beta_j X_j \quad (4.1)$$

First, we present the results of analysis of freight and freight trips generation by distribution centers in the Netherlands at the provincial NUTS2 level for 12 regions, for all palletizable commodities (i.e. commodities suitable for transport on pallets and storage at the warehouses, as shown in FIGURE 4.4) grouped together. TABLE 4.4 shows the determination coefficient R^2 between flows via distribution centers and employment in selected economy sectors.

TABLE 4.4. Correlation between DC throughput and employment in selected sectors

SBS Code	Sector	Description	R^2
G467		Other specialized wholesale	0.98
G466		Wholesale of other machinery, equipment and supplies	0.96
G469		Non-specialized wholesale trade	0.96
G475		Retail sale of other household equipment in specialized stores	0.96
G463		Wholesale of food, beverages and tobacco	0.95
G473		Retail sale of automotive fuel in specialized stores	0.93
H52		Warehousing and support activities for transportation	0.77
H50		Water transport	0.55
H51		Air transport	0.05

The choice of the economy sector as a proxy for distribution throughput is of great importance. For instance, correlation between DC throughput and employment in the sector H52 (warehousing) is weaker than for the sector G467 (other specialized wholesale). R^2 for

the warehousing sector is 0.77, which is comparable to the results of Orsini et al (2009), where this estimate for French warehouses was 0.69. *Table 4.4 shows which economy sectors are linked to the distribution freight generation: it is not employment at warehouses, but employment in the sectors that the warehousing sector serves, such as wholesale.*

TABLE 4.5 shows the Dutch DC throughput and DC trip generation for palletizable commodities per FTE in sector G467 (Other specialized wholesale). The values are the ratio of regional DC outbound ton volumes and trips over regional employment in sector G467. As TABLE 4.5 shows, freight and trip generation per FTE do not vary much between the Dutch NUTS2 regions. The measure of variability, defined as the ratio between standard deviation and the average, is in the interval of 0.12-0.14 for both freight and trip generation. Such small variability allows using the average value as the basis for estimation of freight and trip generation according to the formula 4.1.

TABLE 4.5. Freight and trip generation by DCs for the palletizable commodities, annual ton and tips per Full Time Employee (FTE) in G467 sector

NUTS2 Region		Freight per FTE, ton	DC outbound HGV trips per FTE
NL11		667	65.7
NL12		658	80.3
NL13		758	106.1
NL21		608	78.4
NL22		571	81.2
NL23		787	98.2
NL31		540	77.7
NL32		616	82.9
NL33		690	86.1
NL34		858	95.1
NL41		607	85.7
NL42		718	84.7
	Average:	673	85.0
	Stdev:	94	11.0
	Stdev/av:	0.140	0.124

The stability of DC throughput per FTE (TABLE 4.5) suggests that distribution throughput can be estimated purely on the basis of employment data. As the information over loading and offloading location types is generally not available from public statistics, the strong link between throughput and employment shown for the Dutch case will probably hold true in other structurally comparable economies such as Germany, Belgium, and others. Nonetheless, it is worth suggesting that the average throughput per FTE should be corrected for other economies, adjusting it, for instance, by the total productivity factor.

The results shown in TABLE 4.5 encourage freight flow and trip generation estimation to be made in a linear form, namely in the form of a multiplier to the number of employees. For instance, each employee in the sector G467 generates 673 ton of DC throughput and 85 truck trips annually. Using this very simple approach (conform formulae 4.1, where the intercept a

is set to 0), it is possible to make accurate estimations of the freight flow and trip generation. FIGURE 4.6 and FIGURE 4.7 show the estimated and actual freight and trip generation for the palletizable commodities and employment in sector G467.

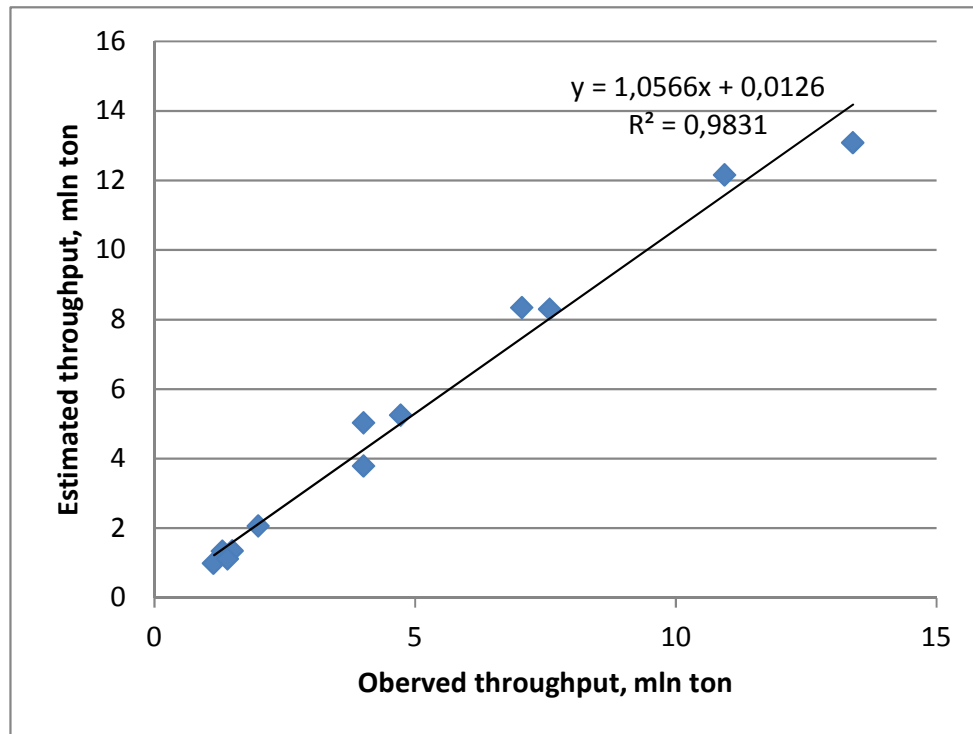


FIGURE 4.6. Observed and estimated DC throughput for palletizable commodities, 12 Dutch regions, sector G467

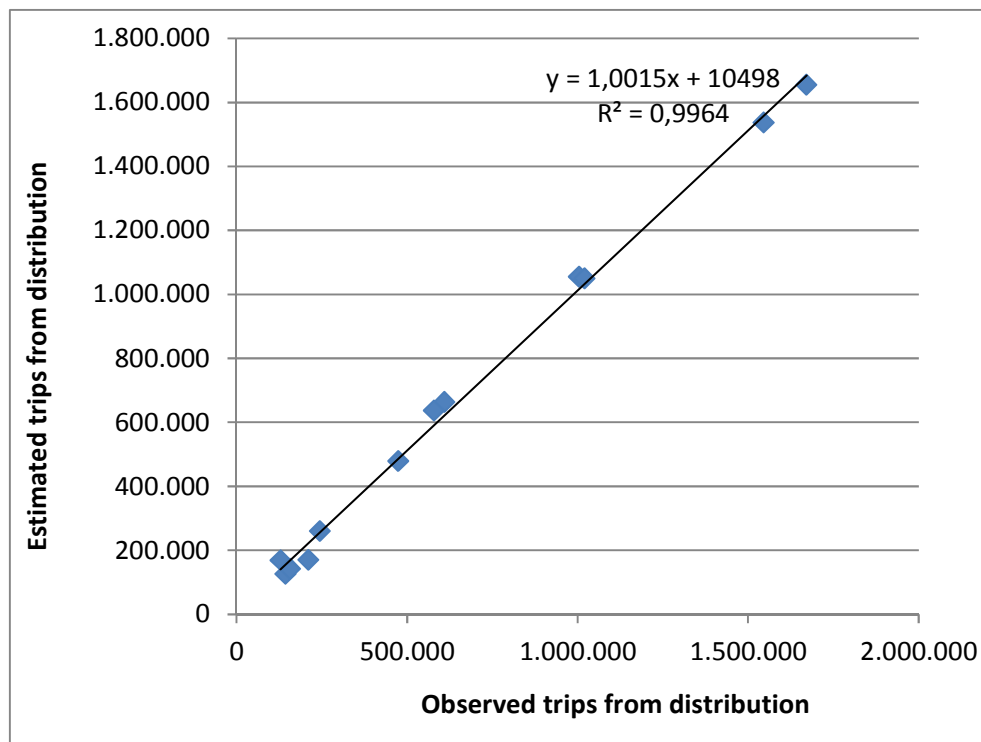


FIGURE 4.7. Observed and estimated DC trip generation for palletizable commodities, 12 Dutch regions, G467

Figures 4.6 and 4.7 show that DC freight and trip generation can be estimated in the form (4.2), omitting the intercept value due to its insignificantly small value.

$$Y = \beta X \quad (4.2)$$

where Y is the estimated DC freight or trip generation, X is the observed employment and β is throughput or number of trips per employee in the chosen economy sector. A more general approach proposed by Holguin-Veras et al (2011) includes an intercept α

$$Y = \alpha + \beta X \quad (4.3),$$

however FIGURES 4.6 and 4.7 show that the α component is very small in comparison to the scale, thus for practical reasons α can be set to zero and omitted. Moreover, the β coefficient can be seen as a proxy to the productivity factor, in this case expressed in FTE productivity with respect to freight volume and trips generation. The main difference to the analysis made by Holguin-Veras et al (2011) is that at the regional level the effects of the economic optimal order quantity, which play a role in companies' decisions on order frequency and size, are hidden by the fact that outgoing vehicles from DCs carry a number of orders and make a substantial number of stops along the route. Our evidence strongly suggests that *if the right economy sector is chosen as a proxy to freight and trip generation by DCs, the freight and trip generation are proportional to employment in that economy sector*. We suggest that the "right" economy sector is the sector which has the highest correlation between employment in it and DC freight and / or trip generation. Note that the data do not show the causal relationship (is it employment that generates flows or is it flows that make it possible for people to be employed?), the causality cannot be determined from this analysis. The following section confirms these criteria on the basis of multivariate analysis.

4.3.1. Disaggregation at commodity level and smaller regions

The disaggregation of the commodity flow per commodity group also shows a high correlation between employment data and distribution freight and trip generation. However, depending on commodity type, different economy sectors perform as the best proxies for the throughput and trip generation. The analysis of 89 SBS sectors and commodities at NSTR-1 level identifies the economy sectors that can best explain distribution throughput and trip generation. TABLE 4.6 shows sectors that explain in the best way distribution freight and truck trip generation. We show R^2 measure for these parameters estimated on the basis of employment data and compared to the observed values.

TABLE 4.6. Estimation of DC throughput and truck trips generation on the basis of sector employment

SBS Economy Sector	NSTR-1 Commodity Group	Throughput estimated through employment, R ²	Truck Trips estimated through employment, R ²
C10: Manufacture of food products	1: Food and animal feed	0.91	0.94
B: Mining of coal and lignite	2: Solid mineral fuels	0.51	0.47
C33: Repair and installation of machinery and equipment	3: Petroleum products	0.69	0.72
I55: Accommodation	4: Ores, metal	0.72	0.73
C33: Repair and installation of machinery and equipment	5: Iron, steel and nonferrous metals	0.91	0.94
G467: Other specialized wholesale	9: Vehicles, machinery and other goods	0.96	0.99

A detailed consideration of the results presented in TABLE 4.6 leads to a number of observations. First, commodities that are often transported as bulk goods such as 2: Solid mineral fuels, 3: Petroleum products, 4: Ores, metals are not very suitable for distribution. Relatively small R² values confirm this conclusion. Second, commodities 1: Food and animal feed, 9: Vehicles, machinery and other goods show high R² values and are logically linked to the explaining sectors. Nonetheless, there still an uncertainty remains that there could be other factors that may influence distribution freight and trip generation.

The high correlation between distribution centers' throughput and trip generation on the one hand, and sector employment data on the other hand, might not fully explain the statistical link between these two variables. Potentially, there could be other variables that explain distribution processes. A combination of economy sectors might give some better results. To clarify this question, a multivariate analysis has been performed using another, more spatially detailed employment statistics dataset.

The detailed employment statistics in SBS classification (see Eurostat (2008) for the SBS classification) are only available at the NUTS2 spatial resolution level, which consists of 12 Dutch regions. The NUTS2 spatial level is not very suitable for a more detailed analysis as the regions are large and many economic activities are lumped together. Therefore, for the detailed analysis the employment statistics in CBS company classification format have been used. These data are available for the NUTS3 spatial level, comprising of 40 Dutch regions. This dataset is different from the SBS dataset, it has a smaller number of economy sectors: 37 economy sectors in comparison to 89 sectors present in SBS data. In other words, CBS company classification employment dataset provides more spatial detail and fewer details over the economy sectors.

To check whether population per region has a substantial impact on DC throughput and trip generation, an extra fictitious sector has been added to the employment dataset: the population size per Dutch NUTS3 region. The multivariate analysis has been performed for the NSTR-2 commodity 12: Drinks. First, the method described above has been applied to find how well various economy sectors can explain distribution of drinks. The distribution process is best explained by the Sector 3: Food and stimulants, with R² value of 0.46. DC throughput of the

Drinks commodity estimated purely on the basis of regional population has an R^2 value of 0.26.

Subsequently 14 economy sectors with the highest R^2 values of estimated throughput have been selected for the multivariate analysis. The results of the multivariate analysis are shown in TABLE 4.7. The R^2 value of the estimated DC throughput based on the results of the multivariate analysis is 0.60, which is larger than in case when only one sector used (0.46). However, the quality of the results are poor: T- and P- statistics values do not provide confidence in the results, as shown in TABLE 4.7.

TABLE 4.7. Results of multivariate analysis for 14 sectors

<i>Sector</i>	<i>T- statistics</i>	<i>P-value</i>
3, Food and stimulants	1.08	0.29
8, Chemical industry	-0.14	0.89
11, Metal products industry	-0.08	0.93
12, Machine industry	0.06	0.95
14, Production of transport means	0.53	0.60
15, Other industry	1.60	0.12
17, Construction industry	-0.35	0.73
18, Soil, water and road construction	-0.48	0.64
19, Other construction	0.79	0.44
20, Wholesale	1.76	0.09
21, Retail automotive and reparation	0.27	0.79
22, Catering	-0.30	0.76
23, Land Transport	-1.24	0.23
38, Total population	-0.72	0.48

The confidence in the results can be improved by lowering of the number of economy sectors. TABLE 4.8 presents the results of a multivariate analysis of 4 sectors, resulting in R^2 value of 0.50 for the estimated throughput. The T- and P- statistics values are in the good range for the sector 3 only (it is the best sector to estimate distribution throughput for the Drinks commodity). Other sectors in the analysis are not within the confidence interval (P- values between 0.01 and 0.05 and T- values of less than -2 or greater than 2).

TABLE 4.8. Results of multivariate analysis for 4 sectors

<i>Sector</i>	<i>T- statistics</i>	<i>P-value</i>
3, Food and stimulants	2.2231	0.0326
7, Petroleum and coal industry	-1.4339	0.1602
15, Other industry	1.5310	0.1345
38, Total population	0.9701	0.3385

The multivariate analysis performed for the estimation of regional distribution throughput for the drinks commodity does not show improvements in estimation of the DC throughput. Although R^2 of the estimated throughput values is higher, the estimation itself does come with a statistical confidence: only the best explaining sector has a significant statistical coefficient.

The evidence points in the direction of the single economy sector usage for the estimation of DC throughput and truck trip generation.

The estimation of freight and trip generation by warehouses and distribution centers has some practical applications. For instance, if detailed information on vehicle movements and flow of goods is needed, such as at the district or street level, transport statistics would not help due to insufficient spatial resolution. In this case, employment data can be used as the input for the estimations of freight trips at the very detailed spatial level under condition of detailed employment data availability. Up to now, there were no data available linking employment and distribution-related transport activities. Holguin-Veras et al. (2012) and CROW (2007) link freight trips and employment for other economy sectors than distribution and warehousing. Freight trips and freight flow estimations of this chapter present an extension to this knowledge with the data on distribution.

4.3.2. Conclusions on sectorial employment and distribution volumes

Transport models that take distribution processes and logistics structures into account need explicit data on distribution-related transport volumes and trips. These data might be used for the purpose of model calibration, as well as an input for gravity models, which estimate deterrence functions and attractiveness of regions for the distribution activities. In practice, transport data containing information over the purpose of transport activity is generally unavailable. Employment data can be used for the estimation of distribution facilities freight and trip generation. DCs' throughput and trip generation is strongly correlated with employment in some economy sectors. It has been shown that DC ton freight generation and DC trip generation are proportional to the employment in the best-correlated sector, the intercept value can be set to zero.

The choice of the economy sector is very important. For instance, it has been shown that for distribution volumes and distribution-related trips the wholesale sector provides much better results than the employment in the warehousing sector. Therefore, the best proxy for the distribution freight and trip generation is the employment in the economy sector, which shows the highest correlation with distribution volumes and trips. This choice criterion for the economy sector holds true at the disaggregated commodity level as well. The multivariate analysis shows that inclusion of a number of economy sectors as explaining variables improves estimations only marginally, while making the results statistically unreliable. Moreover, the data on total population per region does not explain distribution activities: the strongest correlated economy sector provides much better estimations.

Follow up research may be conducted in the two directions. First, DC throughput and number of truck trips per employee might be different in other countries, as the productivity levels vary between the countries. Therefore, a way needs to be found to correct throughput volumes by, for instance, total labor productivity factor, or industry-specific productivity factor. Second, there are indications that regional population might have an impact on distribution activity through labor availability (i.e. people can be employed at DCs) as well as through consumption. However, distribution, consumption and dwelling can be spatially separate, i.e. located in different NUTS3 regions. The research on population proximity factor, which

might be used instead of regional population values, might provide some better results and used in multivariate estimations of distribution throughput and trip generation.

Noting that the distribution throughput volumes have been obtained from the CBS road transport survey by summing up volumes coming in and going out from the distribution location type, it is important to conclude that the location type variable performs very well, at least when the variable takes the “distribution” value. The large and logical correlations observed between employment and distribution suggests that the distribution-related flows are sufficiently properly registered in the road transport survey and, thus, can be reliably used for the purpose of logistics chain modeling.

4.4. German Data

The Logistics Chain Model (LCM) has also been applied and tested on the German data. The data has been kindly provided by Dr. Friedrich (Friedrich, 2010). The German data is related to the trade and commodity flows in the food retail sector, i.e. the products that are sold via food stores and supermarkets.

The German dataset used in the LCM contains essentially 4 types of flow, namely P/C trade flow and three transport sub-flows, PC, PD and DC, which are necessary for the LCM application and calibration. The dataset is at the NUTS2 spatial resolution level for Germany, thus distinguishing 41 regions within the country. Commodity split-up has not been used for the LCM applications: all commodity-specific flows have been treated as the flow of one commodity.

Friedrich 2010 presents a full account on how the German data is obtained and processed to the form necessary for the LCM application. This section presents only the most relevant data accrual and transformation procedures that lead to the data for the LCM model. FIGURE 4.8. conceptualizes the data accrual and transformation procedures for the data on German Food Retail sector.

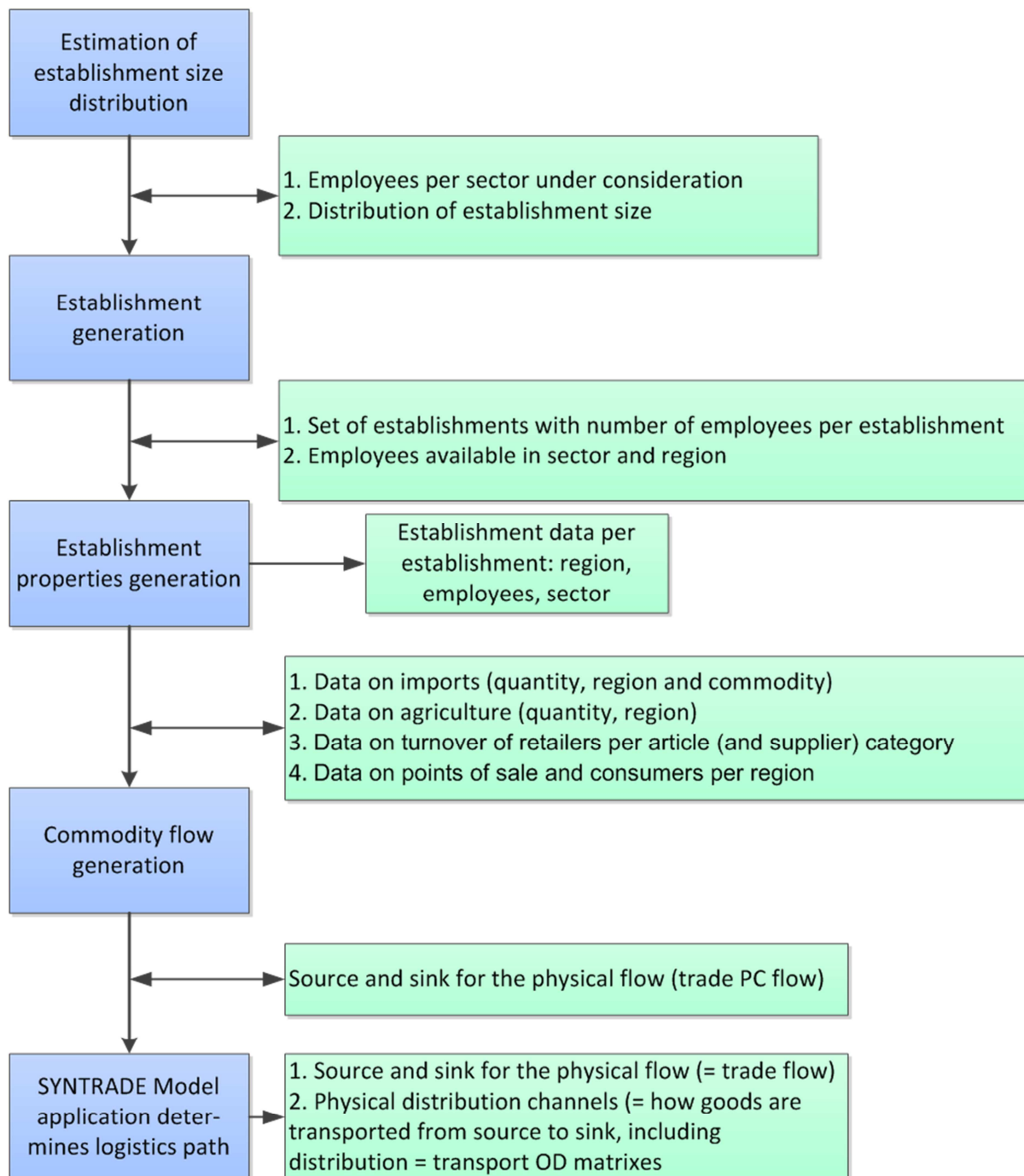


FIGURE 4.8. Data transformation procedures for German Food Retail sector

The procedure starts with the creation of an artificial economic landscape and estimation of the relevant economic activity. The scope has been limited to the activities leading to production and distribution of goods destined for direct consumption or to the sectors that serve consumption, such as the hospitality sector. Two primary data sources have been used for this procedure: inland production and international trade statistics, provided by the German statistics bureau and related to the 2006 data (Statistisches Bundesamt, 2006a) and international trade statistics (Statistisches Bundesamt, 200b).

Further, the distribution of the number of employee per establishment (establishment size) is estimated. The primary data source for this procedure is the number of establishments per establishment size, published by the German statistics bureau at the CPA 4-digit level. The essence of the procedure is the estimation of the parameters of statistical distribution such that the difference between the statistics data and estimated procedure is at the minimum. Friedrich (2010) reports that the average delta for the estimated establishment size distribution is 3% and about 1,5% for the number of employees.

The German employment agency provided the data on the number employees per sector and per region. Combined with the data on establishment size distribution, the result is the dataset on establishments per region per sector with the information on the number of employees.

In addition to the establishment and employment data, data on imports, agricultural production, Logistics Service Providers (LSP) and wholesalers are necessary to establish commodity flows. Using disaggregate data, the model determines turnover per supplier for the retail establishment, thus creating commodity flow tables (P/C trade flow). Subsequently, the SYNTRADE model has been applied to determine the logistics path of the good between the sourcing (production or import) and the Point of Sale (POS) locations, thus making it possible to construct the transport flows. The SYNTRADE model determines the locations of the distribution centers as the nodes linking transport legs.

Thus all the necessary data components have been created for the application of the LCM. The step of SYNTRADE application is to some extent unnecessary, as the LCM model performs the same task of determining logistics path of the goods between the sourcing locations and points of sale. However, the extra data on transport flows allow application of an important LCM functionality, namely, calibration of the model on the known transport flows. The SYNTRADE model output allowed determining of the LCM parameters, such as price sensitivity and regional attractiveness for the distribution, making it possible to decouple the SYNTRADE model from the future LCM applications for the German data.

4.5. European Data

This section briefly describes European level trade and transport flow databases used for the Logistics Chain Model (LCM) estimation at the European level. The data have been provided by the Netherlands Environmental Assessment Agency PBL making it possible to apply the model at the EU level. The data and the method used to obtain it are described in detail in PBL publications, see Thissen 2013a and 2013b. The work with the data has also resulted in a collaborative effort that tested possibilities of LCM application using the European data, see Davydenko 2014a. The model application and estimation is discussed in section 5.3 on Implementation and estimation of the Logistics Chain Model for the Continental Part of the European Union.

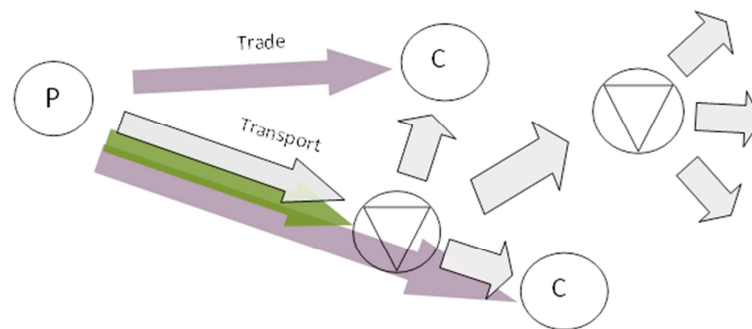
The European Data consists of two full datasets on the industrial products and agricultural products commodities. The datasets contain an interregional trade flow matrix, and corresponding transport flow matrices, which show how the goods are shipped between the

two trading regions. The transport data include direct shipments, where transport origin and destination are the same as the trading regions, as well as flows via distribution regions.

The detailed data construction procedures are described in the PBL reports (Thissen 2013a and 2013b), here we concentrate on the most important aspects of the data construction procedures. It has been observed that the ton-volume of trade is generally smaller than ton-volume of transport, therefore, this disparity should be explained by activities, generally called as transshipments, which may include various types of logistics activities. These activities lead to logistics procedures that require taking cargo from one vehicle and loading it into another one without consuming it. Each vehicle movement between loading and unloading operations is registered in transport statistics. If loading-unloading pair of operations occur on the way from production to consumption, then transport operations would be registered at least twice (and ton-volume transported would be at least two times larger than the trade flow) on the way between production and consumption. Note that this does not distinguish between storage and cross-docking related transshipments.

Construction of trade flow datasets generally faces the problem that trade flows are not registered directly. In Europe, regional Input-Output data and country-level trade accounts are available, but these do not lead directly to the data on interregional trade flows, which show origin of trade (the place where the goods are produced and sold) and the destination of trade (the place where goods are bought and consumed).

The data construction procedure harmonizes trade and transport flows in a way that does not use any resistance or cost based modeling. The European interregional transport flows at the NUTS2 level and country-level trade statistics used as the input. The transport data is used to regionalize the trade flows, as the country-level data presents information over trade between countries, and no information on what regions are trading with each other, see FIGURE 4.9 that conceptualized the procedure that is used to harmonize trade and transport data.



Excess transport flow is attributed to distribution. Distribution-related flows are assigned proportionally to transport flows (i.e. using transport flow patterns)

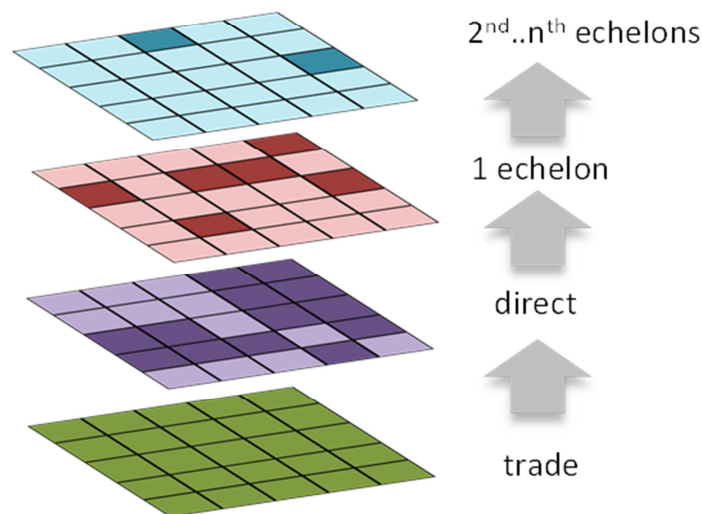


FIGURE 4.9. Conceptualization of trade-transport data harmonization procedure

The trade flow dataset has been constructed using transport flow data. In order to arrive at an unpolluted trade table, except for the handling factor parameter, a parameter free model was used to align transport and trade statistics, resulting in a consistent dataset for trade and transport chains. The regionalization of the trade follows transport flows. As the sums of traded volumes and transported volumes at the country level are almost never equal, the excess of transport is attributed to logistics activities. Hence the procedure assigns extra transport volumes to up to 5-echelon logistics structures, following the direction of transport movements. In (rare) case if trade volume is larger than transport volume, the transport volume is corrected upwards to harmonize trade and transport volumes.

This work has resulted in two datasets: a trade flow matrix (PC flow) and transport flow dataset, which shows how trade is realized by transport in the form of chains (P-C, P-D-C, P-D-D-C, P-D-...-D-C flows of up to 5 intermediary stops at logistics (distribution) facilities). These datasets have been produced for the two distinct commodity groups: agricultural products and industrial products. The data present both transport and trade flows at the NUTS2 level for the continental part of the European Union. These NUTS2 flows distinguish between 256 regions at this level of detail. It should also be noted that trade and transport flows are related to road transport only; trade realized by other transport modalities is not included into the data. Thus, the trade and transport flow datasets are consistent.

4.6. Discussion and Conclusions

This chapter presented and discussed data on which the modeling efforts are based. For the research and modeling design presented in this thesis, three datasets have been obtained and used. The first dataset concerns Dutch road transport statistics extended with the location type variable; the second dataset concerns trade and transport flows for the continental part of the European Union; and the third dataset concerns trade and transport flows of the German food retail sector. These data allowed modeling applications at different spatial and economy resolution levels, thus contributing to the assessment of robustness of the modeling design.

The Dutch road transport survey data have been discussed in detail. The level of detail can be partly explained by the fact that the Dutch dataset has been the first obtained dataset, which influenced the design of the LCM. Due to the fact that Statistics Netherlands does not control for the quality of the location type variable in their statistics, a substantial effort has been devoted to the assessment of the data quality.

To assess its quality, essentially three methods have been used. The first method compared the stability of the data over a series of years. The underlying idea is that transport volumes do not change drastically year-on-year. The comparison has been performed by simply comparing the volumes that pass through certain location types, such as distribution. Another technique used to assess the quality of data was an application of the SCGM to determine regional generation and attraction and to compare the change in these on a year-on-year basis.

The outcome of the first two analysis methods suggested sufficient for our modeling purposes quality of the location type variable. However, without validation of the data by an unrelated dataset, the quality of transport data could be still questioned. For this purpose, the third analysis method has been used: a link between transport flow data and the sectorial employment statistics has been established. A simple regression model has shown that the regional distribution throughput (i.e. the goods volumes that pass distribution location during a year) is indeed strongly linked to the number of people that work at these facilities. As the employment data is directly observed, a strong linear relation between throughput volume and employment figures validates the transport-related data.

The work that links employment data to the throughput volume of regional distribution facilities has positive consequences for the logistics modeling in other countries for which transport statistics extended with location type variable are unavailable. Distribution systems of other countries at this level of detail are not the subject of research presented in this thesis, however, the findings may be useful to other researchers who might be willing to develop or apply modeling techniques described here. Using employment statistics only, it is deemed possible to make estimations of distribution volumes in other countries, especially for those countries where the level of technological sophistication and labor productivity is close to the Dutch ones.

The European Dataset has made it possible to scale up the LCM model to a much larger spatial region, as well as to show that the model can be successfully calibrated on the data

containing a substantially bigger number of regions, 256 regions for the EU dataset, compared to 40 regions in the Netherlands and 41 in the German data. The novel work of Dr. Thissen and his colleagues from the Dutch PBL on parameter-less reconciliation of trade and transport statistics allowed construction of two consistent trade and transport flow datasets for the agricultural products and industrial products commodities.

The German data concerns only one economy sector, namely food retail sector in Germany allowed modeling application at a substantially more detailed level than it is the case for the Netherlands and EU. The German data is based on the study of establishments and employment as the basis for the sourcing decisions, thus making it possible to construct commodity flow (i.e. trade flow in the terminology of this thesis) for this sector for the whole Germany. An application of the SYNTRADE model has generated transport flow matrices, thus making it possible (as we will see in Chapter 5) to calibrate the LCM model on these data.

The extension of the basic Dutch data with the German and EU data allowed showing performance of the LCM at different levels of spatial resolution and coverage, and at different levels of economic activity aggregation. Still, the Dutch data are the most important in this research project as it is directly observed data: it is based on the comprehensive and sophisticated annual surveys that Statistics Netherlands conducts.

5. Model Implementation and Estimation

This chapter builds on the model formulation of Chapter 3 and Chapter 4 on data. It provides details on the model implementation and estimation for the four modeling cases, including estimation of the model parameters in the calibration runs and analysis of the model output in the form of generated transport flows.

This chapter is split into five sections: first is the application of the gravity model on the Dutch data; second is the implementation and application of the full logistics chain model for the Netherlands in two forms: main, based on transport costs, and alternative, based on shipment sizes. The third is the model application at the level of continental part of the European Union. The fourth modeling application is based on the German data for the food retail sector; and the fifth section draws conclusions on the model applications. Each application related section provides an autonomous account of a model application and is closed with conclusions on the application case. The chapter finishes with the conclusions that consider all four model application cases.

Substantial attention in all four modeling cases is given to the empirical validity of the modeling efforts. For the Dutch modeling applications, the model output is compared to the empirical survey data. For the European and German cases, the model output is compared to the available transport data, though not directly observed, but used as the “control” transport flow data set in the same way as the observed data is used in the Dutch cases.

5.1. Separately Calibrated Gravity Model

This section describes the implementation of and experimentations with the Separately Calibrated Gravity Model (SCGM). The mathematical formulation of the model was described in detail in the section 3.5 on the Gravity Model. In this sub-section we first discuss the program implementation and then provide results on the model application to the road transport survey data.

5.1.1. Program implementation

The model has been programmed as a standalone Windows 32 application in Embarcadero RAD Studio, a lineage of the Pascal language extended with objects and visual development environment, developed by Borland Software Corporation, currently known as Embarcadero. The Windows 32 model application provides a simple user interface and parameter input (see screenshot in FIGURE 5.1). The model software functionality can be summarized by the following elements:

1. Data loading and storage operations
2. Data filters
3. Graphical User Interface (GUI) for user-selected parameters
4. Mathematical model and execution interface

1. The model reads and stores data in tab-separated text files. Input data consists of the dataset provided by CBS on road transport survey (see section 4.2 for more details). The model also saves output data in tab-separated format, which is readily accessible by, for instance, office software such as Microsoft Excel. Output data contains parameters estimated by the model, observed and estimated flow matrices, and other data types for further analysis.

2. Data filters allow selection of a sub-set of the road transport survey data to which the model is applied. These filters are directly shown to the user through the GUI. They allow a) selection of a commodity or a number of commodities b) type of loading location type c) type of unloading location type. Data filters allow the model to work only with those records from the survey dataset that are selected by the user. For instance, if only the production loading type is selected by the user, the model will take into account only records, where goods are loaded at the production location type.

3. The GUI allows a) selection of individual commodities or the usage of pre-set commodity groups such as palletizable (i.e. suitable for distribution) commodities b) selection of loading – unloading location type pairs c) specification of output files d) selection of Generalized Logistics Cost (GLC) parameters, as specified in section 3.3.3, providing a possibility to use pre-set values, and to use simple and extended GLC formulations. The interface also provides action buttons, allowing the user to perform pre-programmed tasks.

4. Before showing up the main model screen, the model loads the CBS road transport survey dataset. Once the user selects filtering options and specifies the GLC values, the software is ready to estimate the gravity model. The software executes the model according to described in section 3.5.2 algorithm: each model execution searches for the model sensitivity parameter β . The β parameter depends on the flow type selection (commodity type and loading – unloading location type combination). Once the model completed computations, the output data is saved in a tab-separated file, specified by the user.

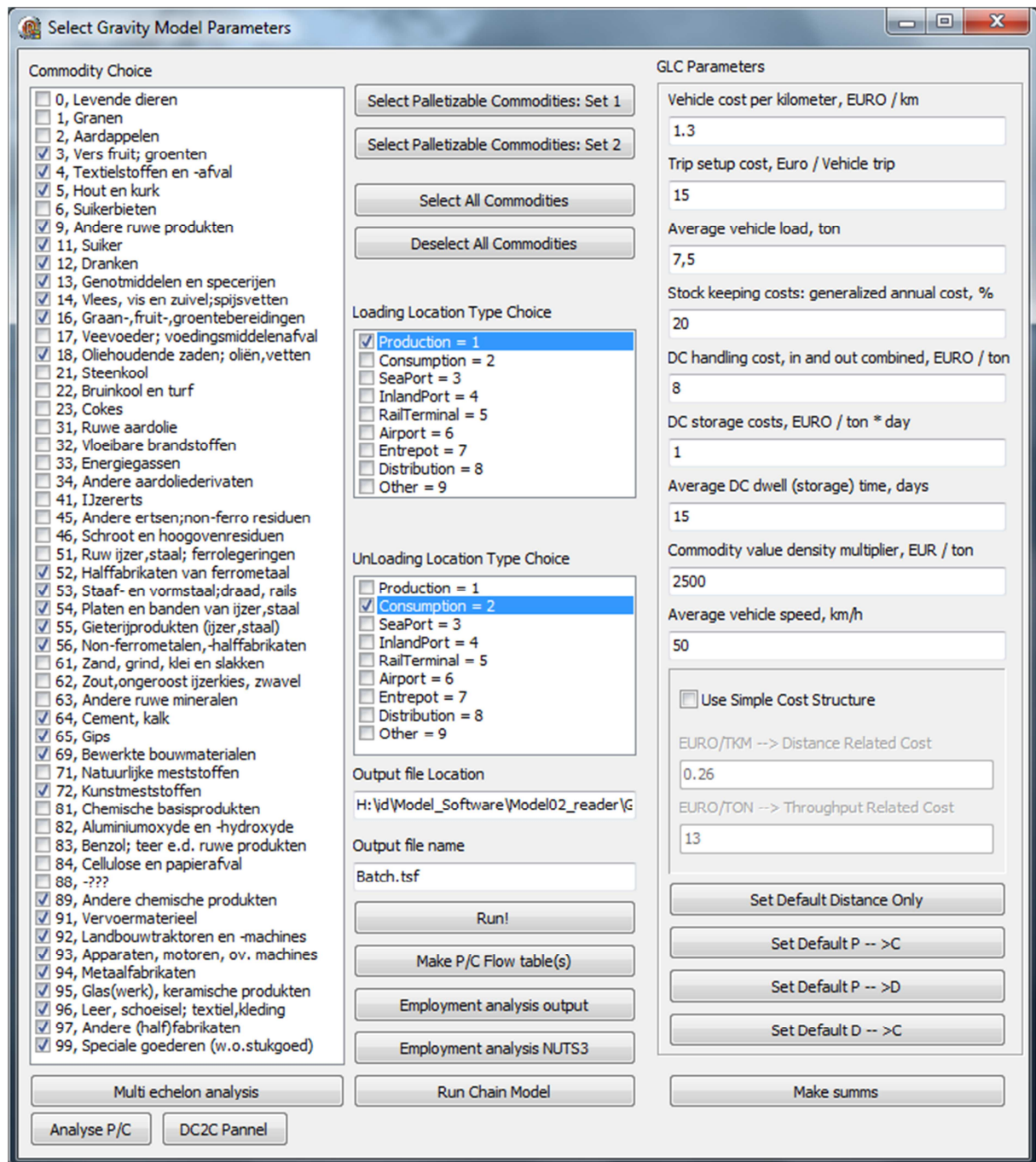


FIGURE 5.1. Screenshot of the SCGM model main screen.

The model output in a text tab-separated format contains data on

- optimum parameter β value
- best (minimum) error value of found for formulas (3.14 & 3.15),
- determined values of p_i and q_j for the optimum parameter β value, see formula 3.4
- computed interregional costs matrix
- exponent values for each i, j combination, formulae 3.4
- estimated (computed) and observed (input data) interregional flow
- values of formula 3.15 for each intermediate parameter β value
- estimated and observed flow values in the vector format for the fit (regression) analysis

5.1.2. SCGM model runs: looking for the properties of different types of flow

The Separately Calibrated gravity Model has been employed for analysis of different combinations of OD transport flows. The model has been applied for various loading / unloading location type combinations and commodity types. The model shows a stable, accurate and fast execution conform minimization (3.14), and subject to feasibility of the solution (3.13).

The main purpose of the SCGM application on the empirical Dutch transport data is to look at the differences between different types of flow. Model application results help pointing out that, for instance, Production-Distribution (PD) flow is different from the Production-Consumption (PC) flow: the differences can be observed in terms of distance decay values and quality of the fit. The results further support the need to distinguish these flows in transport modeling efforts.

Although the main purpose of the gravity model application in this research project has been to obtain P/C trade flows, which are the primary input for the logistics chain model, the gravity model can provide a number of useful insights into the patterns of transportation if it is applied to O/D flows. We consider two main questions that can be dealt with using the SCGM model.

The SCGM model provides an answer on the boundaries of transport flow estimation quality. At the very basic level, each region has a given outflow and a given inflow of goods by physical transport. The gravity model estimates where the regional outflows end up, and conversely, from which regions the inflow is coming. To do this, the described implementation of the gravity model uses a cost function to estimate these flows. When the output of the model is compared to the flows observed in reality, a conclusion can be drawn on how well a certain function can represent reality. In other words, how well the used transport costs function explains the flows. The answer for this question is not only interesting in itself, it also suggests whether modeling techniques that use costs as the decision factor would perform well.

More advanced modeling techniques, such as the logistics chain model described in this thesis, use the same underlying modeling principles, however, in a more sophisticated set up. The logistics chain model estimates three types of flow based on essentially two sets of choices, (1) direct shipment versus indirect shipment, and (2) the choice for the warehousing region in case of indirect shipment. Such a model estimates different types of flows at the same time. The discussed implementation of the gravity model does not test various types of flow simultaneously. Only the spatial distribution of single transport flows is tested, based on the assumptions concerning transport and logistics costs. Thus, the Separately Calibrated Gravity Model (SCGM) should provide a test on how well a logit function with transport cost based disutility estimates the transport flows. Practical cost definitions are given by formulae 3.16 and 3.17, where transport costs are proportional to the distance between the regions, hence the ton-kilometer transport costs are also linearly proportional to the distance. The gravity model tests whether these cost formulations adequately represent reality, in other

words, test the hypothesis that this form of the costs defines resistance to the flow. Using the specified cost functions, the model generates a flow that is compared to a flow observed in reality; the comparison is in the form of a fit between the observed and estimated flows. If the fit is considered to be good, then the model and the discussed cost functions can be used for the modeling of the flow.

5.1.3. Estimation quality for the various flow types

This section looks at several relevant types of flow and draws conclusions about model estimation quality. We consider various flows represented in the source CBS transport flow survey data. The most interesting flows are the following (consult FIGURE 3.4 on a graphical example of P/C flow conversion into transport O/D flow):

1. Total flow between O's and D's, which is the sum of flows in points 2-4 of this list
2. Production-Consumption flow (direct)
3. Production-Distribution flow
4. Distribution-Consumption flow

Each of these four flows can be split into sub flows, taking into account the following reasons. First, the flows can be split into various commodity types. As it can be seen in FIGURE 5.1, there are 52 types of good distinguished (NSTR/2 commodity classification). Second, different location types can be treated as production and consumption locations. For instance, let us consider production location. The survey data has an explicit location type "production": it is assumed that the respondents use this type in case if a transport operation originated or finished at a location where production activities take place. The imports can also be treated as production: goods can be thought of "to be created or made" at the entry points such as ports and terminals. In this situation, location types such as sea port, rail terminal, airport, inland port and entrepot are considered as production locations. Furthermore, in certain instances the consumption location type can also be considered as production location due to the fact that consumption location type also generates some outflows. The latter could be due to returns, garbage, unused products and other types of flows / activities. The gravity model experimentation includes explicit cases when production and consumption locations are considered to be strictly defined, and cases when production also includes import. Section 5.1.6 defines the flow cases.

Looking at the consumption end of the flows, we use the explicit location type "consumption", assuming that people participating in CBS surveys meant deliveries to the shops, retail points, restaurants, etc. In case of consumption points, contrary to the production location type, points of aggregated export, such as sea ports and rail terminals, should not be considered as the points of consumption. This is mainly due to the fact that consumption generates smaller transport batches, while export movements involve aggregated shipments. We can generally treat the production location type on the receiving end of flows as a consumption location because goods are used (consumed) there for further rework. This logic will be applied in considering different types of flow in experiments with the gravity model as well as in case of the logistics chain model.

Among others, the following visual technique of comparing flow values observed and estimated by the model is used throughout the thesis. In a graph on X axis the observed flow values are plotted and on Y axis the estimated flows are plotted such that each cell of the corresponding flow matrix is represented by a dot with (x, y) coordinates. For instance, an interregional flow in the Netherlands is at NUTS3 (COROP) level, which is the most used spatial resolution in this thesis. There are 1600 (40x40) dots on the graph each representing an observed and an estimated flow value. The graphs are rendered in Microsoft Excel, which computes the quality of the fit in the form of the coefficient of determination R^2 .

5.1.4. All flows put together

This model run estimates gravity model on a combined flow, i.e. “all flows put together”. The flow origin side includes all 9 location types, the flow destination also includes all 9 location types and all 52 commodity types are included. In other words, this model run estimates all road transport flows. FIGURE 5.2 shows the estimation quality.

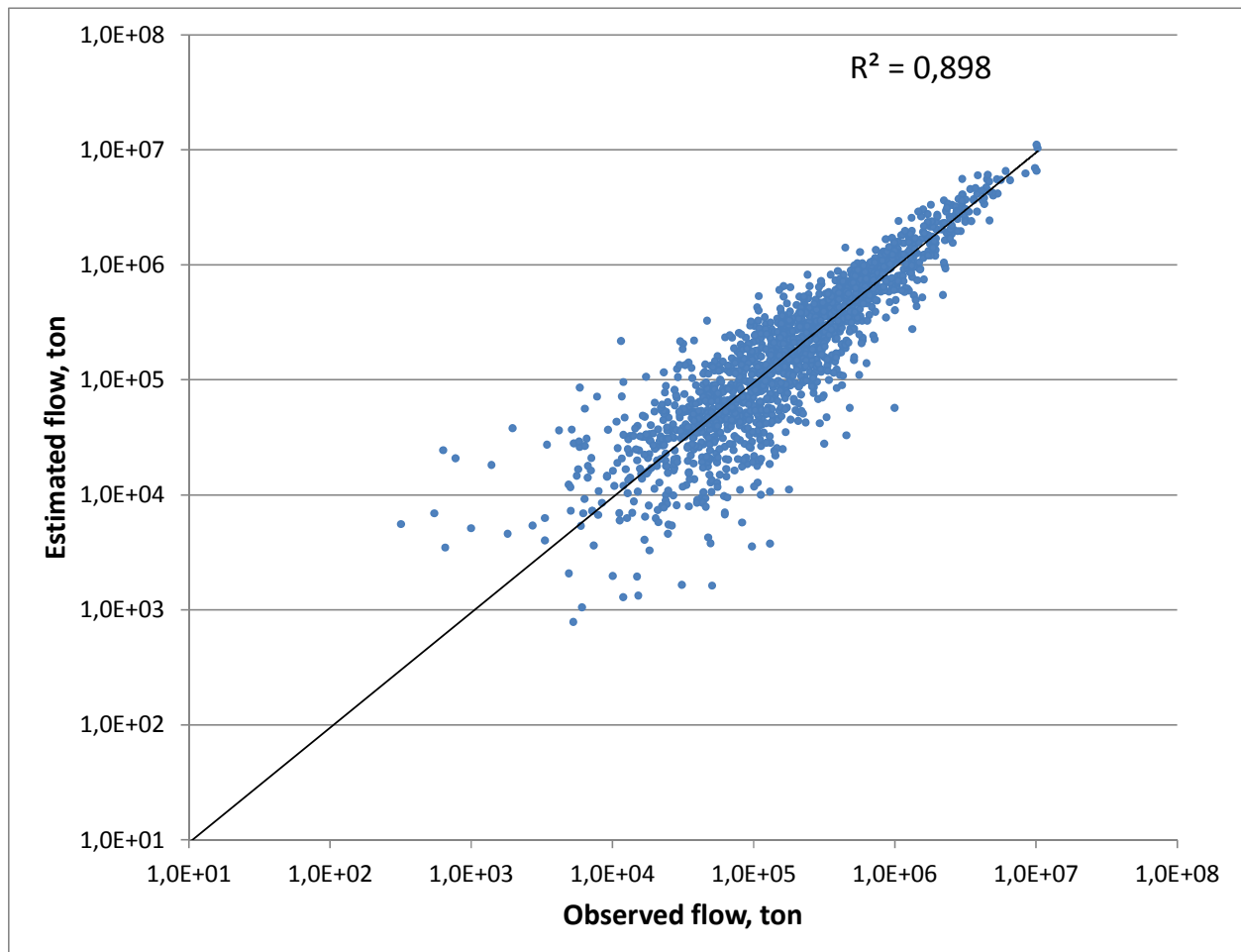


FIGURE 5.2. SCGM output: estimated and observed flows, all commodities, all flow types together

FIGURE 5.2 shows the SCGM estimation quality for all flows grouped together. The determination coefficient reaches the value of 0,90, which is the maximum value among all model runs. The model also estimates gravity model sensitivity parameter β conform model

definition (formula 3.4) in the grid search procedure, described in section 3.5.2. The model shows strong sensitivity in respect to the coefficient value, which allows making a conclusion that the coefficient can be used for distance decay estimations. FIGURE 5.3 shows the relationship between sensitivity coefficient β and the gravity model output quality, measured as Mean Square Root Error (MSRE) conform formula 3.15. It can be seen that a global minimum is easily found.

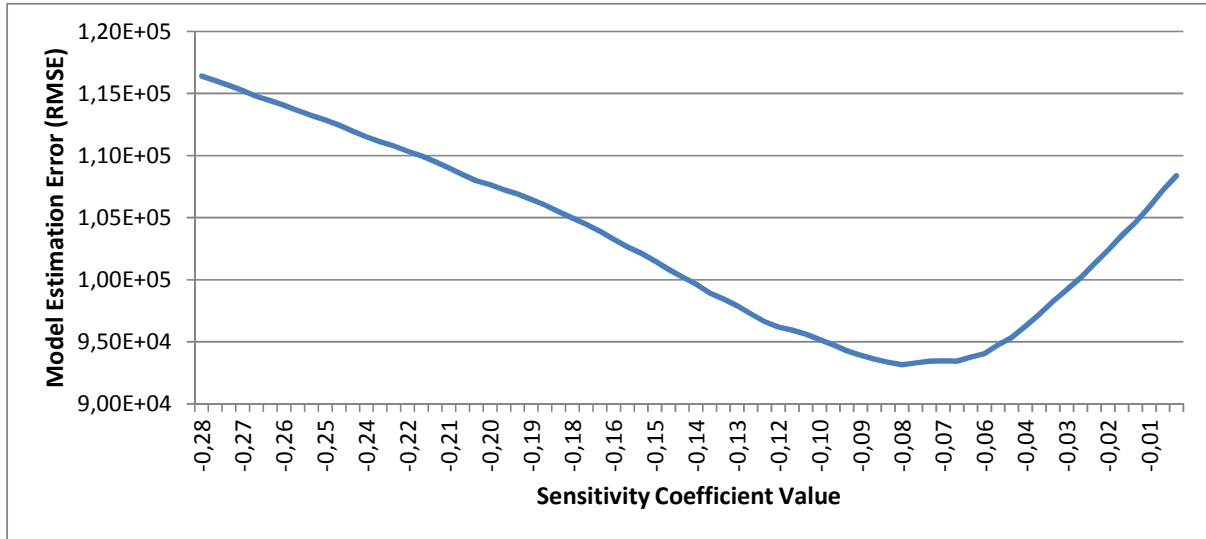


FIGURE 5.3. Relation of SCGM output quality to the model sensitivity parameter value

For this model run, the optimum value of the sensitivity coefficient is -0,08. In this experiment we used a very basic definition of transport costs, equaling 1 ton kilometer to 1 monetary unit (e.g. Euro, Dollar, or Kroner) – the choice of the monetary unit does not matter in this case. For instance, if 1 ton kilometer is set to cost 20 Eurocent (a realistic estimation), then the optimal value of the sensitivity coefficient would be 5 times greater.

5.1.5 Distance Decay

We have estimated the gravity model for O/D flows of the four types and compare the spatial patterns of these flows using the SCGM parameters. Using model definition (formula 3.4), and assuming constant multiplication of production and attractiveness values $p_i q_j$, we define $-\beta c_{i,j}$ of the exponent term of (3.4) $e^{-\beta c_{i,j}}$ as the flow resistance coefficient. FIGURE 5.4 shows the relationship between distance and resistance coefficient in the gravity model. The resulting flow $t_{i,j}$ is strictly proportional to it. Therefore, if the exponent value reduces by the factor 2, the resulting flow will also decrease factor 2. The exponent value for the shipment of zero distance is 1; in case of $\beta = -0,02$ the exponent becomes 0,5 for the cost value (i.e. distance) of 35. In other words, the flow will decay factor 2 with distance 35 kilometers, it will further decay factor 4 with distance 70 kilometers and so forth. FIGURE 5.4 shows distance decay for this sensitivity coefficient for the realistic range of origin-destination relations.

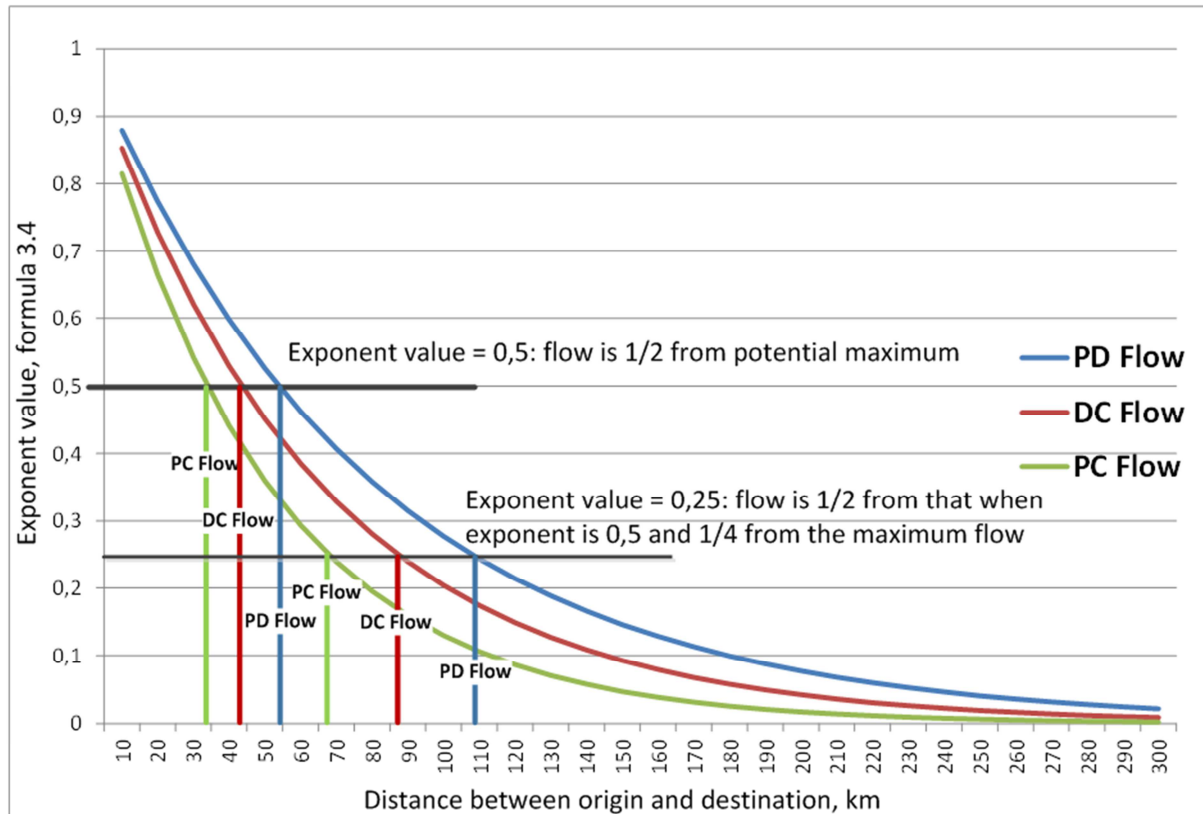


FIGURE 5.4. Distance decay: relationship between distance and flow resistance

FIGURE 5.4 shows that for any value of the sensitivity coefficient β , the $e^{-\beta c_{i,j}}$ term is equal to 1 if the cost is 0, which is the case if distance is 0 km. For any given fixed production and attractiveness values $p_i q_j$, the flow will be reduced by 50% (i.e. factor 2 or halved), when the exponent value reaches the value of 0,5, which corresponds to the value of -0,695 of the $-\beta c_{i,j}$ term. Given the cost function that is proportional to the distance, the exponent value (and the flow size) is determined only by the sensitivity coefficient β . A higher absolute value of sensitivity coefficient β leads to a steeper distance decay, as PC flows decline faster with distance than the PD flows. The PC flow of FIGURE 5.4 (green line) is first halved at the distance of 34km; a doubling of the costs at the distance of 68 km leads to a further halving of the flow at this distance: the flow declines factor 4 at the distance of 68 km compared to zero-distance. The same pattern can be observed for DC and PD flows, however, the halving of the flow happens not at 34km, but at 43 and 54 km respectively.

Frances Cairncross (1997) has advocated an increasing irrelevance of the physical distance, elegantly introducing the notion of the death for the distance. Fifteen years on, his predictions about telecom costs and costs of communication has certainly come true, but is it still true for transport and distribution systems? There is enormous observed goods flow in international trade, which bridge over huge distances. A quick look at the international trade statistics, at for instance ComExt Eurostat database (ComExt intra- and extra-EU trade data), might suggest that the distance has indeed become less relevant.

However, a look at FIGURES 5.2 and 5.4 suggests that physical distance play indeed an important role in the observed physical flows. Without any further generalization, it can be

concluded that interregional road transport in the Netherlands is governed by transport costs, which are generally proportional to the distance. A cost function that is proportional to the distance explains physical flows very well, with the fit of 0,9 (determination coefficient value).

This sub-section looked in detail at SCGM estimation quality, relation between model error (MSRE difference between observed and estimated interregional volumes) in respect to sensitivity coefficient value. The distance decay has also been discussed. The following sub-sections will consider model application to various other types of flow.

5.1.6. Estimation of sub-flows

To make the description of the SCGM experiments shorter, 4 subsets of production/consumption locations combinations and commodities have been defined. **Subset I** uses the strict definition of production and consumption location types and includes all 52 commodities in the total flow. **Subset II** uses the strict definition of production and consumption locations, but restricts commodities selected to the so-called palletizable commodities. The palletizable commodities are shown as selected commodities in FIGURE 5.1. The choice for the palletizable commodities has been somewhat arbitrary: we deem palletizable those commodities that are suitable for transport on pallets (consult Annex 1 for the English list of (palletizable) commodities). **Subset III** uses a broader definition of production and consumption locations. We consider port and terminal location types as production locations too, in other words, mostly import related flows are considered as production related flows. On the consumption side, we use consumption and production location types as consumption points. This choice is justified by the fact that if a production location receives goods, those goods are most probably used as input, i.e. consumed for production purposes. The flow to terminal facilities is not considered as a consumption-related flow, because it has a different nature, namely goods are mostly grouped and put into containers, which is an opposite of disaggregation for the final consumption. **Subset III** includes all commodity types. **Subset IV** uses the same broader definition of consumption and production locations, but includes only palletizable commodities.

TABLE 5.1 provides a summary on the transport flow estimation quality in the SCGM in the form of determination coefficient R^2 , which shows the quality of fit. TABLE 5.2. provides a summary on the distance decay, measured as the distance in kilometers for which the flow reduces by the factor 2, for the considered flows.

TABLE 5.1. Flow estimation quality per flow type, determination coefficient

Flow type	Flow subset type			
	Subset I	Subset II	Subset III	Subset IV
PC	0,78	0,71	0,79	0,75
PD	0,50	0,47	0,81	0,76
DC	0,81	0,81	0,81	0,81

TABLE 5.2. Distance decay per flow type (flow reduction by factor 2), km

Flow type	Flow subset type			
	Subset I	Subset II	Subset III	Subset IV
PC	31,5	45,1	34,3	54,2
PD	54,3	58,4	53,9	47,9
DC	41,9	45,3	43,6	46,4

SCGM modeling of production – consumption flows

The model shows a stable and good quality production-consumption flow estimation for all 4 subsets. However, there are some slight differences in estimation quality and distance decay factor. The determination coefficient is in the range 0,71 – 0,79: it can be observed that when all commodities are considered, it tends to be higher (0,78 and 0,79 for all commodities, subsets I and III, and 0,71 and 0,75 for the flows restricted to palletizable commodities, subsets II and IV). Distance decay shows that if all commodities are included, the flow decays factor 2 with the distance 31,5 and 34,3 km for subsets I and III respectively. If we restrict flows to palletizable commodities the distance decay becomes less steep, 45,1 km and 54,2 km for subsets II and IV respectively. We suspect that these observations are due to the fact that palletizable commodities are less sensitive to transport costs by having greater monetary density value, in other words transport costs constitute a smaller share of the total product costs such that other considerations than transport costs play a role in determination of flow. Less expensive commodities are relatively more expensive to transport and thus show tendency for shorter production-consumption legs. The flow of less expensive commodities is also easier to predict using a cost function, as they are more sensitive to transport costs. This results in somewhat larger determination coefficients.

SCGM modeling of production – distribution flows

Considering the quality of the production-distribution flow estimation, a firm conclusion can be made that the choice of production location type has a strong impact on estimation quality. If the strict criterion is used for production locations (subsets I and II), the value of determination coefficient is around 0,50. If the broader definition of production locations is chosen, then the determination coefficient is substantially higher (0,81 and 0,76 for subsets III and IV respectively). This dualism suggests three plausible issues related to the data. First is that production-distribution flow is “less dense” in a sense that flow matrix contains a substantial number of zero flows. This could be attributed to the possibility that not every location has a production-distribution relationship; another possibility is that the data is a survey sample and not all transport movements could be registered. Second is that the goods coming from the terminals show greater sensitivity to the transport price signals, and thus can be very well modeled by a cost function based on the distance. The ports and terminals are fixed locations and distribution systems and warehouses might be located to serve these facilities efficiently. Third is the possibility that goods produced in the Netherlands are less sensitive with respect to transport costs from production to distribution. Production is more disperse than fixed locations of ports and terminals; the choice for distribution center locations might be more influenced by the consumption patterns, then production patterns: this is confirmed in the analysis of distribution – consumption flows in the next section.

The observation that distribution-consumption flows are more accurately modeled by the transport costs than production-distribution flows has an impact for the design of the logistics chain model, discussed in the section on the LCM model. The chain model assumes that all goods brought to distribution will be further shipped to consumption. Transport costs for the leg from distribution to consumption are more determining for the choice of location for distribution facilities. This is logical given the fact that distribution facilities serve not only the function of goods storage, but also split larger incoming shipments into smaller outgoing shipment batches. Thus, transport costs per ton-kilometer unit are higher on the distribution-consumption flow.

SCGM modeling of distribution – consumption flows

The gravity model provides the best flow estimation results for the distribution-consumption type of flow among the three studied types of flow. The determination coefficient is 0,81 for all 4 experiments, deviating from average not more than by a half percentage point. The distance decay does not vary significantly between the four experiments neither, showing slightly higher values for the flows that include only palletizable commodity types. This suggests that the flows become more homogenous as they approach consumption sites. It can also be interpreted in a way that the flows from distribution to consumption are well optimized, showing a nice estimation fit realized by a function based on distance-proportional costs. The distribution-consumption flows include substantial volumes generated by large retailers, which have probably optimized their logistics operations, as for any given consumption location, large chains have an (optimal) choice of the distribution location supplying the consumption point.

5.1.7. Transport demand price elasticity

Price elasticity of (transport) demand E_d is a measure showing responsiveness of the system in respect to price signals. The dominant response of demand for certain good or service is that it decreases if the price of the said good or service goes up. The transport demand price elasticity is a very practical concept. The literature overview on this subject provides a wide range of values, see Significance & CE Delft (2010) and de Jong et al (2010). The most used definition of transport elasticity of demand is the following:

$$E_D = \frac{dQ/Q}{dP/P} \quad (5.1)$$

Where Q is quantity of demand and P is price

Essentially formula 5.1. measures the relation of change in demand (numerator) in response to the change in price (denominator). For instance, if a change in price of 5% leads to a change in demand of -3% (e.g. in the realm of transport demand is expressed in tones shipped or tone-kilometer transported), then demand price elasticity would be -0,6.

Similarly to the distance decay, it is possible to compute transport demand price elasticity based on the outcomes of SCGM runs, using only one variable, namely sensitivity parameter β defined in formula 3.4. Keeping regional production and attraction values p_i and q_j constant,

a change in price c_{ij} would lead to a change in the interregional flow $t_{i,j}$. In such definition, only the sensitivity parameter β will influence the elasticity of transport demand.

Transport demand elasticity in this definition also depends on the distance between origin and destination. Demand is less elastic for the price signals on short distances, where per kilometer transport cost represents a relatively small share of total costs. On the longer distances transport demand is more sensitive to the relative change in transport price, as it is much bigger in absolute terms.

For the purpose of analysis, transport demand price elasticity for the four flows is considered for the three distance classes: 25, 50 and 100 km. TABLE 5.3 shows the elasticity of the flows.

TABLE 5.3. SCGM transport demand price elasticity per flow type and distance class

Flow Type / Distance Class	25 km	50 km	100 km
PD	-0,316	-0,622	-1,205
PC	-0,498	-0,971	-1,848
DC	-0,389	-0,764	-1,469
OD	-0,484	-0,944	-1,799

Due to the fact that price elasticity depends on the relation distance, elasticities of TABLE 5.3 should be used with caution. The data indicates that average distances are larger for PD flows and smaller for direct PC flows, with DC flows somewhere in between (see FIGURE 4.3). The figure also shows that approximately 35% of the flow is realized for the relations with the distance of less than 30km and more than 50% of the flow is realized over the distances of up to 50 km (these data does not take into account intraregional flows, where the distances are even shorter). The logistics chain model applications of Chapter 6 present a more clear-cut estimation of transport demand price elasticity at the level of trade, transport and logistics system of the Netherlands. Section 6.3.2 will further compare and discuss SCGM and LCM elasticity estimations.

5.1.8. Regional transport production and attraction

The SCGM estimates regional flow production values (p_i) and regional flow attraction values (q_i) as specified in the gravity model definition formula 3.4. The values generally say how large some certain activities are in regions generating and attracting transport flows. In the case of logistics modeling, it is interesting to look at the constituting parts of the total OD flow: if the constituting parts are similar, then there is no strong need to treat them separately; if they are sufficiently different, then it is worth to undertake an effort to model the underlying factors behind the parts. In case of transport OD flows, they consist of PD, PC and DC flows. If regional level generation and attraction of these three sub flows are different, then, the flows should not be lumped together into a single O/D flow (as it is the common practice in transport modelling currently), but treated separately by, for instance, a logistics chain model.

The p_i and q_i values are unitless in formulation 3.4. Implementation of the SCGM normalizes these values by fixing p_{30} to 1. Due to the fact that the PD, PC and DC flows are not equal in size (direct flows PC are much larger than the flows via distribution), the p_i and q_i values should not be compared directly for the different flow types. A better way of comparison is to determine the level of proportionality between these vectors. TABLE 5.4 shows *determination coefficients* (R^2) between production vectors in a matrix form; TABLE 5.5 shows the values of *determination coefficients* (R^2) between attraction vectors for the four types of flow.

TABLE 5.4. *Determination coefficients* (R^2) between production vectors for the four flow types

	OD	PD	PC	DC
OD	1,000	0,676	0,946	0,429
PD	0,676	1,000	0,530	0,169
PC	0,946	0,530	1,000	0,349
DC	0,429	0,169	0,349	1,000

TABLE 5.5. *Determination coefficients* (R^2) between attraction vectors for the four flow types

	OD	PD	PC	DC
OD	1,000	0,516	0,753	0,682
PD	0,516	1,000	0,394	0,305
PC	0,753	0,394	1,000	0,614
DC	0,682	0,305	0,614	1,000

TABLES 5.4 and 5.5 show a substantial similarity between the attraction and production vectors for the total OD and PC flows. This can be explained by the fact that PC flows are dominating by volume in the total OD flows. However, regional flow attractiveness and generation for the other flow types shows very substantial dissimilarities. For instance, at the regional level flow production from production locations destined to the distribution does not have much similarity with the flow production from distribution locations to the consumption. The comparison of estimations of the P and Q vectors in the gravity model for the different flow types clearly shows that for different flow types regional flow production and consumption differ substantially. This observation warrants modeling techniques that treat PD, DC and PC sub-flows individually, without lumping them all together into the total O/D flow. In other words, the data strongly supports the need to develop logistics chain models.

5.1.9. Conclusions on the Gravity Model

This section has discussed in detail the separately calibrated gravity model's application cases. The gravity model deserved considerable attention in the framework of this research due to the following three major reasons. The first reason is that the model allowed looking at the CBS survey data from a relatively simple transport modeling perspective: before going into complexities of logistics chain model, gravity model experiments allowed assessing the possibility to model transport movements using very simple cost functions. The experiments confirm that CBS road transport survey data captures major transport-economical laws such as distance decay and general negative dependence of the flow volume on the transport costs.

This fact follows a major practical conclusion: transport costs proportional to ton-kilometer volume shipped is a good predictor for the flow intensity.

The second reason is that SCGM applications show that PD, DC and PC flows have different properties, such as distance decay (price elasticity), attraction and generation. This fact helps arguing that a logistics chain model is indeed desirable in the transport modeling framework, as opposed to the current practice of lumping all constituting transport sub-flows into a total OD transport flow, and thus losing important details related to the function of distribution centers as the nodes in transport networks.

The third reason is that we cannot avoid a gravity model implementation for the greater purpose of this study, namely modeling of the logistics choices in the aggregate freight flow. As we argue in the sections on data and logistics chain model, there are no trade flow data available in observed form at the origin-destination level, but these data are required for the logistics model as the input. Therefore, a gravity model has been used to obtain these data. The implementation and exploitation of the gravity model has made it possible to use the model as a part of the combined gravity and logistics chain model.

The implementation of the model is pretty straightforward and realized in Borland Delphi. The model is solved up to a certain small error margin ϵ ; it is calibrated using a grid search procedure described in section 3.5.2. The model implementation is very fast: it takes some 20 seconds to go through grid search procedure containing 500 grid nodes. In other words, it takes approximately 1/25 of a second to solve the gravity model. It is also interesting to note, that the model's computational complexity is polynomial, it does not grow exponentially with the number of regions contained in the source data.

The model has been applied to the three types of flow: production-consumption, production-distribution and distribution-consumption. There have been 4 production / consumption and commodity subsets used. Generally speaking, the more inclusive subset is, the more accurate the model estimations are. However, it can be generalized that the flows to final consumption are estimated more accurately than the flows to distribution. We suspect that the cost of shipments to the final consumption is higher than to distribution per ton-kilometer transportation unit, and thus these shipments are more governed by the transport costs, and to a lesser extent by other factors. For the flow to distribution, there could be other factors at play, for instance, attractiveness of the distribution location. The location's attractiveness will be discussed in more detail in the section on the logistics chain model.

The experimentation with the separately calibrated gravity model strongly supports the need in development of the logistics choice class of models. The standard practice now in transport modeling is to look at the total OD flows, which is confirmed to be insufficiently good as parts of the transport OD flows have different natures and properties. The applications of the SCGM have shown that PD, DC and PC flows have different price elasticities and at the regional level, regions have different propensities to generate and attract these flows. This confirms the assumption that PD, PC and DC flows are fundamentally different, leading to a conclusion that there is a strong need for a model that would explicitly distinguish these flows, such as the logistics chain model discussed in the next section.

5.2. Logistics Chain Model for the Netherlands

This section looks into implementation and estimation of the combined gravity and logistics chain model, explained in the chapter 3.3 on the combined model design and chapter 3.6 on mathematical specifications. The emphasis of this chapter is on empirical validation of the proposed model. It shows that the calibration effort leads to an estimation of the model parameters, which subsequently results in the model output that is close to the regional distribution throughput volumes and interregional OD flows observed in reality. The quality of the estimated model is presented in the form of a match between observed and estimated output values. Finally, the estimated model is critically discussed and conclusions are drawn on the estimation results.

5.2.1. Estimation of the model's parameters

The combined gravity and logistics chain model essentially performs two important steps. In the first step, the gravity model takes regional production and consumption volumes in the form of two one-dimensional vectors and finds corresponding trade flows. The producing regions are, therefore, functionally matched with the consuming regions, as explained in the section 3.5. The flow is computed according to the formula 3.4, where there are two main determining parameters for the size of the flow between regions i and j , namely resistance or generalized logistics costs between i and j , $c_{i,j}$ and the price sensitivity parameter β . The generalized logistics costs are computed in the logistics chain model and are more sophisticated than simple distance-based costs of the SCGM explained in section 5.1: these costs depend on the composition of the logistics chains, which are used to transport goods from the producing region i to the consuming region j . The price sensitivity parameter β intrinsically belongs to the gravity model and influences spatial distribution of the trade flows. In practical terms, if this parameter has a large value, the gravity model will redistribute trade flows in a way that the relations with the least resistance will get a relatively large share of the flow. Other relations get smaller flows. Conversely, if the sensitivity parameter in the gravity model is relatively small, the model will not react strongly to the resistances between producing and consuming regions, redistributing trade flows more evenly between the relations.

The gravity model sensitivity parameter can be treated as unknown, as there are only general estimations for its value coming out from the separately calibrated gravity model estimations, documented in chapter 5.1. These experiments provide a plausible range for the possible values, however, those values are not sufficiently good for the purpose of an accurate estimation of the trade flows. Therefore, the gravity sensitivity parameter β (formula 3.4) is one of the estimated model parameters, which is performed in the calibration process according to formulae (3.30 and 3.31) of section 3.6.1. In the process of calibration the model searches for such a value of the gravity model's sensitivity parameter so that the total modeling outcome is closely matched by the flows and regional distribution volumes observed in reality.

In the second step, the logistics chain model takes the P/C trade flows from the gravity model and translates them into transport O/D flows. The logistics chain model is mathematically

defined in section 3.6. The essence of the logistics choice model is in computation of two choices. For each production-consumption relation i, j the model computes a) the probability that the goods are delivered directly from i to j (formula 3.19) and b) for the part of the trade flow that is transported indirectly via a distribution center the model determines the probability that the distribution center is located in region k (formula 3.22). Therefore, for each trade relation, the top-level choice has two alternatives and the nested choice, which determines the region of distribution, has n choices (for the Netherlands at NUTS3 level $n = 40$).

The top level choice is made in a logit formula (3.19), which uses logit sensitivity parameter α' and (dis)utility for each of the choices for determination of the trade flow split between direct shipments and shipments via distribution. The utility of the choices are computed according to (3.20 and 3.21), which are based on the total logistics costs. At the top choice level, the logit sensitivity parameter α' is unknown (only the workable range of the parameter is known), therefore it is estimated in the calibration run.

The three sensitivity parameters discussed above define model sensitivity towards (dis)utility of the choices. The utility of the choices is determined by the generalized logistics costs, which consist of the transport costs, distribution costs and regional distribution attractiveness. The transport costs and distribution costs are global parameters, i.e. their specific values are applicable for all regions that the model takes into account. The regional distribution attractiveness parameter is related to the local situation in the region to which it refers.

The transport costs are defined in formula (3.16), namely for an arbitrary pair of regions i, j it computes ton-kilometer costs for 3 types of transport: PD, PC and DC. The denominator in (3.16) is the average vehicle load, which is different for these three types of transport. We fix the vehicle-kilometer cost c_{vkm} for all three types of transport to 1,30 Euro per vehicle-kilometer. Indeed, the costs associated with carrying out transport activities do not depend on how well the vehicle is loaded, but defined by the distance driven. The loads, however, depend on the mission or type of transport activity.

Therefore, the vehicle loads L^{PD} , L^{PC} and L^{DC} are the model parameters and estimated in the calibration procedure. This design has two properties: first the model parameters are grounded in reality through a single model constant, c_{vkm} . Setting it to 1,30 Euro per vehicle-kilometer is an assumption, however, a realistic one. This constant provides all other parameters with a link to the reality: the distribution costs and estimations of regional attractiveness are based on real Euro estimations. The second property is that this design allows simultaneous estimation of the average vehicle loads and ton-kilometer costs, as these are directly linked in (3.16).

The model formulation that uses only transport vehicle-kilometer cost as a real-world parameter can be successfully estimated, provided a realistic range of the parameters within which the variable values are searched for in the calibration run. However, such model is not normalized. For normalization, we have chosen to fix the average vehicle load on production-distribution segments L^{PD} to a value estimated in a not normalized calibration run. A re-run of calibration with the fixed L^{PD} value led to a similar estimation of other model parameters, confirming robustness of the model in respect to the chosen normalization.

The warehousing and distribution costs are also treated explicitly in the model. If a logistics chain uses a warehouse or distribution center, there are some extra costs linked to the usage of distribution. These costs are related to vehicle loading and unloading activities, storage of the goods, capital invested in the goods and possible value depreciation of the goods. All these cost components are included in a single cost of distribution c^w introduced in (3.24). The distribution cost is estimated during calibration procedure. Therefore, the model has 4 global cost-related parameters: three related to the transport costs per transport type and the fourth is related to distribution costs.

In principle, the combined gravity and logistics choice model can be estimated using only the seven discussed global parameters. Experiments on transport flow estimation using only global parameters result in computed flows that have an R^2 measure of around 0,3, when compared to the observed transport flows. However, it is important to consider the fact that the experiments with the separately calibrated gravity model, discussed in section 5.1, generally produce flow computations with R^2 measure in the range 0,5-0,8. Therefore, the combined model is deemed to be capable of a better performance. Using analogy with the gravity model, which in formulation (3.4) includes regional production and attraction variables p_i and q_j , the location dependent model parameters can substantially improve the model output. In the context of the logistics chain model, the A_k variables are the location-specific parameters. A substantial model estimation improvement is achieved by the introduction of the location-specific extra costs A_k , showing the necessity of these parameters.

Indeed, the flows via distribution centers are guided not only by the transport costs and costs of distribution. There are some location specific factors at play as well, for instance, accessibility to backbone networks, quality and availability of logistics facilities, ground prices, availability of labor and the labor compensation rates. These objective factors can be extended with non-quantifiable subjective factors as well. The location specific attractiveness factors are not known, however, can be estimated in the model. Formula (3.24) introduces location-specific attractiveness variable A_k , which is expressed as the extra Euro cost per ton of distribution throughput in region k . A negative value of this variable reduces the total cost of the corresponding logistics chain, and a positive value increases it, thus negative values lead to a more attractive distribution in region k . For the Netherlands, the number of regions is $n = 40$, therefore, there are 40 values of the parameter A_k to estimate. In total, the combined gravity and logistics chain model has 47 parameters for estimation.

5.2.2. Model implementation and calibration

The combined gravity and logistics chain model has been implemented in Borland Delphi (object Pascal). Similarly to the implementation of the separately calibrated gravity model, this implementation has a very simple user interface, which allows selection of commodities included into computation and selection of location types belonging to production and consumption, expanded with some extra research functionality. The model input is loaded as a tab-separated text file; the software generates output in the form of two tab-separated text files. The first file contains detailed flows, including regional distribution throughput. The second file contains data on the values of estimated parameters, if calibration procedure has been performed. The screenshot of the software implementation is given in FIGURE 5.5.

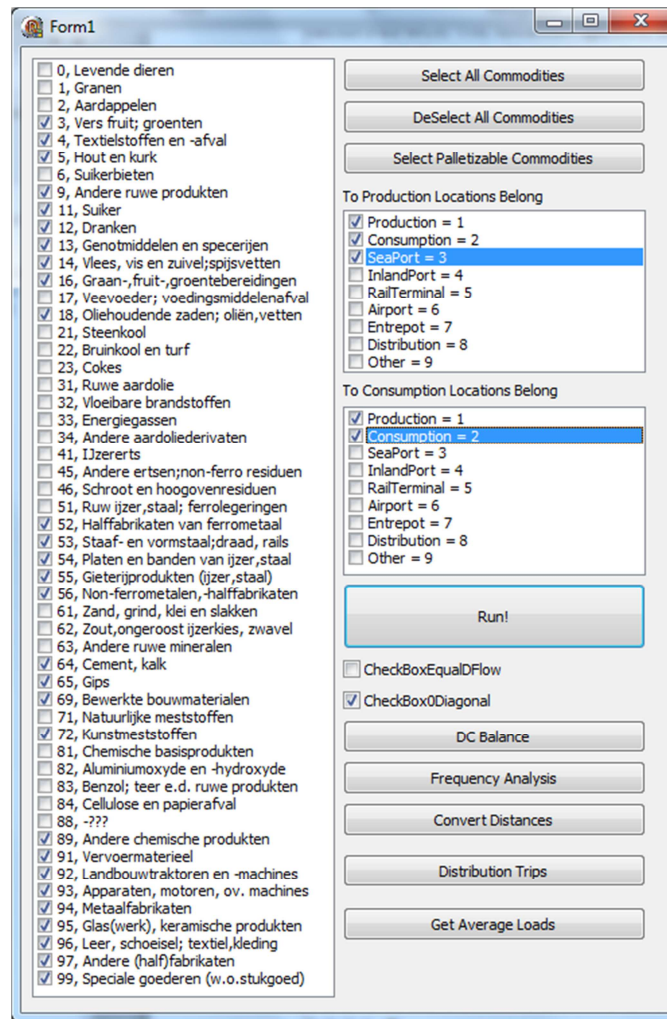


FIGURE 5.5. User interface implementation of the combined gravity and logistics chain model.

The model relies on tab-separated files for input and output. This implementation choice has four advantages: first, the process of loading and saving data can be implemented in a simple way, without reliance on third-party software functionality. Second, the loading and storing of data takes a few seconds for file sizes of more than 10 MB. Third, interfacing with Microsoft Excel is simple and is in the form of copy-paste of textual information. Fourth, the data analysis can be done in Excel or any other appropriate software, such that the model is not burdened with the development of the own output analysis tools.

The model takes CBS road transport survey data as the input. As in the case of the separately calibrated gravity model, a road transport survey dataset for the period 2007-2009 has been used, the reasons for the data summation for these three years are given in the section 4.2.

5.2.3. Calibration procedure

The calibration procedure consists of a number of iterative steps, in which the best values for each of the 47 model parameters are estimated (see section 3.6 on mathematical definition of the calibration procedure). The best value of a parameter is estimated in a single variable grid

search procedure. For each model parameter, a range between the lowest and highest possible value is defined. This range is consequently split into a number of points for which the model is computed (see TABLE 5.6 for the parameter ranges and number of steps).

TABLE 5.6. Model variables' estimation boundaries and number of estimation points.

Parameter	Lowest value	Highest value	Number of grid points (steps) in base estimation
Gravity sensitivity β	0,001	0,300	200
Top level choice logit α'	0,001	1,000	200
Nested choice logit α	0,001	0,500	200
PD vehicle load, ton	3,000	20,000	200
PC vehicle load, ton	3,000	20,000	200
DC vehicle load, ton	3,000	20,000	200
Distribution cost, Euro / ton	3,000	30,000	200
Regional distribution attractiveness A_k , Euro /ton	-60,000	60,000	150

The following procedure is used to find the best value of a parameter. Let P_i denote the value of parameter i , $i = 1..47$, where 47 is the total number of model's parameters to be estimated. We consider the best value P_i^b to be the value for which the RMSE difference between observed flow and model estimated flow (3.31) is at minimum, or the RMSE between observed distribution throughput and model estimated throughput (3.30) is at minimum, depending on the calibration objective (flow or throughput).

To find the value of P_i^b , the lowest P_i^L boundary and the upper boundary P_i^H of parameter P_i are defined (TABLE 5.6 presents these values): the best value of parameter P_i^b is deemed to be within these boundaries. The range between P_i^L and P_i^H is split into S same incremental steps. The P_i^b is found in the following steps:

1. Set Quality = M (sufficiently big number)
2. Set $P_i = P_i^L$
3. Compute Quality' as RMSE between estimated and observed flows (3.31)
4. If Quality' < Quality then perform items #5 and #6, otherwise go to #7
5. Set $P_i^b = P_i$
6. Set Quality = Quality'
7. If $P_i < P_i^H$ then set $P_i = P_i + (P_i^H - P_i^L) / S$, go to item #3. Otherwise stop, the best value P_i^b is found

Note that a smaller value of the Quality variable represents a better value of P_i parameter in respect to model estimation (3.22), as Quality is computed as RMSE between observed and estimated flows. The smaller this difference is, the better is the result.

The presented algorithm finds an optimal value for one of the 47 model parameters. Once this value is found, the index i is increased by 1 and the procedure repeats for the next parameter, until for all parameters the best values are found. There two issues to be considered: first, once the best value for parameter $i + 1$ has been found, the current "best value" of parameter i might cease to be the best. This is due to the fact that there is a functional relationship

between the model parameters. For instance, the average transport loads for PD transport leg may depend on the warehousing costs. The second issue is that this heuristics finds a local optimum, which may depend on the order in which the parameters are estimated.

The first problem can be overcome relatively easy. If the estimation procedure is repeated 5-10 times for all model's parameters, the parameters settle around some stable values and do not tend to change with the subsequent iterations. The second problem cannot be solved completely, as there is no practical way to find the absolutely best values for all parameters, as the size of the problem is too large. However, searching for the best values of parameters in random order suggests that the solutions found by the described optimization procedure do not tend to be settled in 'bad' local optima. The experiments with the random order of parameter estimation lead to the similar results as it is in case of strictly sequential parameter search procedure.

The calibration procedure has been performed conform formula 3.31, however, some extra details have been included. Formula 3.31 is applied to the total road transport OD flows, which are obtained as the sum of the constituting PD, DC, and PC flows. Considering the fact that the majority of the shipments are direct shipments (some 85% of the total ton volumes), if the model is estimated on the total OD flows, the impact of direct flows on model estimation would be too large. Considering also the fact that the root mean square error uses quadratic difference between observed and estimated values, the larger flow is even more prominent in error minimization. Therefore, to make distribution-related flows more prominent in the model estimation, which is indeed the goal of the model development, it is possible to give distribution flows a larger weight in the model calibration. Formula 3.31 can be re-written in the following form:

$$\min \quad a \sqrt{\frac{1}{n} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (f_{ij}^{G,PD} - f_{ij}^{O,PD})^2} + b \sqrt{\frac{1}{n} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (f_{ij}^{G,PC} - f_{ij}^{O,PC})^2} + c \sqrt{\frac{1}{n} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (f_{ij}^{G,DC} - f_{ij}^{O,DC})^2} \quad (5.2)$$

Where a , b and c are the weight coefficients for the production-distribution, production-consumption and distribution-consumption flows respectively. The base model calibration has been performed with coefficients $a = c = 1$ and $b = 0,1$, thus RMSE of the distribution-related flows has been considered 10 times more important than the RMSE of the direct flows. Indeed, if the distribution-related flows are not given a substantial importance in the logistics chain model, the total flow estimation may be better performed by the gravity model, which does not account for the logistics chains.

5.2.4. Results of the base model calibration

For the base model calibration all 52 NSTR commodity types have been included into the relevant flows. A broad definition of production and consumption locations have been used, equivalent to the definition of Subset III (see section 5.1.6) used in the estimation of the gravity model. A broad definition of consumption and production locations generally leads to a better estimation result, as well as to a more comprehensive transport modeling. FIGURES

5.6-5.9 show the model estimation quality in respect to the match of PD, DC, PC and the total OD flows.

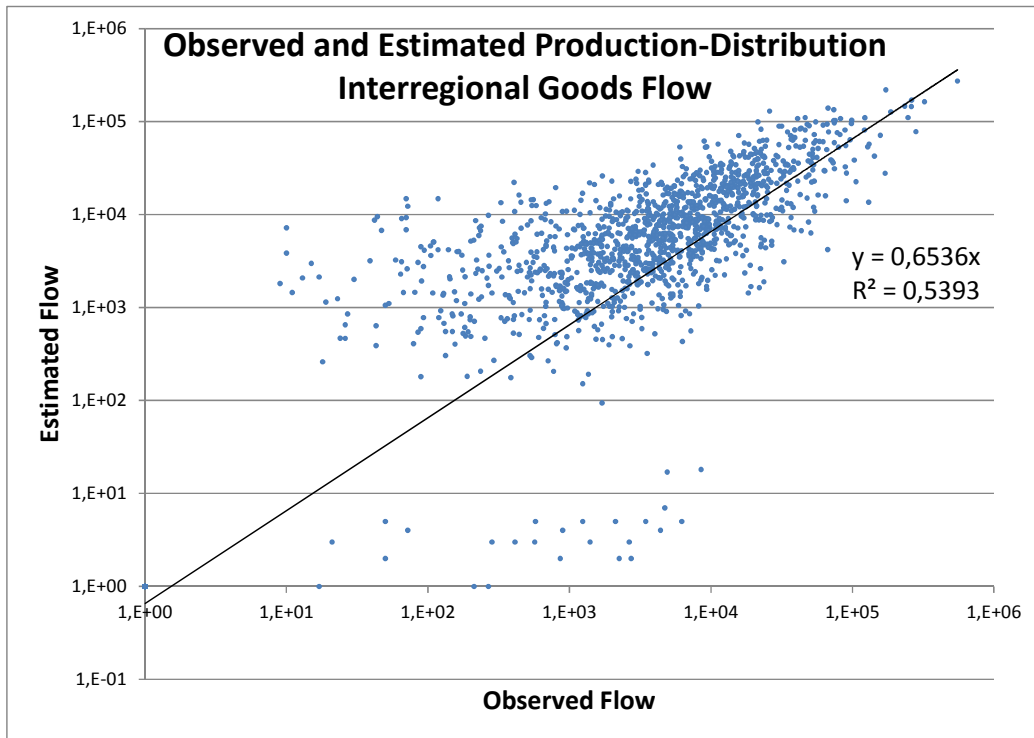


FIGURE 5.6. LCM model calibration result in respect to PD flows
NB dots with small estimated flow volume are related to Haarlem region

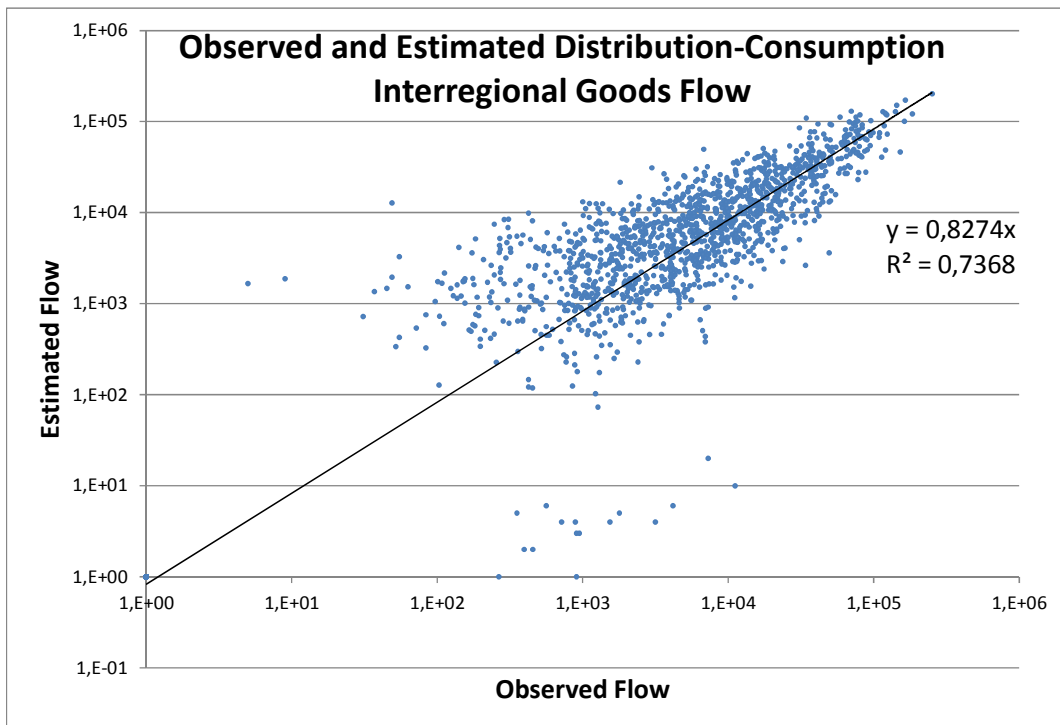


FIGURE 5.7. LCM model calibration result in respect to DC flows

Please note a phenomenon of flow underestimation in FIGURES 5.6 and 5.7. There is a region along the X axis of the graph, where the model estimated flow values are between 0

and 25 ton, while observed flow values can be as high as 11 000 ton. This DC-related flow underestimation in the LCM concerns the region around the city of Haarlem (NL324 Agglomeratie Haarlem), which presents an interesting case on how the model works. The Haarlem region is exceptional: it generates only 33 000 ton of distribution outflows per year, while other 39 Dutch regions generate 19 508 000 ton of annual distribution outflows. Thus, Haarlem generates only 7% of the average distribution volumes, and in this sense is an outlier.

Given the location of Haarlem in the densely populated Randstad region, one would expect the region to be substantially involved in the distribution activities. Based on the transport and warehousing costs, Haarlem would get substantially more volumes than the road transport survey data of CBS are showing. The model would consider it very suitable for distribution, based on the transport and warehousing costs. But in practice the region is not actively used for the distribution activities. To make the flow estimations more realistic, the LCM model assigns a large $A[k]$ value for the region of Haarlem, to ensure large costs of the region and to divert the flows to other regions and to direct shipments. A higher value of $A[k]$ for Haarlem makes the total RMSE of the model estimation smaller (better).

The total RMSE of the estimated model is at the minimum when the model assigns high region-specific costs to the Haarlem region. At the level of the flows it means that the model estimates very small flows (0-25 ton) related to the distribution around Haarlem. The CBS data also shows small distribution flows in the Haarlem region, but they are comparatively larger, thus well visible on the figures showing the quality of fit. The CBS data also shows that there are a substantial number of OD-relations showing zero flows (i.e. CBS did not observe any flows in its sampling) for specific relations. The Haarlem effect is also present in the estimations of direct PC flows (FIGURE 5.8), and by implication in the total OD flows (FIGURE 5.9). The model “sends” a share of the flows related to distribution in Haarlem via distribution in other regions, and it also “sends” the flows directly, thus it visibly overestimates direct flows in respect to Haarlem.

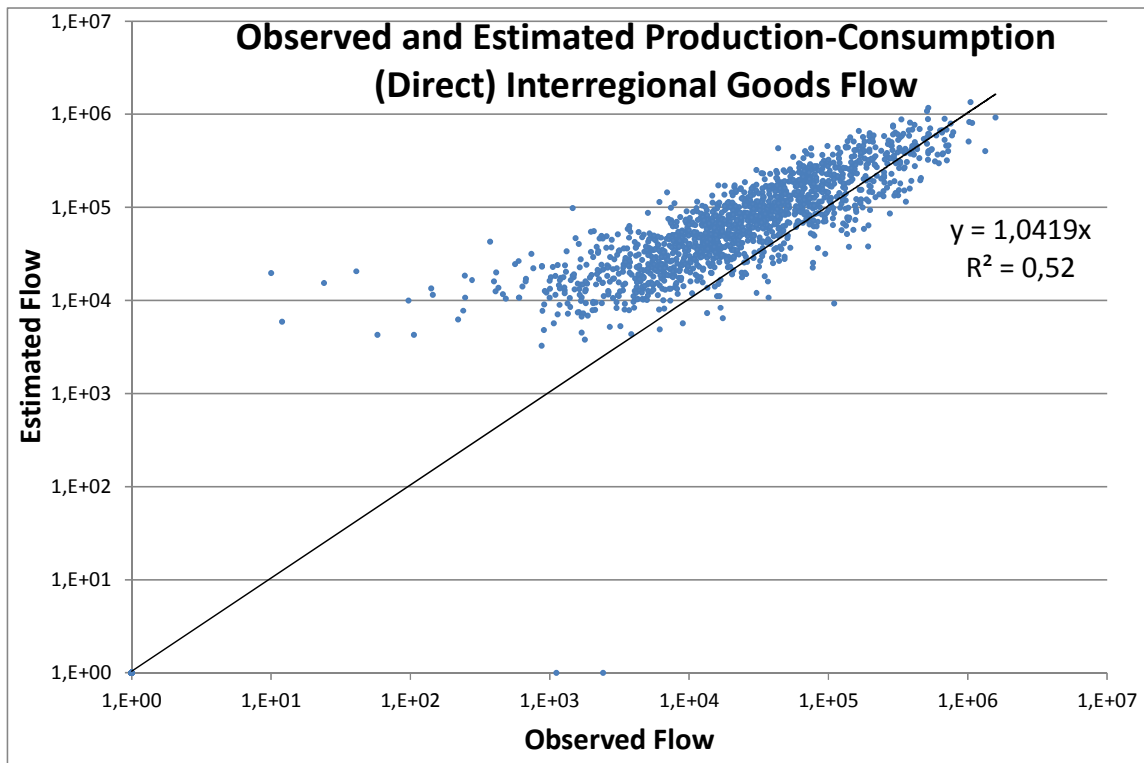


FIGURE 5.8. LCM model calibration result in respect to PC flows

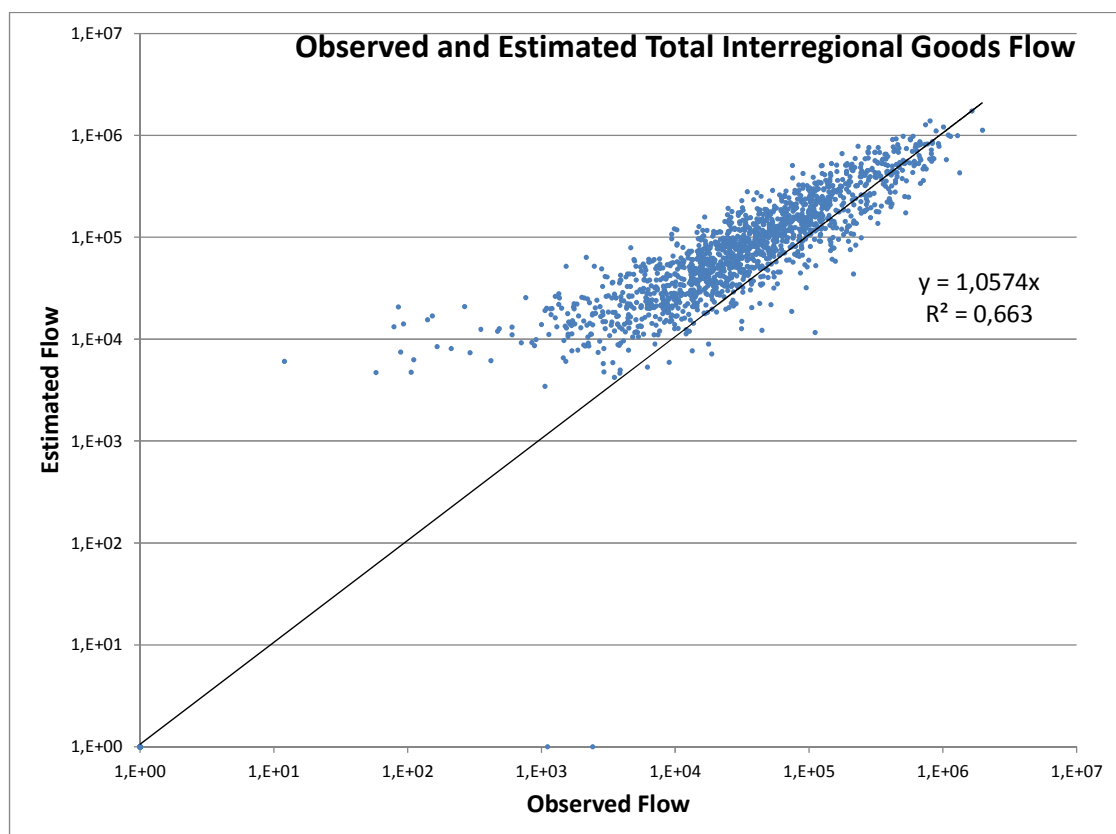


FIGURE 5.9. LCM model calibration result in respect to total OD flows

It is important to observe that the logistics chain model reaches a good estimation of the total flows, with the R^2 measure of 0,66 and a small bias ($y=1,06x$). The quality of the total OD flow estimation (FIGURE 5.9) is significantly better than that of the estimation of PC flows

(FIGURE 5.8), especially taking into account the fact that PC flows constitute some 75% of the ton flow volume.

When estimation quality of constituting sub-flows is considered separately, the same pattern can be observed as in the case of SCGM estimation (section 5.1). TABLE 5.1 on quality of the SCGM estimation shows the pattern: the best estimation results are achieved for DC flows, while PD flows are generally estimated with the smallest R^2 measure. As it has been discussed in the conclusions on SCGM estimation (see 5.1.10 Conclusions on the Gravity Model), there could be two factors explaining this phenomenon. First, as DC flows are generally more expensive than other types, companies pay more attention to optimization of these flows, thus they are well governed by the cost functions. Second, the locations of DCs are probably more rational with respect to consumption patterns than location of production facilities and ports.

The bias in the PD and DC flows should be attributed to the two main reasons. The first reason is the inequality of PD and DC volumes in the CBS survey data. The survey data shows that DC ton volumes are larger than PD ton volumes. The logistics model assumes that PD volumes are equal to the DC volumes: there is no production or consumption activity assumed at the distribution centers and warehouses. The distribution centers are just nodes in the chain model, thus total PD flow is equal to total DC flow in the model. We have made no attempt to correct for this anomaly and used the data as they were. Consequently, we find the results in the quality of fit.

The second reason is that in the model the diagonal flows, i.e. the flows originating and destined within the same region are not taken into consideration. The diagonal flows are not considered because it is not possible to find an accurate transport cost function for them (the distance within a region is unknown or too small), as well as due to their relatively large size. If taken into account these flows would prevail over interregional flows, thus making chain-level modeling a bit less accurate. However even if not taken explicitly in computation of the RMSE, the diagonal flows have an impact on the quality of fit and bias. The diagonal flows may “consume” a larger or a smaller share of flows, thus introducing an uncontrolled error into the estimation of the flows.

The model performs very well in respect to estimation of the regional distribution volumes, see FIGURE 5.10. The observed and estimated regional distribution throughput volumes have been obtained as the sum of all incoming and outgoing volumes to distribution location per region: there are 40 observed and 40 estimated values, equaling the number of regions in the system under consideration (see formulae 5.3 and 5.4.

$$w_i^G = \frac{1}{2} \sum_{j=1}^n (f_{i,j}^{G,PD} + f_{i,j}^{G,DC}), \forall i \quad (5.3)$$

$$w_i^O = \frac{1}{2} \sum_{j=1}^n (f_{i,j}^{O,PD} + f_{i,j}^{O,DC}), \forall i \quad (5.4)$$

Where w_i is the regional distribution throughput, $f_{i,j}^{PD}$ is the interregional production-distribution road transport volume (incoming volumes to the distribution facilities) and $f_{i,j}^{DC}$

is the interregional distribution-consumption road transport volume (outgoing volumes from the distribution facilities). For the model estimated flows (upper index G), the sum of incoming volumes is strictly equal to the sum of outgoing volumes per region, which is enforced by the model design. For the observed flows (upper index O), the incoming volumes are not equal to outgoing due to sampling error.

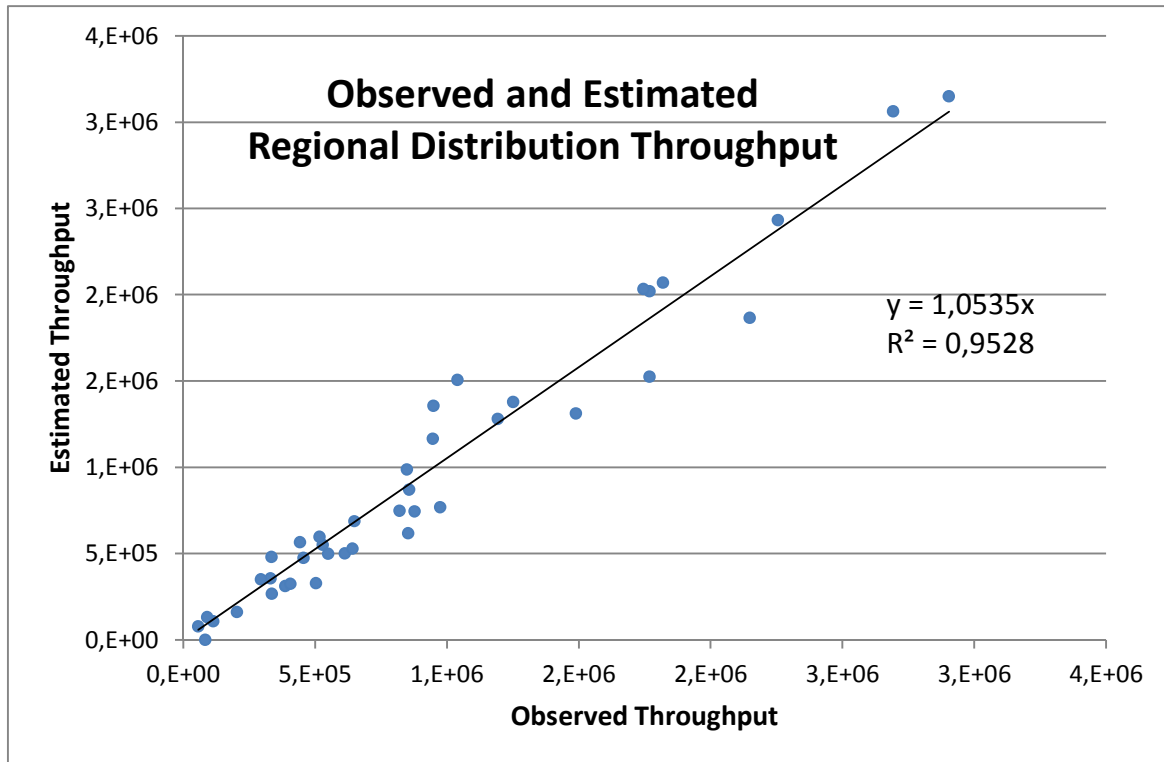


FIGURE 5.10. LCM model calibration result in respect to implied annual regional distribution volumes

It should be noted that the results presented in FIGURE 5.10 have been obtained in model calibration according to formula 5.2, namely it was not optimized for minimization of the difference between observed and estimated distribution throughput volumes. This good fit has been obtained by implication of difference minimization between observed and estimated flow volumes. The experiments with the explicit minimization of the difference between observed and estimated distribution volumes result in similar estimation quality results (Davydenko, Tavasszy, 2013b).

In this chapter the attention has been primarily devoted to the estimation of the interregional flows, similarly to described in (Davydenko et al., 2013), which is more difficult than estimation of regional distribution throughput only. The model described here has 47 variables (of which 7 are related to sensitivity and costs and 40 are related to regional distribution attractiveness). Thus, estimation of 40 regional distribution throughput values is much less challenging than the estimation of $40 \times 40 = 1600$ interregional flow values, especially taking into account the fact that the total flows are composed of the three underlying PD, DC, and PC sub-flows. This result is a confirmation that the model is capable of a good performance with respect to estimation quality of the interregional OD flows, as well as with respect to the estimation of implied distribution throughput volumes.

5.2.5. Base model calibration: estimated model parameters

TABLE 5.7 presents values of the logistics chain model variables estimated in the basis calibration run. These variables' values realize interregional flow estimations and distribution throughput discussed in the previous section (see section 5.2.4 for more details). TABLE 5.7 shows the global model variables estimated in the model.

TABLE 5.7. Estimations of the global model variables in the base calibration

Variable	Value	Comment
B:	-0,050	Gravity model sensitivity parameter
Alpha:	-0,141	Nested level logit sensitivity parameter
NAlpha:	-0,157	Top level logit sensitivity parameter
PDLoad:	9,163	Average production-distribution load, ton. Implied ton-kilometer cost 0,1419 Euro / ton-km
DCLoad:	8,631	Average distribution-consumption load, ton. Implied ton-kilometer cost 0,1509 Euro / ton-km
PCLoad:	8,313	Average production-consumption load, ton. Implied ton-kilometer cost 0,1564 Euro / ton-km
WHTonCost:	9,244	Euro per ton of warehouse / DC throughput

In addition to the global variable on warehouse / distribution per ton costs, the model estimates regional distribution attractiveness variables A_i , which is essentially the extra cost of distribution in region i (note that negative values of A_i make distribution in region i more attractive, reducing the total chain-level cost). The values of the A_i parameter can be interpreted in the following way: based purely on transport costs and global warehousing costs, a positive value of A_i suggests that the region would get more distribution volumes than it actually gets. In other words, from the transport cost point of view, the region is more attractive for distribution in the naïve model than it factually is. A positive value of A_i reduces flows through the region by increasing disutility of the chains linking production and consumption via a warehouse or distribution center in region i . Similarly, a reverse conclusion can be made for the interpretation of the negative values of A_i variable. A negative A_i increases attractiveness of distribution in region i by decreasing disutility of the logistics chains with a facility in the region. The region gets more distribution-related volumes than the transport costs would suggest.

FIGURE 5.11 shows a map of the Netherlands colored according to the region-specific attractiveness values. More dark regions present positive values of A_i parameter, meaning that the relative cost of distribution is increased in those regions. Light colored regions take negative values of the A_i parameter, thus increasing relative attractiveness of the regions for distribution.

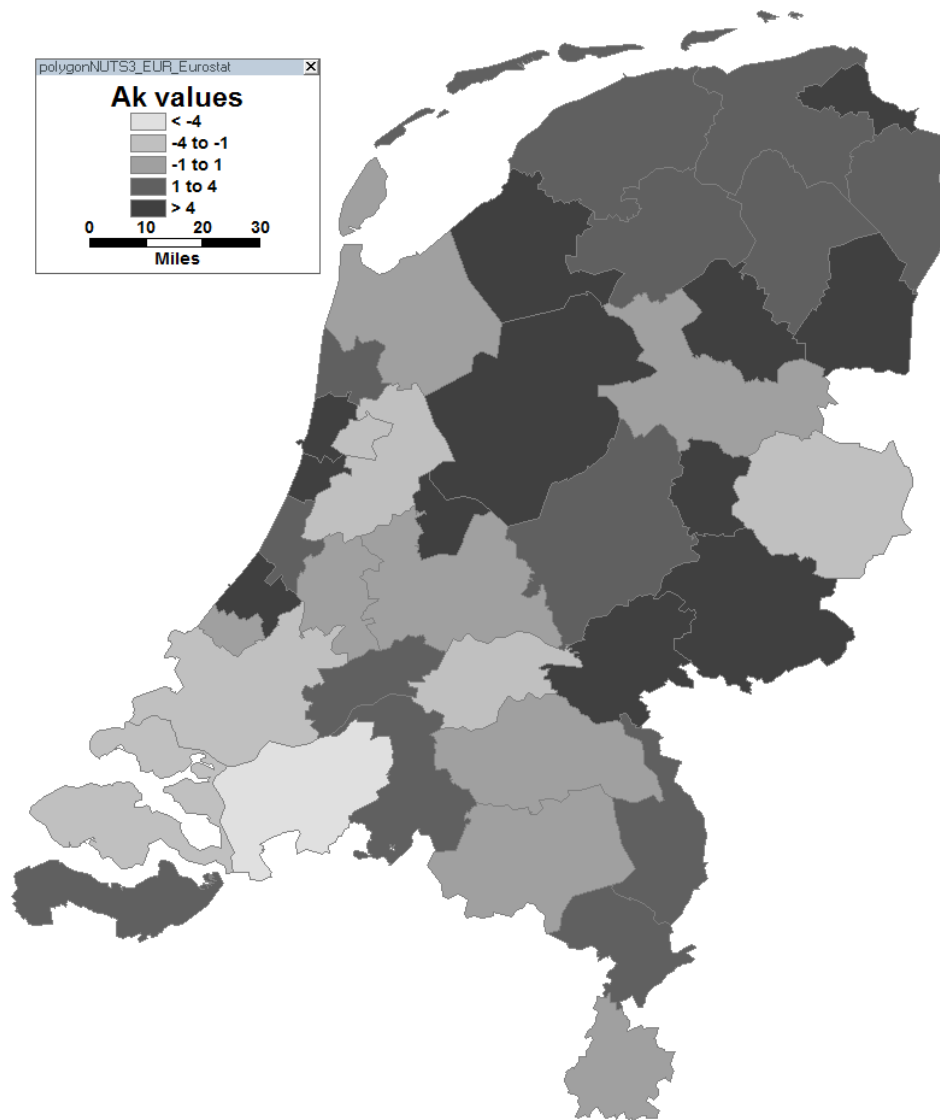


FIGURE 5.11. Regional attractiveness parameters per NUTS 3 region

The estimated average vehicle loads and implied transport costs suggest that production-distribution flows are slightly cheaper than distribution-consumption and direct production-distribution flows expressed in the cost per ton-kilometer shipped measure. However, this difference in vehicle loads and ton-kilometer costs is not large and is within 10%, which is in accordance with the source CBS road transport survey data with respect to vehicle loads.

It should be noted that although production-distribution transport flows are estimated to be slightly cheaper than other transport flow type on the basis of ton-kilometer costs, the overall costs of indirect shipments are larger than the costs of direct shipments. It is mainly due to two factors: first is related to the warehouse and distribution related costs, which add to the transport costs in case of shipments via distribution centers. The second factor is that the total distance via distribution is larger than in case of direct shipments: a shipment takes a detour to a warehouse, which in most cases is not located directly on the route from the production point to the consumption point.

The relatively higher cost of shipments via warehouses or distribution centers is justified by the size of the flows. At the top choice level, namely at the choice of direct-indirect shipment routing, approximately 85% of the ton volumes are shipped directly. Therefore, if we model all commodities together, on average the costs or disutility of the indirect shipments are justifiably higher than those of direct shipments. More detail would be needed to distinguish between commodities as to their cost structure. The available data do not allow model estimation at the commodity level due to the fact that flow matrices are sparsely populated at the level of individual commodities; there are too many zero flow values to allow a reliable model estimation. Therefore, the model is calibrated at the level of total flow, with the flows of all commodities put together.

The attractiveness variable A_i can also be considered as the shadow cost for distribution in region i . It accounts for some unknown factors, which make distribution more or less attractive in the region. For instance, such factors as historical development of the warehousing sector, availability of suitable labor force, accessibility, perception and other factors maybe masked in the shadow cost.

Ideally, the sum of the attractiveness factors A_i should be equal to 0, thus allowing for the estimation of the global warehouse or distribution throughput cost. In practice, it is impossible to achieve a zero sum due to the two reasons. First, the regions are not equal in distribution volumes. For instance, distribution volumes passing the region of West Nord Brabant are more than 45 times larger than those going through Delfzijl. Therefore, a proper normalization of the attractiveness factors would require some weighting of the regions. The second reason is that the values of A_i have an impact on the modeling outcome: model estimation and normalization of A_i values might be an infeasible task. The estimated A_i values are of sufficient quality in respect to normalization such that they do not have a substantial impact on the global warehouse throughput cost estimation. A simple (un-weighted) sum of the A_i values is close to 0.

5.2.6. Alternative logistics chain model implementation and estimation

The essence of the alternative model design is that the base logistics chain model parameters responsible for the vehicle loads for the three transport stages (and corresponding ton-kilometer costs) L^{PD} , L^{DC} and L^{PC} are substituted by the normally distributed shipment sizes, see section 3.7 on alternative definition of the chain model. The shipment sizes are represented by two model parameters S_m (mean shipment size) and S_d (shipment size standard deviation), which are estimated in the model calibration run.

The model's logic remains the same: the physical flows are determined in the nested logit procedure, where at the top level we determine whether direct routes are followed or the shipment will follow a route via a distribution center. At the nested choice, the distribution region is determined. As in the main model case, the same disutility function (total logistics cost) is used.

In the calibration run, the model searches for such values of S_m (mean shipment size) and S_d that the difference between observed and estimated interregional transport flows is at minimum, conform formula 3.31. The model's logic is that if the shipment size is small, then

shipments via distribution centers become more attractive. This is under assumption that production-distribution (PD) flow is realized in a Full Truck Load (FTL) shipment; and the direct production-consumption (PC) and distribution-consumption (DC) shipments are equal to the shipment size. In other words, distribution centers allow aggregation of the inbound shipments to the FTL levels; outbound shipments from distribution centers consist of only one disaggregated shipment without bundling.

In the main model definition, transport ton-kilometer costs are determined by the estimation of the vehicle loads for each of the three types of flow. In the case of alternative definition, transport costs are determined by the shipment size, which is represented by two parameters, shipment size mean and shipment size standard deviation. As these parameters are not known, the model estimates them in calibration run. However, it should be noted that alternative definition reduced the number of model variables responsible for representation of transport cost component of the generalized costs (disutility).

Alternative model implementation is realized “on top” of the main model implementation. In the programming, shipment size is converted into vehicle loads and then the model follows the standard path. Due to the stochastic nature of the shipment size definition, there is no direct conversion of the shipment size into vehicle loads. Instead, shipment size is sliced into intervals and each interval gets its probability (or share in total) according to the values S_m (mean shipment size) and S_d (shipment size standard deviation). Thus, the alternative specification is translated into the main model through discretization.

FIGURES 5.12-5.16 show the quality of the estimated by the alternative model definition of the transport flow fit. The flow fit can be directly compared to the estimations made in the main definition of the LCM (see FIGURES 5.6-5.9). TABLE 5.8 compares fit quality of the main and alternative formulations of the LCM.

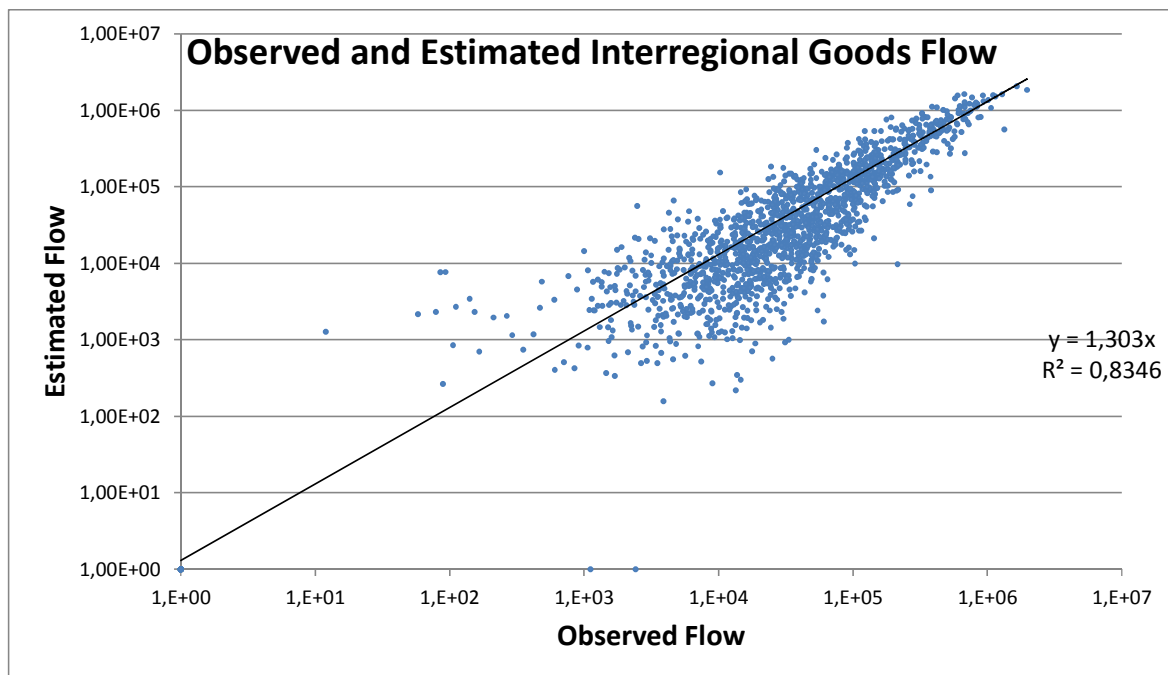


FIGURE 5.12. Alternative LCM model calibration result in respect to total OD flows

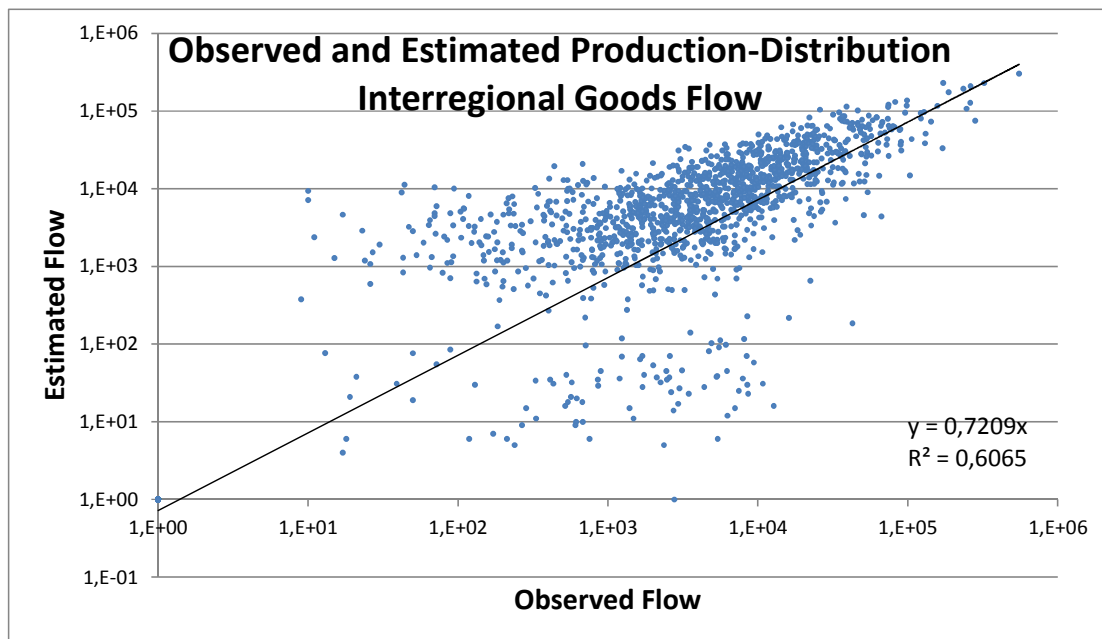


FIGURE 5.13. Alternative LCM model calibration result in respect to PD flow

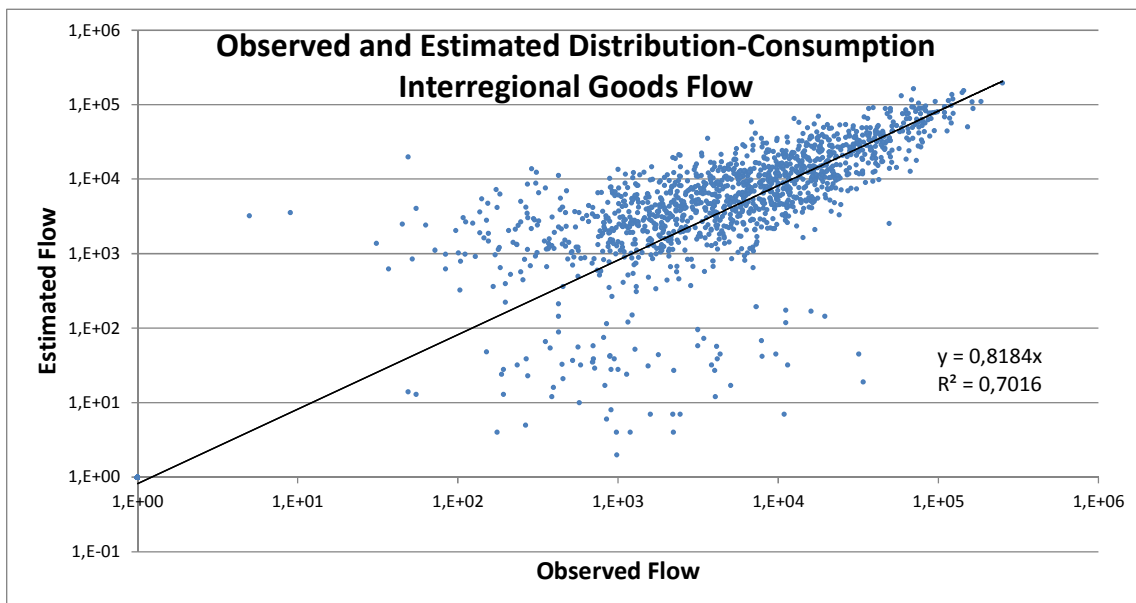


FIGURE 5.14. Alternative LCM model calibration result in respect to DC flow

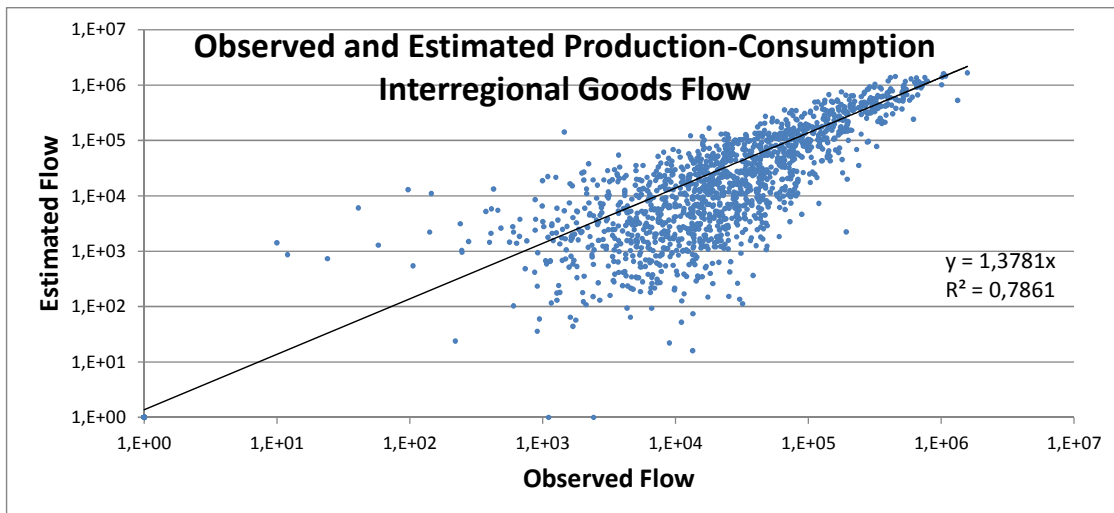


FIGURE 5.15. Alternative LCM model calibration result in respect to PC flow

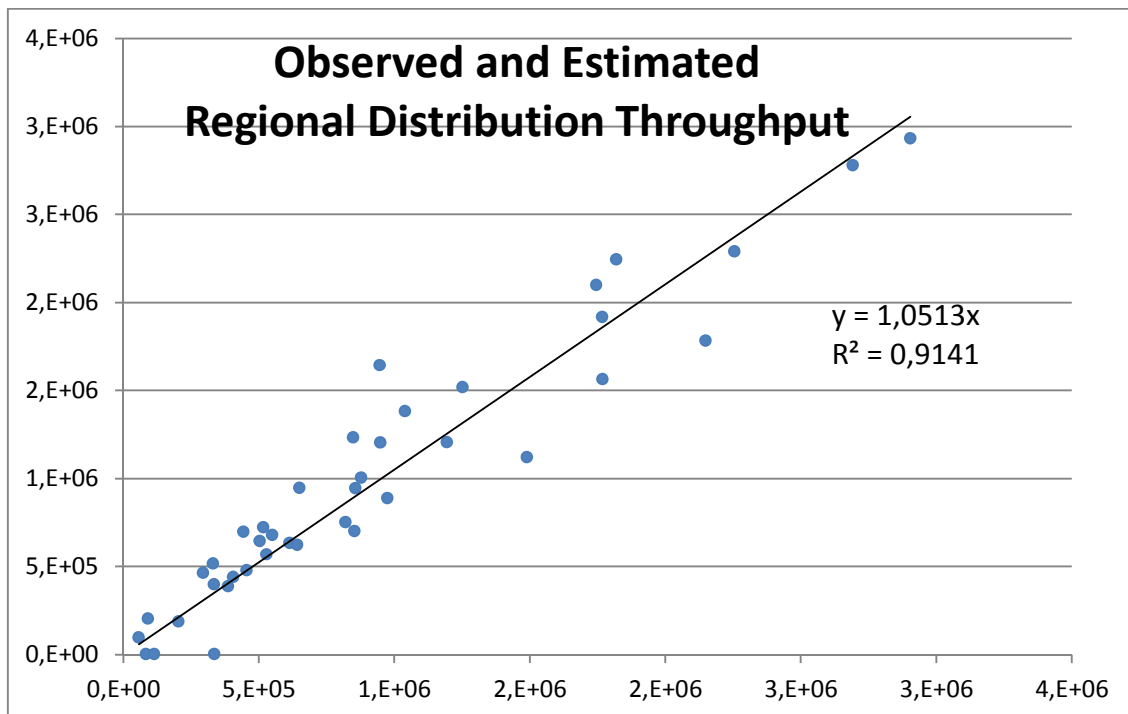


FIGURE 5.16. Alternative LCM model calibration result in respect to implied regional warehouse throughput

TABLE 5.8. Comparison of the flow estimation fit between main and alternative definition of LCM

Flow type / parameter	Main LCM fit (R ²)	Alternative LCM fit (R ²)	Main LCM bias	Alternative LCM bias
PD	0,535	0,603	Y = 0,654x	Y = 0,721x
DC	0,735	0,699	Y = 0,827x	Y = 0,818x
PC	0,519	0,786	Y = 1,042x	Y = 1,378x
OD	0,663	0,835	Y = 1,057x	Y = 1,303x
WH Throughput	0,953	0,914	Y = 1,053x	Y = 1,051x

The alternative definition of LCM performs on par with the main definition of the LCM. On average, it has a slightly better fit of the flow estimation: the fit of the largest flows (PC and OD) is better. The estimation bias is also on par with the main LCM definition, however, for largest flows (PC and OD) is the bias larger than in the case of the main LCM definition. Based on these considerations, it can be concluded that the alternative LCM definition transport flow estimation performance is close to that of the main LCM definition and of sufficient quality.

It is also possible to conclude that these two LCM definitions can be used interchangeably, depending on the emphasis of the modeling effort. In case if estimation or use of transport costs is leading, then the main definition of the LCM is the most suitable. In case if the emphasis is on consolidation function of the distribution centers and warehouses, then the alternative definition of LCM is more applicable, as it estimates shipment sizes. TABLE 5.9. presents the global model parameters estimated in alternative LCM calibration run.

TABLE 5.9. Important global parameters of alternative LCM definition

Variable	Value	Comment
Mu:	24,91	Estimated average shipment size, ton
Sigma:	8,48	Estimated standard deviation of the shipment size, ton
FTL:	15,00	Assumed average Full Truck Load capacity, ton

These parameters should be interpreted as follows. If shipment size is sufficiently large, it is cheaper to transport it directly from the production site to consumption site. Direct shipment is the shortest one and does not incur extra kilometers travelled to the location of the distribution center and costs related to the distribution activities and storage. However, if shipment size is sufficiently small, it becomes costly to transport it directly as the costs of ton-kilometer become large. In this case a route via a DC becomes more attractive: inbound transport costs to the DC are relatively small as the model assumes FTL shipments from production to distribution facilities. In the calibration run, the model has determined that the average shipment size is 24,91 ton (under assumption of 15 ton FTL) and the standard deviation is 8,5 ton. The large average shipment size explains why we observe that only 14% of the flow is realized via distribution centers; the majority of the flow is transported directly.

The average shipment size is a modeling abstraction, which has probably only loose representation in reality. It should also be taken together with the assumed FTL shipment size. However, the average shipment size and its variability provide a good insight in the function of distribution systems. Upstream of DC facilities the transport sizes are large; downstream of DC facilities transport batches are smaller. The alternative definition of the LCM thus empirically supports the notion of the consolidation function of the distribution facilities.

5.2.7. Conclusion on logistics chain model estimation

Section 5.2 has demonstrated that the 4-step freight modeling framework can be extended with a logistics chain model, estimated on real world observations of freight flow. The model has been applied to the interregional goods flow carried out by road transport within the

Netherlands. The input data for the model is based on the annual surveys conducted by Statistics Netherlands. The same data set is used for the model calibration and model validation.

The logistics chain model takes interregional trade flow as the input. In the context of distribution systems, these trade flows should be related to road transport, as distribution in the Netherlands is an (almost) exclusive business of road transport. The search for the good quality trade flow data related to road transport has not been successful; therefore the problem of trade flow data is overcome by estimating of the PC flow data together with the logistics chains. A gravity model has been applied to regional production and consumption volumes in order to estimate the PC flow. The gravity model uses resistance factors applicable for all production-consumption relations from the logistics chain model, thus ensuring consistency between these two types of model.

Main model parameters, such as transport batch sizes for PC, PD and DC flows, regional attractiveness for warehousing and model cost sensitivity parameters, have been estimated on empirical transport flow data. In the model calibration runs the values for these parameters have been found such that the RMSE between estimated and observed interregional goods flows is minimized.

Section 5.2 presented and discussed in detail the model estimation results. For the total OD flows in the main model formulation, R^2 between estimated and observed flows is 0,73. For the constituting PC, PD and DC flows R^2 values are between 0,52 and 0,66. The model also performs well in respect to the estimation of the implied regional warehouse throughput, achieving R^2 measure of 0,95 between the observed and estimated values of the throughput. It is important to notice, that the regional warehouse throughput is estimated in this context by implication: the difference between observed and estimated interregional flows has been explicitly minimized, while the difference in throughput values did not take part explicitly in the calibration process.

The alternative definition of the LCM also performs well with respect to the quality of transport flow estimation. The alternative model definition explicitly uses the assumption that one of the functions of the distribution facilities is the flow consolidation: incoming flow to the facilities has larger vehicle loads than outbound. This assumption is confirmed in the model and average shipment size (together with its variability) is estimated.

The model can be used for policy-related studies, as estimated in calibration run variables can be substituted with other values in order to assess the impact of policy-related measures, changes in cost structures and economic environment. These applications are demonstrated in Chapter 6 of the thesis. It is also worth concluding that future efforts need to be directed at the inclusion in the model of the multi-echelon distribution structures, i.e. logistics chains that include distribution-to-distribution flows.

5.3. European Logistics Chain Model application

Additionally to the Dutch application case, the logistics chain model has been also applied to the European trade flows. The work of the Netherlands Environmental Assessment Agency PBL (www.pbl.nl) has made this modeling application possible. The PBL has estimated interregional trade and transport flows for the EU, see Thissen 2013a and 2013b. The data used in model application at the European level is discussed in section 4.5 on European Data. The content of this section is based on a collaborative contribution between TNO, TU Delft and PBL, presented at TRB, see Davydenko et al. (2014).

There are seven important conceptual differences between the logistics chain model application for the Netherlands and for Europe. The following list summarizes the differences.

1. The PBL data consists of two datasets, describing both trade and transport flows. Therefore, there is no need to estimate trade flows in a gravity model coupled to the logistics chain model.
2. The PBL data is not directly observed data, meaning that some procedures have been applied to obtain it. The construction of the PBL datasets is outside of the scope of this contribution (section 4.5 provides a conceptual overview of the procedure); a detailed description of the procedure can be found in (Thissen et al., 2013a) and (Thissen et al., 2013b). The procedure is also presented in (Davydenko et al., 2014a) for the sake of consistency.
3. The PBL dataset covers whole (continental) Europe, but at the NUTS2 spatial resolution level, thus providing a much larger geographical coverage, but at a less spatially detailed level. For instance, the Netherlands is divided into 12 regions (provincial level), while the Dutch data is at the NUTS3 level, dividing the Netherlands into 40 regions
4. The available PBL datasets do not cover all commodities. These datasets represent trade and transport flows for two distinct types of goods: industrial products and agricultural products. Therefore, in comparison to the Dutch data, the model is applied to a part of the total flow. This makes the modeling more detailed, however less comprehensive.
5. The data provided by PBL does not make any distinction between loading and unloading location types, but ensures that trade flows are harmonized with the transport flows. In other words, if there is a stop between production and consumption locations, this stop may be related to warehousing activities, distribution activities (similar to warehousing, but without an explicit goal on supporting stocks) and cross-dock function (similar to distribution, but without sophisticated sorting processes). Therefore, the notion of a stop for distribution activities in the PBL data is somewhat broader in respect to the location type, but it is narrower in interpretation: the node is explicitly used as a loading / unloading point connecting production and consumption.
6. The total number of regions in the PBL datasets is 256, compared to 40 regions in the Dutch data. The size of the flow matrices is, therefore, a square of this number. For the modeling implementation it does not represent an insurmountable difficulty, however, it increases the number of regional attractiveness parameters in 6,4 times and the size of flow matrices by the factor 41. The computational complexity of the calibration

procedure is, therefore, some 260 times bigger. This is partly compensated by the absence of a need to estimate the trade flows in a gravity model.

7. The PBL datasets come in the form of chains: the production and consumption locations are linked by transport legs via the nodes. Thus, essentially one file contains both the data on trade flows (the ends of the chain) as well as transport flows (each link between production, intermediate node and consumption point). These data have been relatively easy to fit into data structures of the logistics chain model developed for the Netherlands. However, the PBL estimates that there are some deeply echeloned supply chains exist, of up to 5 distribution-distribution nested flow levels. The logistics chain model described in this thesis works only with 1-echelon depth of the distribution, therefore the more than 1-echelon deep structures have not been taken into account in the modeling effort. Indeed, if these structures are really present in practice, they do not have an impact on the modeling of one echelon or direct supply chains, as neither trade flow nor transport flows of deeply echeloned chains are taken into account.

5.3.1. Implementation of the European Logistics Chain Model

The model implementation is very similar to the implementation of the Dutch Logistics Chain Model, which can be thought of as the most sophisticated basis model (see section 5.2 on Dutch Logistics Chain Model implementation). The implementation is conform mathematical definition provided in section 3.6. The main difference from the Dutch model is that trade flow estimations are included into the PBL datasets, thus there is no need to estimate them. The gravity model functionality is, therefore, not included into the European model.

FIGURE 5.17 shows the conceptual representation of the European logistics chain model. It takes trade flow (P/C) matrix as the input, splits trade flows into direct shipments and shipments via the distribution structures. In case of shipments via distribution structures, the model determines location of the distribution. The output is in the form of three transport flow matrices, production-distribution, distribution-consumption and production-consumption. The sum of these flow matrices gives the total transport OD flow.

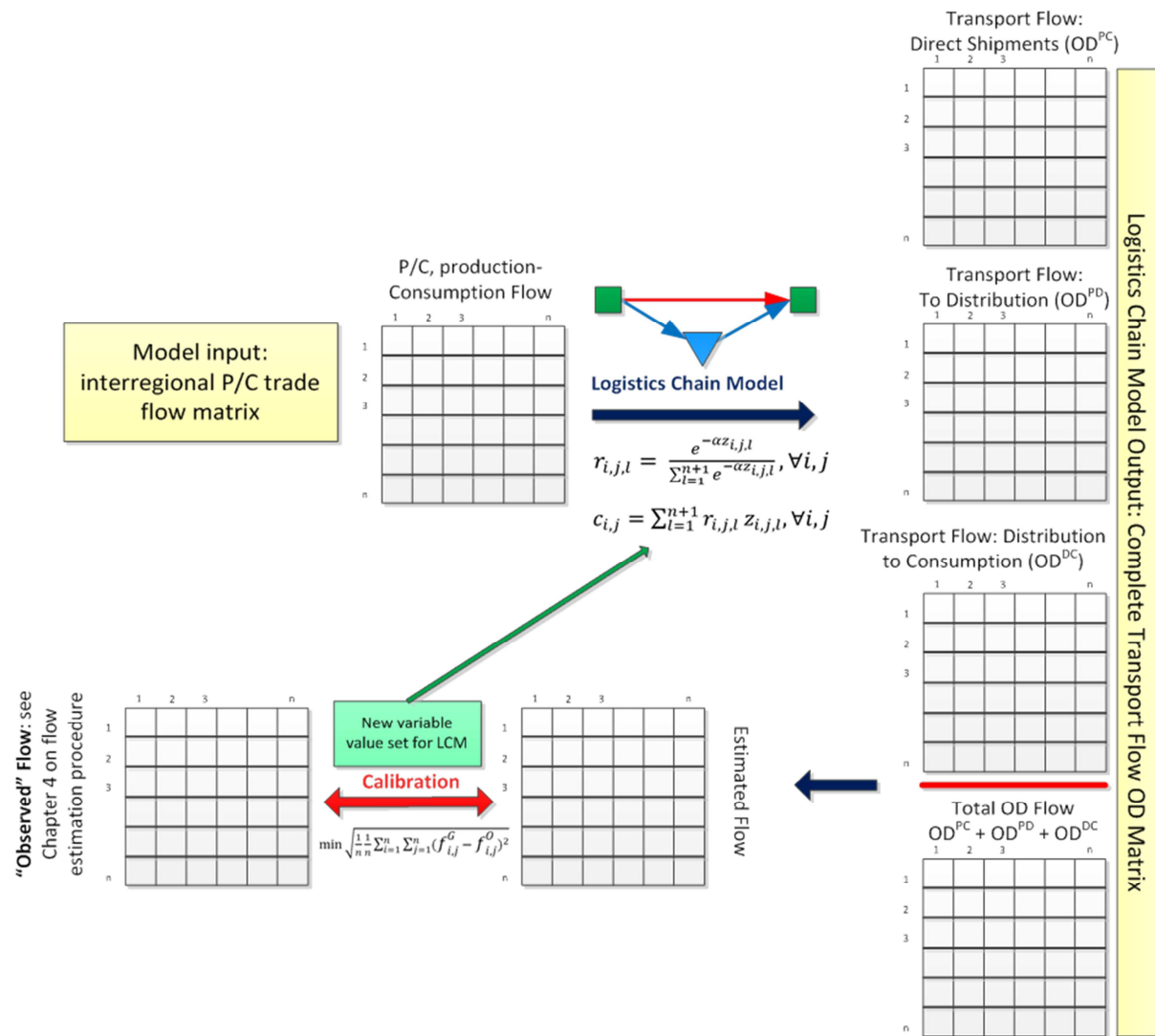


FIGURE 5.17. Conceptual representation of the European and German logistics chain models

The calibration procedure searches for the model parameters' values such that equation (5.2) is satisfied in respect to minimization of the root mean square error between observed (estimated by PBL in this context) interregional transport flow values and estimated by the LCM model. It has been observed that the estimated model variables stabilize after 5-6 iterations of the grid search procedure runs.

At the European level, the ratio between the ton volume of direct shipments and shipments via the distribution structures is different than it is the case in the Netherlands. At the European level, the relative volume of direct shipments is smaller than the volume of shipments via the distribution structures. This observation confirms the assumption that shipment consolidation becomes more important on larger distances, where transport costs play a more important role. This fact has also had an impact on the weights in the formula 5.2: in contrast to the Dutch model calibration procedure, where more weight has been given to shipments via distribution (constants $a, b > c$), in the European model more weight has been assigned to direct shipments (constants $a, b < c$), as the volume of direct shipments is smaller. This holds true for data on both commodity types that have been provided by the PBL, the agricultural and industrial products.

5.3.2. European Logistics Chain Model calibration

The European Logistics Chain Model has been estimated and calibrated for two separate cases. The first case is based on the Industrial Produce dataset and the second case is based on the Agricultural Produce dataset. Both source datasets have the same structure: they have been split into two flat databases (files) with the following structure.

1. *Direct flow database.* This file contains information on direct shipments, i.e. the shipments where trade flow is equal to the transport flow in terms of regions where goods are loaded to the road vehicles and unloaded from them.
 - a. Production region code. This is a unique code identifying a region at the NUTS2 level. There are 256 regions represented in the dataset.
 - b. Consumption region code, in the same classification as the production region code
 - c. Volume in ton. This is the volume shipped between producing region and consuming region in the reporting year.
2. *Flow via distribution.* This file contains information on indirect shipments, i.e. the shipments where trade flow is not equal to the transport flow. The trade flow is physically realized by the means of two transport legs: the first leg is from the production location to the distribution location and the second leg is from the distribution location to the consumption location. The dataset contains the following fields and uses the same coding for the geographic NUTS2 regions.
 - a. Production region code
 - b. Distribution region code
 - c. Consumption region code
 - d. Volume in ton, which is shipped annually via the corresponding production-distribution-consumption chain.

Therefore, the dataset contains complete information on the trade flows and transport flows, including direct shipments and shipments via distribution structures. The dataset also contains a distance matrix, matching origin and destination regions in a table 256x256. The distance matrix is necessary for the calculation of road transport costs, which are proportional to the distance between loading and unloading locations.

The calibration results of the European Logistics Chain Model are presented in the same manner as the calibration results for the Netherlands: on the X axis the “observed” values are plotted and on the Y axis are the estimated in the model flow values are plotted. However, there is a conceptual difference between the Dutch LCM and the European LCM model calibration results. In the case of the Dutch model calibration, the estimated flow values are compared to the observed flow values, which come from the CBS survey results. In the case of the European model calibration, the estimated in the model flow values are compared to the values provided by the PBL. Therefore, for the sake of consistency we will still call the estimated by PBL flow values as observed flow (in respect to the logistics chain model), however, the term “control” values might be more appropriate in this case.

5.3.3. European model calibration

This section provides details on the model calibration quality for the European flows. FIGURE 5.18 matches estimated by the European Logistics Chain Model transport flow values and the observed (control) flow values. The section presents the quality of fit for the total flows only, as flow matrices are rather large (256x256 elements), which makes figure presentations somewhat awkward.

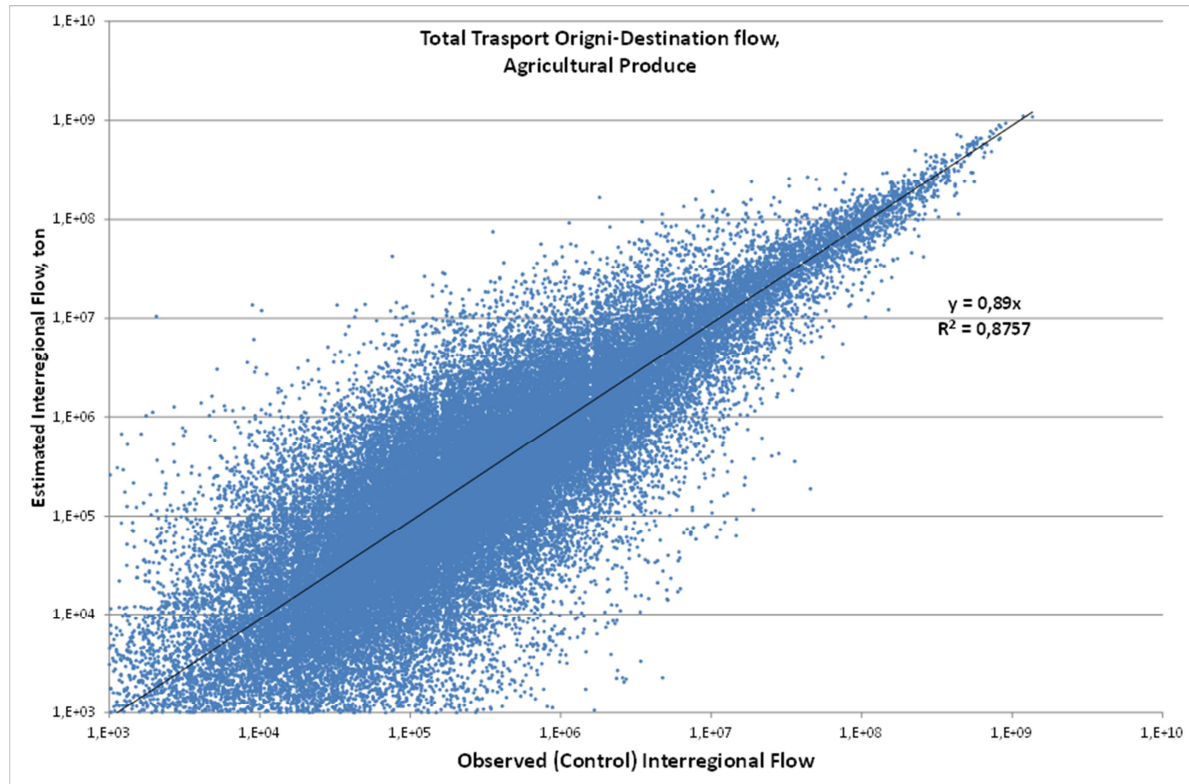


FIGURE 5.18. European LCM model calibration result with respect to PD flows (agricultural products)

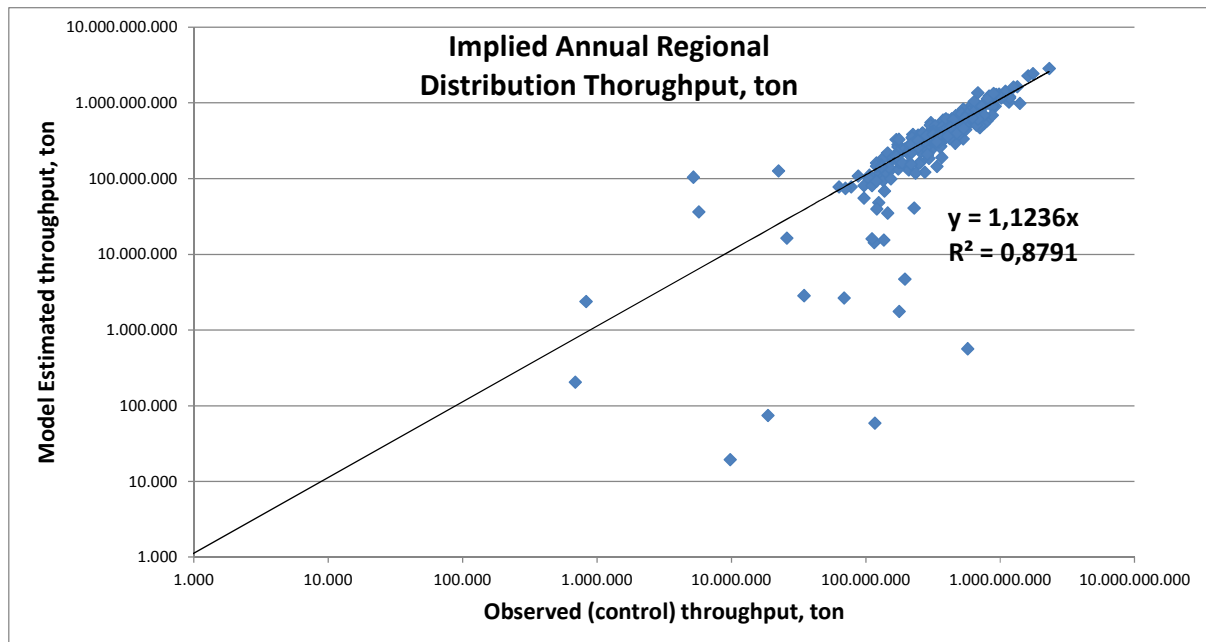


FIGURE 5.19. European LCM model calibration result in respect to implied regional distribution volumes (agricultural products)

Application of the European Logistics Chain Model for the trade flows in Agricultural Produce and Industrial Produce commodities provides good model estimation results, see TABLE 5.10 for more details. The R^2 measure for the flow estimation quality is between 0,78 and 0,90 for the constituting sub-flows of the agriculture commodity and between 0,72 and 0,86 for the industrial commodity. The implied regional distribution volumes are also very close to the control values. This result is considerably better than the flow fits for the Dutch model application. The model estimation has the same quality pattern for both commodities with a minor difference for the distribution-consumption flows (R^2 measure for industrial products is 0,72 and for agriculture is 0,78). This is probably due to the fact that the agricultural products have a lesser value density than the industrial products and, therefore, more expensive to transport in terms of the share of transport costs in the total product cost. This fact implies that there is a more clear distance decay pattern in case of agricultural goods. The implied regional distribution throughput is also very close to the observed (control) values (FIGURE 5.19). We do not have an explanation for the underestimation of small flows, as these concern the region of Valencia in Spain and two Greek regions (the central and Epirus).

TABLE 5.10. European LCM flow estimation results

Flow / parameter	R ²	Bias
Agriculture PD	0,83	Y=0,95X
Agriculture DC	0,78	Y=0,80X
Agriculture PC	0,90	Y=0,92X
Agriculture OD	0,88	Y=0,89X
Agriculture Implied DC throughput	0,88	Y=1,12X
Industrial PD	0,82	Y=0,85X
Industrial DC	0,72	Y=0,75X
Industrial PC	0,88	Y=0,88X
Industrial OD	0,86	Y=0,84X
Industrial Implied DC throughput	0,90	Y=1,05X

TABLE 5.11 shows model variable values that have been estimated in the calibration run for both agriculture and industrial commodities. The industrial products show a clearer differentiation in average vehicle loads per sub-flow than the agricultural products. While the average load for the PD flow is almost the same in both cases, the DC loads are different (6,325 versus 8,318 ton) and PC loads differ substantially as well (5,923 versus 7,031 ton).

TABLE 5.11. Estimated model variables in the calibration of the European LCM on Agricultural and Industrial Produce data

Variable	Estimated variable value (calibration)
Agriculture PD load, ton	9,529
Agriculture DC load, ton	8,318
Agriculture PC load, ton	7,031
Agriculture PD cost, Euro cent / ton-km	13,64
Agriculture DC cost, Euro cent / ton-km	15,63
Agriculture PC cost, Euro cent / ton-km	18,49
Agriculture Warehouse throughput cost, Euro / ton	2,547
Industrial PD load, ton	9,491
Industrial DC load, ton	6,325
Industrial PC load, ton	5,923
Industrial PD cost, Euro cent / ton-km	13,70
Industrial DC cost, Euro cent / ton-km	20,56
Industrial PC cost, Euro cent / ton-km	21,95
Industrial Warehouse throughput cost, Euro / ton	3,619

The difference in the average transport batch sizes and implied transport costs suggests that distribution in industrial products has a more profound function of storage and order decoupling than it is the case for agricultural products. The businesses in industrial products try to save costs on line haul, and then send smaller batches from distribution centers to the end customers. While agricultural commodities transported in big batches to the distribution, the direct flow and the flow from distribution is realized in a relatively bigger batches than for

the industrial goods, thus suggesting that the function of distribution is less for storage and probably more to the cross-dock function of operations. Still, a relatively low value density should also suggest the higher share of transport costs in the cost of the end product, thus serving as a stimulus for a better utilization rate of the vehicles. The patterns with respect to average vehicle loads and transport costs are consistent for these two commodities.

The costs related to the distribution processes are also higher in case of the industrial products, however it should be noted that for both commodities the distribution-related costs are relatively low compared to transport costs, especially taking into account that the distances are large in the European model application. At the European level, the delivery lead time (i.e. costs vs service level) can play a bigger role due to larger distances.

5.3.4. Conclusions on the European Logistics Chain Model

This section has provided a definition and implementation results of the European Logistics Chain Model. A unique dataset constructed by the Dutch Environmental Agency (BPL) has made this model application possible. Functionally, the PBL European data is better than the Dutch dataset, because it contains both transport flow data and trade flow data, which are both necessary for the estimation of the LCM. There is also an important aspect of the PBL data: it is internally consistent in respect to distribution-related flows, meaning that the inflow into regional distribution systems is equal to the outflow.

The ELCM model estimated on the European trade and transport flows performs better than the Dutch application of the model in respect to estimation quality, showing higher values of the R^2 measure between the estimated by the model flows and observed flows (in this case better described as control flows). There are a number of reasons that probably influence a better estimation quality of the European model.

First is that it has been applied at the NUTS2 level, in comparison to the Dutch application case at the NUTS3 level. A more coarse spatial resolution of the model conceals goods flows over small distances, which are large in volume and difficult to model with a cost function proportional to the distance (distance between two adjacent regions is a very rough estimation in itself; the costs of short distance transport movements are less influenced by the distance itself, but by the (overhead) time spent loading, unloading and repositioning of the vehicle).

The second probable reason of the better estimation quality is that trade flows have not been estimated in a simple gravity model, but come from a novel estimation method. The trade flow statistics have been indeed worked upon by the PBL to make them consistent with transport flow statistics. The process did not involve any modeling based on the cost functions, such as for instance gravity model. It is very probable that the trade flow estimations are of a better quality than it is the case for the Dutch model application.

The third reason is that the distances in the area under consideration (continental Europe) are much larger than in the Netherlands and, therefore, transport flows are more suitable for the estimation using transport cost functions proportional to the distance travelled.

The fourth reason is that flows via distribution centers are not based on the location type variable, but come from a modeling effort applied to the transport statistics. This makes data internally consistent with respect to equality of inflow and outflow for the regional distribution systems (the “niceness” of the data does not, however, imply that it is more empirically accurate). The transport flows estimated by the LCM are also consistent with respect to inflows and outflows, thus the estimations provided by the PBL do not introduce an imbalance in respect to inequality of distribution inflow and outflow, which minimizes the difference between LCM estimated flows and control flows.

The European Logistics Chain Model estimation is computationally challenging, although not so much in terms of computational complexity. The model parameters estimation is done in polynomial time: computational time grows $O(n^3)$ with the number of regions under consideration in the model. In the European case, the flow matrix size is 256x256 and the model contains 261 parameters, which need to be estimated. For a given fixed set of parameter values, the flow estimation takes less than a second to compute on mainstream PC, so practical ELCM applications are not computationally challenging. However, estimation of all 263 model parameters is much more computationally demanding than it is the case for the Dutch model: a 560% increase in a number of model parameters and a 4100% increase in the flow matrix size compared to the Dutch model. Still, the model parameters have been estimated using the same grid search procedure within reasonable time (36 hours). The absence of the need to estimate the gravity model has also contributed to the performance of the ELCM model estimation.

At the European level there is a stronger need to design a logistics model that incorporates multi echelon nature of some European supply chains. This model uses only one distribution echelon: the flow originating from the distribution location can only be transported to the consumption location. Both European and Dutch data suggest the presence of multi echelon distribution structures, which is especially valid for spatially extended supply chains such as those at the European level. In practical terms it means that in addition to the PD, DC and PC transport sub-flows, a future LCM model should be able to model DD flows, namely flows originating at distribution locations and terminating at the distribution locations as well. Such a model will be more computationally complex, but it will better capture the richness of real world complex logistics chains.

5.4. LCM Implementation and estimation on German Data

The LCM model has been applied to the German data, which represents trade and transport flows of the German food retail sector. The data is at the NUTS2 spatial resolution level (see more details on how the data is obtained in section 4.4) and does not distinguish between commodity types. The data distinguishes between 41 regions in Germany, provides a distance matrix for the estimation of transport costs, and importantly, accounts for trade flows in the form of a PC flow matrix between the sourcing locations and the Point of Sale (POS) locations. The transport flow data are split into the three transport flow categories: PC, PD and DC.

The data are similar in their nature to the European Data: they contain information on both trade and transport flows. Therefore, there is also no need to estimate trade flows, as it is the case in the Dutch model; the model needs only the LCM part. Therefore, FIGURE 5.17 accurately represents the model design and data flows within the model. Thus, the German LCM application is “easier” in computational terms than the Dutch model as there is no need to estimate trade flow in the gravity part of the model; it is also computationally easier than the European model application, because the flow matrices have dimension 41x41 regions, compared to the 256x256 size of the European data.

5.4.1. LCM calibration on German Data

The calibration of the LCM on German data has been carried out in the same way as the LCM Dutch and European applications: a grid search procedure has been used for estimation of the model parameters. The values of parameters have stabilized after 5 iterations; the resulting parameters values do not depend on the order of parameter estimation in the grid search procedure, suggesting that the model is not trapped in a local minimum. FIGUREs 5.20-5.23 show estimation fit quality of the German LCM model.

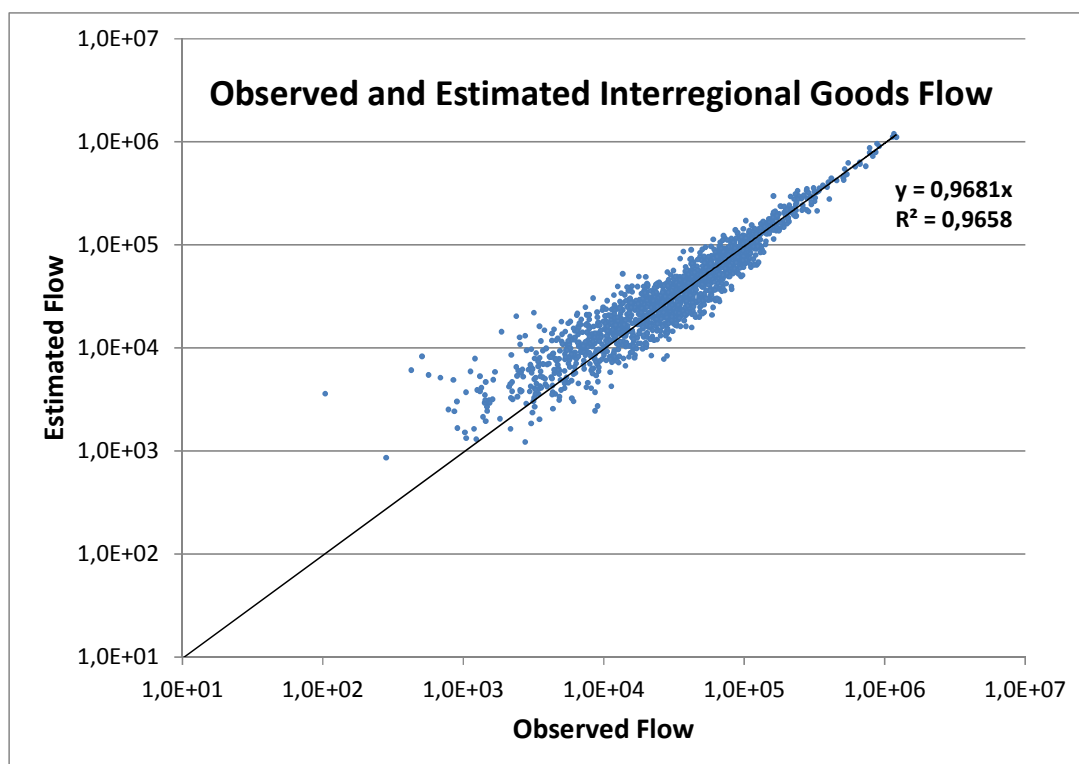


FIGURE 5.20. Estimation fit of the total OD transport flows in German LCM

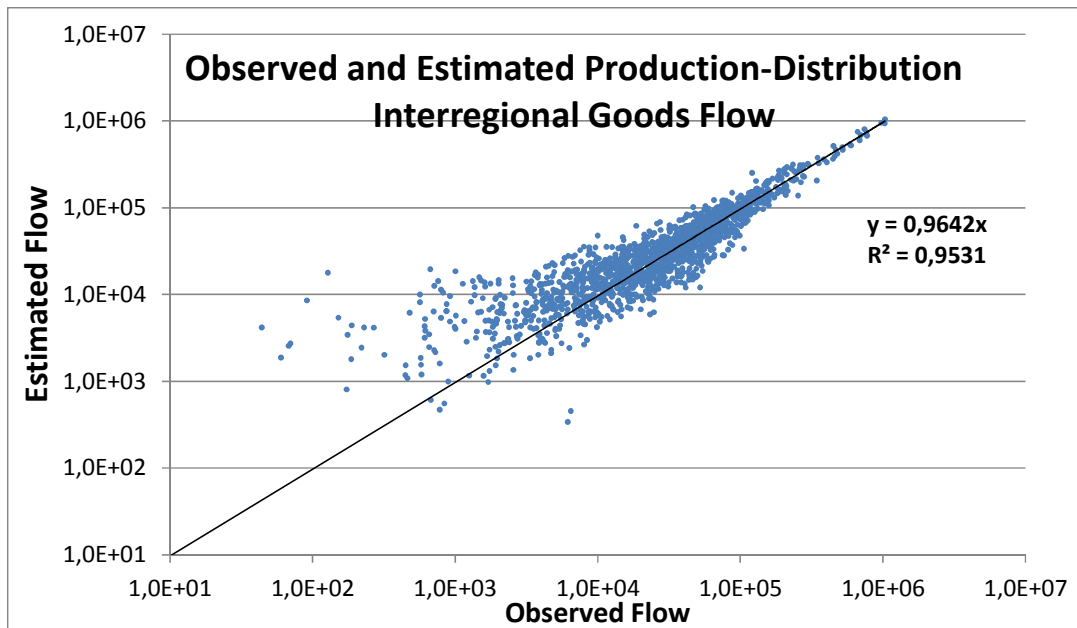


FIGURE 5.21. Estimation fit of the PD transport flows in German LCM

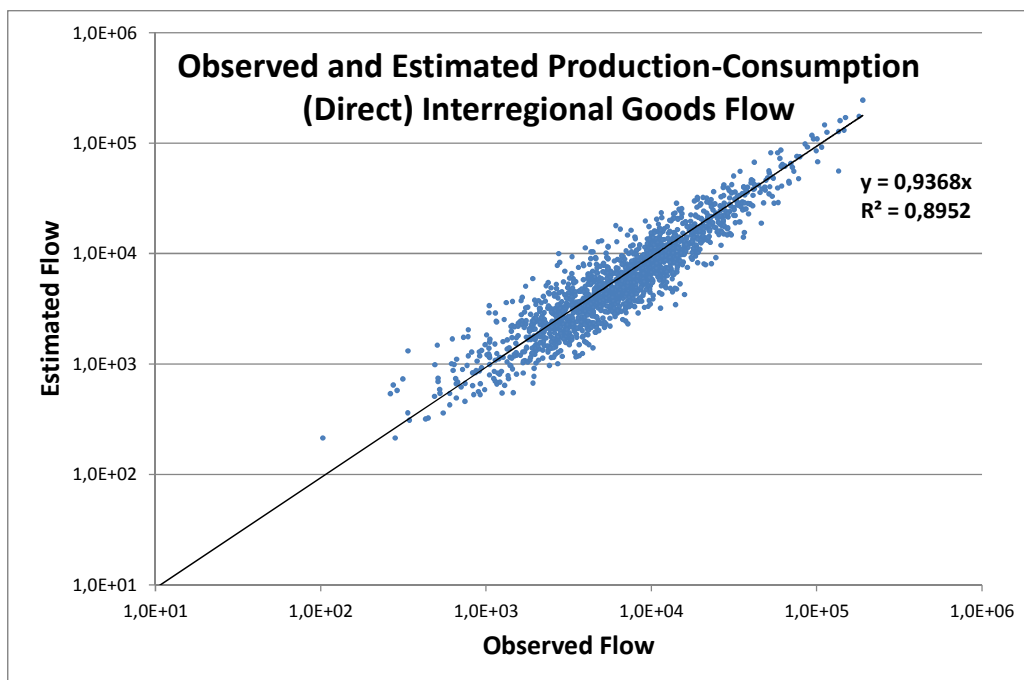


FIGURE 5.22. Estimation fit of the total PC transport flows in German LCM

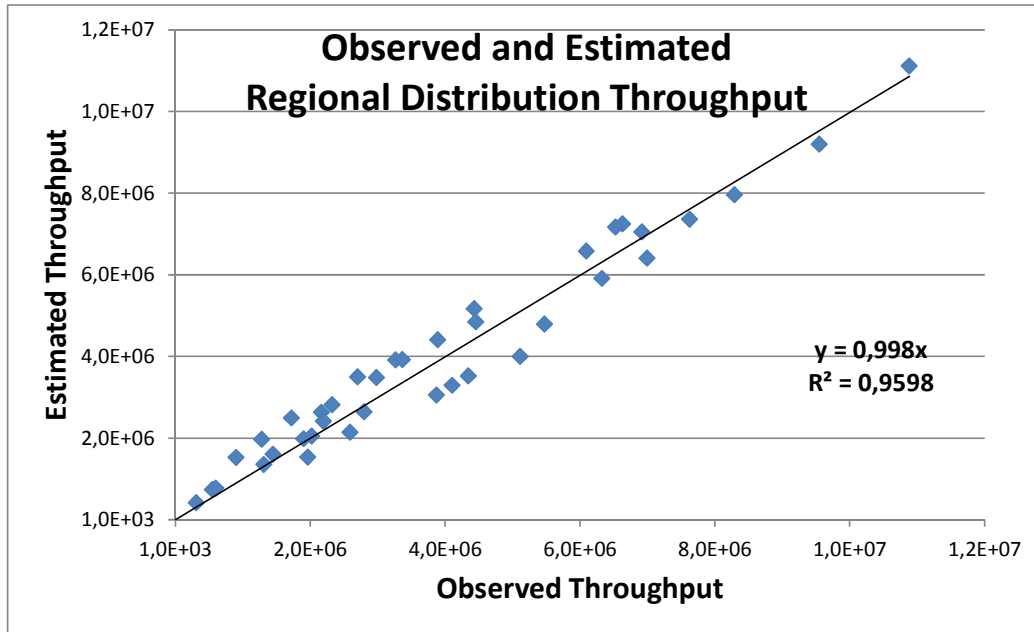


FIGURE 5.23. Real and implied by LCM flows regional distribution throughput fit

The LCM model reproduces flows estimated by the SYNTRADE model rather very well, at least in comparison to the fit quality of the Dutch and the European model. The model has also made plausible estimations of the transport costs and load factors, which are presented in TABLE 5.12.

TABLE 5.12. Transport loads and transport costs estimated in German LCM application

Flow Type	Average vehicle load, ton	Average transport cost, EUR cent/tkm
PD	10,283	12,64
DC	5,450	23,85
PC	7,867	16,53

The vehicle loads estimated by the LCM model imply that transport to the distribution centers (inbound flow to the warehouses) is carried out as fully loaded vehicles as possible. A load of more than 10 on average can be considered as a Full Truck Load (FTL). The flow from warehouses to the consumption shows a higher degree of disaggregation: frequency of delivery has more importance than transport costs on shorter transport legs, especially taking into account the nature of products, many of which are perishables. The direct shipment loads are almost an average between the inbound and outbound loads of the distribution facilities. If a shipment is sufficiently large, it makes sense to send it directly in order to avoid the extra costs of the distribution facility and network detour to the facility (i.e. extra ton-kilometers to visit the warehouse). Also service level considerations play a role in the choice of the path: a high service level requirement may lead to the choice of a more expensive routing.

5.4.2. Conclusions for LCM application on German Data

The LCM application on the German data has two specific properties compared to the Dutch and European LCM applications. First, the quality of fit is better than in two other LCM applications. Second, the estimated vehicle loads and transport costs per ton are “logical” and better follow the theoretical explanation of the warehouse and distribution function.

The good fit and estimations of the average vehicle loads may be partly attributed to the fact that the SYNTRADE model, which generated transport flows via the logistics structures (these transport flows have been used as control or “observed” flows in model calibration), follows a certain logistics optimization logic in determining of the transport flows. The fact that the SYNTRADE model uses micro level optimization logic in its procedures is captured by the macro model. Apparently, the same modeling outcome can be reproduced by the aggregate choice model that determines the flows based on disutility of the choices. The application of both micro level SYNTRADE model and macro level logistics chain model leading to a similar outcome can be seen as the bridge between micro and macro levels of freight transport and logistics modeling, which is not a trivial result.

The good fit of the transport flows by the LCM raises the question on the right choice of modeling techniques. The SYNTRADE model works at the micro level, simulating and optimizing decisions of actors at the company level. The SYNTRADE approach requires substantial data at the micro level to be successfully implemented. It is no coincidence that only the food retail sector is considered in that model: a broader model application is too costly as it would require manual collection of the company- and sector- related data. On the other hand, the LCM has shown that it can reproduce the flows estimated by the SYNTRADE model with the great accuracy. This reasoning naturally leads to the question on if we indeed need sophisticated transport models that incorporate detailed data at micro level to estimate logistics chains including storage and distribution, if a simple macro model can reproduce these flows very well? In other words, the complex logic of the micro model has been reproduced with the relatively simple LCM.

The answer to this question is not within the reach of this research effort. First, there is only one application case of the LCM, which compared its output with the output of the micro-level model SYNTRADE. In other cases the outcome can be different. Second, without output data of SYNTRADE, the calibration of the LCM would not be possible, meaning that the transport flows are needed for the estimation of regional attractiveness, sensitivity and vehicle loads in the LCM. Nonetheless, it is safe to draw the conclusion that both approaches are interesting and viable. The micro approach allows scenario applications beyond the reach of the macro models; the macro models are less data hungry, give a broader scope and can be constructed faster using smaller amounts of data.

5.5. Discussion and conclusions

In this section four freight modeling applications have been considered. First, the separately calibrated gravity model has been applied to the transport O/D flow matrices. Second, it has been demonstrated that the logistics chain model can be empirically proven on the Dutch road

transport data. Third, the LCM has been shown to perform well at the European level of freight systems. And finally fourth, the LCM showed a good performance in estimation of the transport flows and distribution volumes at the level of a single sector, German food retailing.

The main conclusion from the SCGM applications at the level of transport origin-destination flows is that different components of the total transport OD flows show different properties with respect to regional freight generation and attraction, as well as sensitivity towards price signals. In other words, the studied sub-flows PD, DC and PC of the total OD flow are different in their nature and properties. Up to now the dominant freight modeling practice did not treat these flow components separately; in many instances trade flow has been assumed to be equal to the transport OD flows, possibly increased by a certain factor to reconcile detours related to distribution (e.g. transport ton volumes are generally larger than trade ton volumes). The SCGM applications show that such practices, however satisfactory in some instances, lead to a structural error in flow estimations, as the components of the total OD flow are structurally different. Thus, SCGM applications present an empirical case for the idea that a logistics model should be incorporated into the classical 4-step modeling framework.

The modelling applications of the Dutch road transport data have shown that the combined gravity-logistics chain model can reproduce flows resulting from regional distribution systems, including transport O/D flows at the level of total transport, as well as at the level of the PD, DC and PC components of the total flow. The problem of the absence of trade flow data is overcome in a practical manner, using a gravity model that matches computed on the transport data regional production and consumption volumes into a trade flow matrix. The model further estimates the average ton vehicle loads and ton-kilometer transport costs per flow type.

It has also been shown that two formulations of the LCM are possible: one where transport costs are the direct decision variables and another one is the formulation where the shipment size is the direct decision variable. Both formulations lead to a generalized logistics cost as the determining variable in the discrete choice, however, these two modeling approaches have a different perspective. The main model formulation is driven by transport costs, which are estimated in the model and present a realistic estimation of transport costs. The main model formulation is better positioned for the scenario-wise analysis of complex transport systems, such as impact of a change in transport costs and the costs of distribution facilities. The alternative formulation, which is driven by shipment size, is more fictive in the sense of shipment size estimation, as there are currently no empirical data available to confirm the correctness of the shipment sizes. However, it allows better capturing the logic of warehousing and distribution systems: namely storage and aggregation functions, concerning our hypothesis that inbound warehouse flows tend to use FTL shipments, and outbound warehouse flow represents the shipments as they are ordered by the consumption points.

Both LCM formulations and applications for the Netherlands show sufficient empirical validity of the modeling, as it has been extensively shown in the fit analysis. Model application on the real survey data, with its intrinsic sampling errors, can never achieve a 100% fit. Still, the fits of the generated flows by the model and presented in section 5.2 present a firm conclusion on empirical validity of the Dutch LCM.

The application case on the European data has shown that the LCM can also perform well on a much larger geographical scope. There are two main differences between the Dutch and EU application cases. First is the computational complexity: the Dutch model deals with the flow matrices 40x40, and the European one works with the 256x256 dimensions. The difference is the number of regions also implies the number of parameters to estimate in the calibration runs: 47 in the Dutch case and around 260 in the EU case. Due to the fact that the trade flows are available for the European case, there is no need to couple the gravity model to it. As the result, the European case has shown that the model can be calibrated within reasonable time: it took around 36 hours of running for the grid search procedure to find stable values of the model parameters.

The second difference is that the EU model covers larger distances, where different logistics properties are at work. Indeed, at the European level much larger share of the flow is transported via distribution structures. At the long distance transport, the transport costs dominate over distribution costs. The LCM model captures this fact and generates the flows that show a good fit with the control flows. Moreover, the European case shows that the model is ready for answering policy-related questions on behalf of the European policy makers.

The application case for the German food retail sector data has also shown that the LCM model can be applied not only at the level of the whole economy, but at the sectorial level as well. The most interesting result of the German case is that both micro-level SYNTRADE model and macro level LCM model can generate the same transport flows for a given trade flow. This fact fuels the discussion on what approach for incorporation of the logistics model into transport modeling framework is the best one: aggregate or disaggregate.

Finally, the main conclusion from the four application cases is that an extension of the 4-step modeling framework with a logistics chain model presents an empirically proven improvement of the classical freight modelling approach. The logistics chain model presents an intrinsic value on its own, as it can be used for analysis of various policy-related questions, as the following Chapter 6 will show. The LCM can also form an important part of the industrial-scale models such as the TRANS-TOOLS model, ensuring that the logistics systems are properly represented in these comprehensive models.

6. Logistics Chain Model Applications

This chapter describes the Logistics Chain Model (LCM) application areas and application cases. Section 6.1 discusses possible application areas, provides detailed information on what can be influenced in the model for the purpose of scenario analysis and what useful indicators can be constructed based on the model output. Section 6.2 presents an application case related to a study of logistics sprawl in the Netherlands. It is based on the work presented at the METRANS I-NUF congress (Davydenko 2013) and sheds light on what incentives are necessary to increase or decrease concentration of logistics facilities in the Randstad region of the Netherlands. Section 6.3 discusses price elasticities derived in the logistics sprawl case and compares them to those estimated in the separately calibrated gravity model (see section 5.1 for more details on the gravity model implementation).

6.1. Introduction

The LCM can be applied for scenario analysis, exploring freight systems and the place of logistics facilities in them. To perform scenario-based analysis with the LCM, the model variables need to be understood and a set of simple and useful indicators for the modeling outcome has to be defined.

6.1.1. Main areas of LCM application (spatial logistics, large-scale models)

The LCM model has two broad application areas. The first application area is the possibility to embed the model into comprehensive modeling suits such as the TRANS-TOOLS model (Chen 2011). The LCM model would substantially improve the modeling of translation of trade flows into transport flows. The incorporation of the LCM into more comprehensive modeling suites is context-dependent, i.e. depends the model it is embedded into and the goals that are being pursued. The second application of the LCM model is to use it as a standalone research tool for scenario analysis. This chapter describes standalone LCM model applications.

The LCM model is a macro model. Its applications are at the level of interregional freight flows and regional distribution volumes. This macro nature of the model makes its applications suitable for the use by various government agencies at different levels. For instance, the model can be used for the analysis of the impact of changing transport prices on the interregional flows. The change in transport price can be “natural” due to variation in fuel price and labor costs, as well as through government imposed taxes, tolls and subsidies. The model can also be used in spatial policies, as the impact on spatial patterns of distribution systems can be determined by the model.

The LCM applications for individual companies are also possible. These applications can be related to property markets, which may be relevant for the scenarios on logistics facility asset price development. Another application area is understanding of the logistics chains leading to or departing from the facilities, such as ports, terminals and business parks. The analysis of the chains would help better positioning and marketing of the facilities, as the model may show the types of flow most suitable for the facilities under consideration.

6.1.2. Influence of the decision variables

The combined gravity and logistics chain model is controlled by a number of variables (see formulae 3.23 and 3.24 on the formal specification of the Total Logistics Cos (TLC)). These variables can be split into three classes. The first class concerns logistics costs; the second class concerns price sensitivity of the model; and the third class concerns regional attractiveness. The values of these variables are determined during model the calibration procedure. The model is linked to the real world cost values through a constant vehicle-kilometer cost and a normalized value of the average PD load and corresponding ton-kilometer transport costs. The calibration procedure finds specific values of the model variables so that the transport flows estimated by the model best represent the same flows, as observed in reality.

The logistics cost variables consist of the two components. The first component is the ton-kilometer transport costs for the three different transport flows: PD, DC and PC. These transport costs are defined as vehicle-kilometer cost divided by the ton load carried by the vehicle. In this way, the transport cost is linked to the average loads transported and depends on the flow type. The second component is the warehouse throughput cost. This cost is related to vehicle unloading at the distribution location, handling and storage costs at the distribution, and the costs related to the loading of the departing vehicle at the distribution.

The logistics costs have three levels of influence on the transport flow estimated by the model. First, logistics costs in the form of TLC influence gravity model outcome. Second, logistics costs influence decisions on whether the goods are delivered directly or whether a chain with a distribution center is chosen. Third, for the chains that include distribution, logistics cost influences the flow patterns over alternative distribution centers. The following list presents more details on how the TLC influences at these three levels. In more detail:

1. Gravity model level. The gravity model matches two single-dimension vectors of production and consumption into a two-dimensional trade flow matrix. The

probability that two regions will trade with each other is influenced by the resistance that is met when goods are shipped between them. The TLC is the resistance: for a fixed set of the price sensitivity values, a higher TLC would lead to a trade flows estimation that favor shorter distance production-consumption relations than it would be the case for a smaller TLC. Therefore, any increase in the TLC, be it transport costs of any flow type or warehousing costs, will lead to a shift in spatial distribution of trade in the direction of shorter distances between the trading regions.

2. Top level choice in the LCM model is whether the flow is direct or follows a distribution structure. The components of TLC are influencing this choice in different ways. An increase in PD, DC and distribution costs would lead to a larger share of flow going direct; an increase in PC transport costs would lead to a smaller share of direct shipments. The reverse is also true: a decrease in PC costs would lead to a larger share of direct shipments.
3. Nested level choice in the LCM models determines the locations of distribution centers that are handling indirect flow. The costs of direct shipments and the per-ton handling cost (the same global value applicable to all regions) do not influence this choice. The PD and DC transport costs influence the probability that a distribution center is used in a certain region in an interesting way. First, the higher the PD and DC costs, the more likely that the distribution would be close to the optimal route between production and consumption, making long detours less likely. Second, an increase in the DC cost would favor distribution close to the consumption points, while an increase in PD costs would favor distribution close to the production locations. However, it should be noted that PD costs are generally not higher than DC costs (otherwise distribution loses one of its primary functions). A relatively high PD cost (this is the case if DC shipments are large and there is no substantial ton-kilometer cost difference between PD and DC shipments) leads to the choice of the central locations for the distribution facilities.

The second class of variables is the price-sensitivity parameters. These parameters determine how strongly the model would react to the difference in the (dis-)utility of the choices. A higher value (absolute value as all sensitivity parameters in the model can take only negative values) of a sensitivity parameter makes the choice function to react stronger to the difference in, for instance, costs associate with each of the choices. This means that an alternative with a better choice gets more flow assigned to it than it would be the case with a smaller sensitivity parameter value. A smaller value makes model reaction less strong, if the value of the parameter is 0, the choice function will not distinguish the choices at all and each choice will get the same share of the total flow.

A higher value of gravity model sensitivity parameter makes it more sensitive to the price signals and essentially works in the same way as a higher value of the TLC: trade flows between nearby regions would increase and the trade flows between far away regions would decrease. A higher value of the sensitivity parameter at the top level choice in the LCM will assign more flow to the choice with the smaller disutility. In the context of the Dutch model implementation, it would mean more direct flows, as the disutility of direct flow is generally smaller. Finally, a higher value of the sensitivity of the nested choice in the LCM would lead to more flow via the “cheaper DC’s”. In the context of the Dutch model application it would

mean that closest DC's to consumption will get more volume and that the regional attractiveness factor will be more prominent in determining of the PD and DC spatial flow realization.

The third class of variables is the regional distribution attractiveness. A positive value of this parameter increase the costs of the logistics chain going through the region to which it apply. A negative value decrease the distribution costs associated with the region and increase its attractiveness for the distribution. The attractiveness variable includes all location-specific costs (real and fictive) that influence the distribution volumes of a region.

6.1.3. LCM output indicators for scenario analysis

The output of the logistics chain model is the four transport flow matrices: PD, DC, PC and the total transport flow OD matrix. For the Netherlands, each of these matrices contains 1600 data values, thus the matrices themselves are not a good means for scenario analysis. A meaningful scenario analysis should be based on a smaller number of indicators that are useful for the purpose of the analysis.

In the context of transport systems, policy makers are generally interested in welfare and environmental effects. For instance, transport causes adverse effects on infrastructure load, death toll, and is directly linked to congestions and other externalities. For example, the European statistics agency (Eurostat) publishes data on transport flows measured in ton-kilometers per mode, energy use and pollution at the level of EU countries (see European Union, 2013). The EU policies are based on the future projections of these parameters, see for example European Commission (2013).

The ton-kilometer indicator is a good one for the estimation of the environmental effects caused by the transport. For instance, the TREMOVE model (De Ceuster et al (2005) and Annema et al (2006)) can convert ton-kilometer values into CO₂, NO_x and other emissions. Advanced economic models, such as for instance EXIOMOD (Ivanova 2014) can use ton-kilometer transport values as an input for the economic analysis.

Direct transport emissions and direct load on infrastructure networks depend more on the number of vehicle-kilometer travelled than on ton-kilometer shipped. For instance, the COPERT emission database can help using right emission factors (Ekström et al 2004). Therefore, the following the ton-kilometer and vehicle-kilometer output measures are important for economic and environmental analysis. Moreover, if a transport model can estimate vehicle-kilometer indicator directly, such estimation would generally be of a better quality than it would be the case if another model uses ton-kilometer values as the input for environmental performance analysis. The estimations of emissions can be even further improved, if the weight of the vehicle and its cargo is taken into account (Ligterink et al., 2012). The LCM model provides the data on cargo weight and vehicle kilometers travelled, thus making possible very detailed scenario analysis with respect to emissions.

Additionally to the ton-kilometer and vehicle-kilometer indicators, the regional warehouse or distribution throughput is a useful indicator. First, it indicates directly the level of economic

activity linked to the logistics; second there is a clear relationship between employment in the wholesale sector and distribution volumes, as shown in Davydenko (2011 and 2012). The employment related to the distribution activities can be of direct interest for the policy makers at the regional and state levels. Third, regional warehouse throughput may be important for spatial planning purposes. Land use related to freight transport is expanding and may involve major investments (Dablanc et al., 2014). We discuss this main topic of application in the next section. TABLE 6.1 summarizes the proposed indicators.

TABLE 6.1. Indicators for assessment of the LCM output

Indicator	Explanation
PD_{tkm}	Production-Distribution flow expressed in ton-kilometers at the system level
DC_{tkm}	Distribution-Consumption flow expressed in ton-kilometers at the system level
PC_{tkm}	Production-Consumption flow expressed in ton-kilometers at the system level
OD_{tkm}	Origin-Destination flow expressed in ton-kilometers at the system level. This flow is the sum of PD, DC and PC ton-kilometers
PD_{vkm}	Vehicle-kilometers travelled to satisfy Production-Distribution flow
DC_{vkm}	Vehicle-kilometers travelled to satisfy Distribution-Consumption flow
PC_{vkm}	Vehicle-kilometers travelled to satisfy Production-Consumption flow
OD_{vkm}	Vehicle-kilometers travelled to satisfy the total flow
D_i	Distribution activity in region i , measured as annual number of ton volume of goods departed from distribution location in region i

6.2. Logistics Sprawl in Randstad

Several studies show that logistics facilities have spread spatially from relatively concentrated clusters in the 1970s to geographically more decentralized patterns around large urban areas (see e.g. Dablanc, 2014 for Los Angeles and Seattle). This phenomenon of ‘logistics sprawl’ is defined by Dablanc and Rakotonarico (2010) as the movement of logistics facilities outside the cities towards suburban areas. Although logistics sprawl has not been studied directly for the Netherlands, it is probable that the phenomenon also took place in the country, especially in the Randstad region.

The Dutch Randstad mega-region, where some 7 million people live and where some major European entry ports are located, is heavily populated with logistics facilities. The spatial development and the locations of logistics facilities determine not only decisions on infrastructure investments and usage of heavy goods vehicles, but also affect commuting transport and job availability. Next, the locations of logistics facilities influence the environmental consequences of freight transport operations. This is not only the case for CO₂ emissions (as a derivative of the number of truck kilometers and fuel used), but also for the more local pollutants and as a result the air quality in the populated urban areas. Only a part of the freight trips from these logistics facilities concern distribution activities, in contrast to long-haul transport. But, especially for these distribution trips the impacts on both the environment and infrastructure are severe, since receivers’ demands (e.g. opening hours of

stores or other facilities) and local regulations (e.g. time-windows) narrow down the available time for these trips to the peak period in the morning. As a result, these trips contribute even more to the negative environmental impacts of freight transport (Quak and De Koster, 2009).

Logistics facilities in suburban areas are often the origin of these distribution trips (or the destination of pick up trips), and therefore the locations influence the effectiveness of local policies, such as low emission zones, or city logistics initiatives, such as consolidation centers. Van Duin et al. (2013) show an example of the importance of finding the right location for a transshipment hub (for transshipment of parcels from conventional trucks to electric vehicles for the final urban deliveries in Amsterdam) in relation to the conventional truck's destination hub in order to make this type of zero emission urban distribution also financially viable.

Therefore, understanding the possible future patterns is important for effective regional as well as local policies concerning spatial development and planning, social policy, transport and infrastructure investments. The understanding of future transport flows is also important for the environmental consequences in these densely populated areas. The (future) locations of logistics facilities are uncertain, as they are determined by many factors that are exogenous to the logistics system and are difficult to predict. In the case of the Netherlands, besides the expected growth of port throughput (a tripling of container flows in the highest scenario by 2030), changes in the geography of consumer markets also heavily impact the spatial patterns of distribution systems. Up until recently, there were no adequate quantitative methods available that could help explore future patterns of regional logistics and distribution systems (logistics sprawl) in an empirically proven way, in contrast to describing the historical development of the phenomenon.

This analysis of the logistics sprawl effects for the Randstad region is based on the work presented at the METRANS conference (see Davydenko et al., 2013). The LCM allows for evaluating future scenarios and the effects on the locations of logistics facilities. This section shows how sprawl-related scenarios can be defined and analyzed with the LCM model. These outcomes can contribute to making better decisions on both local and regional spatial planning, infrastructure investments, and as a result can help to reduce negative environmental consequences of freight transport. The results of this analysis are also used to determine price elasticity of transport and warehousing activities in the Netherlands (see section 6.3 for more details on elasticity).

6.2.1. Logistics Sprawl Scenario Elaboration

The Netherlands is a relatively small country with a high average population density of 404 people per km² (CBS 2013). The Randstad region of the Netherlands is an urbanized area, comparable to areas such as greater Los Angeles. The Randstad region occupies approximately 20% of the total country's area and has a population density of 1170 people per km². The Ports of Rotterdam and Amsterdam and Schiphol Airport are located in the region.

The Randstad region plays an important role in logistics and distribution, see FIGURE 4.1 for the visualization of regional distribution volumes in the Netherlands. In the context of

logistics sprawl question, we consider centralization of logistics activities if the share of Randstad in total logistics activities grow and the share of logistics activities outside of Randstad decreases. The logistics chain model presented in this thesis provides a way of estimation for various policy options on distribution systems and transport volumes generated by them. Another purpose of the model is to understand side-effects of the policies, autonomous trends, etc.

For the analysis of the logistics sprawl, the LCM estimation on the Dutch road transport survey data has been used as the reference (basis) scenario. The model variables such as transport and warehousing costs, as well as regional attractiveness reflect current reality in the Dutch distribution and warehousing industrial sector. Adjusting these parameters scenario-wise (see section 6.1.2 on influence of the model variables), provides an insight into the changes in distribution system that would be observed in reality if underlying factors change in line with the assumed in scenarios changes. In the context of logistics sprawl study, the following scenarios have been defined:

1. *Current situation.* It is the base reference scenario that is equal to the outcome of the calibrated on the Dutch road transport survey data LCM model. All other scenarios are compared to the current situation scenario.
2. *Push towards centralization.* As a policy measure, a push towards centralization (counteracting logistics sprawl) can be realized by an increase in attractiveness of the Randstad region or decrease in attractiveness of non-Randstad regions. In practice, it can be realized through local tax incentives, regulations on ground prices and other measures. Given the estimated distribution costs of 9,2 Euro per ton of throughput, an increase or decrease of regional attractiveness parameter A_k by 1 Euro represents a change in distribution costs by around 11%. Note that parameter A_k represents extra costs related to distribution in a region. Negative parameter values increase attractiveness of the region and positive values decrease attractiveness. Therefore, an increase in value of this parameter decreases attractiveness of the region k , and conversely, a decrease in value of the parameter increases attractiveness of the region k for the distribution activities. The push towards centralization can be achieved in two ways: a) increasing attractiveness of Randstad or b) decreasing attractiveness of non-Randstad regions:
 - a. *Increase in attractiveness of the Randstad region.* This instrument is realized by a reduction of A_k values by 1 Euro per ton unit of throughput, where index k belongs to the Randstad set of regions.
 - b. *Decrease in attractiveness of the non-Randstad regions.* This instrument is realized by an increase of A_k values by 1 Euro per ton unit of throughput, where index k belongs to the non-Randstad set of regions.
3. *Push towards decentralization.* Similarly to the scenario's 2a and 2b, the decentralization (helping logistics sprawl) is realized by a change in the regional attractiveness parameter A_k .
 - a. *Decrease in attractiveness of the Randstad region.* Parameter A_k is increased by 1 Euro per ton of throughput for the regions belonging to Randstad.

- b. *Increase in attractiveness of the non-Randstad regions.* Parameter A_k is decreased by 1 Euro per ton of throughput for the regions belonging to the non-Randstad set of regions.
 4. *Increased road transport costs.* A change in transport costs will have an effect on spatial organization of the logistics and distribution systems. Transport costs per ton-kilometer unit are important decision variables in the model. For the scenario on increased transport costs, two sub-scenarios are considered
 - a. *Increase in transport costs and constant loads.* In this scenario, an increase of vehicle-kilometer c^{vkm} cost by 10% is considered. The average vehicle loads L^{PD} , L^{PC} , and L^{DC} are left unchanged, thus all three types of transport become 10% more expensive on the ton-kilometer measure. This scenario does not take into account a possible reaction of the distribution systems in respect to vehicle loads.
 - b. *Increase in transport costs and increase in production to distribution average load L^{PD} .* In this scenario, the vehicle-kilometer cost c^{vkm} is increased by 10%, as in the scenario 4a. However, in this scenario we make an assumption about a possible reaction of the distribution systems to the increased transport prices through an increase in production to distribution loads L^{PD} of 5%. The vehicle loads related to transport to the customers (L^{PC} and L^{DC}) are assumed to remain the same. This implies that ton-kilometer transport costs are increased by 5% for the PD flows and by 10% for the PC and DC flows.
 5. *Decreased road transport costs.* Similarly to the scenarios 4a and 4b, the effects of decreased transport costs are studied on the transport and distribution volumes.
 - a. *Decrease in transport costs and constant loads.* The vehicle-kilometer transport costs are decreased by 10% in comparison to the reference scenario 1. Transport shipment sizes and vehicle loads remain unchanged.
 - b. *Decrease in transport costs and decrease in average loads to customers L^{PC} and L^{DC} .* In this scenario we make an assumption that companies react to a decreased transport cost by improving customer service by increasing delivery frequency to the customers by 5%, which results in a 10% decrease in PD flow transport costs and a 5% decrease in PC and DC type of flow transport costs.

To estimate scenario outcomes, the following parameters are presented (the indicators are discussed in more detail in section 6.1.3)

1. Annual regional distribution volumes, expressed in ton of warehouse of distribution center throughput. These volumes are measured at the NUTS3 level: the Netherlands is divided into 40 regions of which 10 belong to Randstad. The changes in regional distribution volumes for Randstad and non-Randstad regions are presented in an aggregated form: the Randstad and non-Randstad distribution volumes are shown.
2. Country level number of ton-kilometers and vehicle-kilometers. These two indicators are further split into three transport legs, namely from production to distribution (PD), from production to consumption (PC), and from distribution to consumption (DC). The interregional distance table, which shows distances between region centroids, is used to compute these parameters, conform formulae 3.23 and 3.24, as the distances are also used for transport cost calculations.

6.2.2. Logistics Sprawl Scenario Outcomes

TABLE 6.2 presents the main scenario outcomes. The changes in regional distribution throughput are presented as the percentage change in respect to the regional distribution volumes of the current scenario, reflecting the impact of the scenario changes in the system. The changes in the number of ton-kilometers and vehicle-kilometers are presented for production-distribution, production-consumption, distribution-consumption and total flows. The scenario-wise change reflects the changes in comparison to the current situation scenario.

TABLE 6.2. Scenario estimation results

<i>Changes in regional distribution throughput volumes, ton per year</i>								
Scenarios / Distribution throughput change	Randstad			Non-Randstad				
Centralization 2a (Randstad more attractive)	16,1%			-0,4%				
Centralization 2b (Non-Randstad less attractive)	1,2%			-13,4%				
Decentralization 3a (Randstad less attractive)	-13,7%			0,6%				
Decentralization 3b (Non-Randstad more attractive)	-0,9%			15,6%				
Increased transport costs 4a	-0,8%			-1,0%				
Increased transport costs 4b	5,5%			6,2%				
Decreased transport costs 5a	0,8%			1,3%				
Decreased transport costs 5b	6,4%			7,5%				
<i>Changes in country-level number of ton-kilometers and the number of vehicle-kilometers</i>								
Scenarios / Global Parameter	PD tkm	PC tkm	DC tkm	PD vkm	PC vkm	DC vkm	Total tkm	Total vkm
Centralization 2a	4,6%	-0,6%	4,7%	4,6%	-0,6%	4,7%	0,2%	0,1%
Centralization 2b	-8,9%	1,2%	-8,8%	-8,9%	1,2%	-8,8%	-0,3%	-0,2%
Decentralization 3a	-3,7%	0,5%	-3,8%	-3,7%	0,5%	-3,8%	-0,1%	-0,1%
Decentralization 3b	10,4%	-1,4%	10,4%	10,4%	-1,4%	10,4%	0,4%	0,3%
Increased transport costs 4a	-3,4%	-3,1%	-3,9%	-3,4%	-3,1%	-3,9%	-3,1%	-3,1%
Increased transport costs 4b	7,4%	-4,2%	2,1%	2,3%	-4,2%	2,1%	-2,9%	-3,4%
Decreased transport costs 5a	3,9%	3,1%	4,4%	3,9%	3,1%	4,4%	3,2%	3,2%
Decreased transport costs 5b	12,6%	0,5%	8,2%	12,6%	5,8%	13,9%	1,9%	6,8%

The regional distribution volumes can be influenced by the distribution throughput costs. Scenario 2a shows that a decrease in distribution-related costs by 1 Euro per ton of throughput, which represents some 11% of the distribution cost, leads to an increase of Randstad distribution volumes by 16,1%, while taking only 0,4% of the volumes of the non-Randstad regions. A more attractive distribution in the Randstad region leads in this scenario to a shift from direct shipments to shipments via distribution in Randstad. Scenario 2b, where the costs of non-Randstad distribution are increased by 1 Euro per ton of throughput, show a decrease of non-Randstad volumes by 13,4% and some spillover effect to the Randstad region and an increased volume of direct shipments. Similarly, an increase in the costs of distribution in Randstad (Scenario 3a) leads to a decrease of 13,7% of the volumes and a spillover to non-Randstad regions of 0,6%. Scenario 3b shows an increase of non-Randstad volumes by 15,6% in case if distribution there becomes cheaper. FIGURE 6.1 illustrates the effects of

centralization scenario, which makes the Randstad region more attractive for distribution, and thus can be considered as a measure against the sprawl of distribution activities outside of Randstad. The purple color indicates an increase in distribution volumes, the green color indicates a decrease. The coloring is for illustration purposes and is not proportional to the regional distribution volumes, however, intensity of the color provides a basis for perception. The darker the color is, the stronger is the change in regional distribution volumes. For the interested reader, Appendix II presents the source data of FIGURE 6.1.

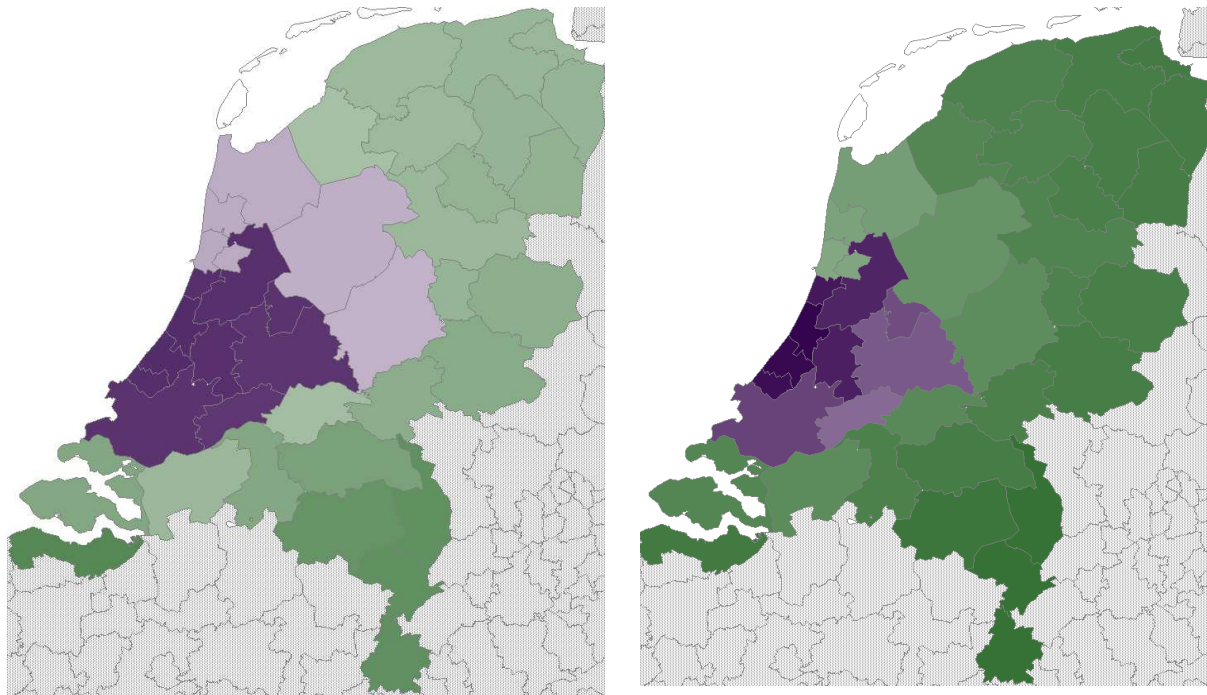


FIGURE 6.1. Effects of Scenario 2 (2a – left and 2b – right) on regional distribution volumes

The visualization of FIGURE 6.1 shows that in case if Randstad is made more attractive, it attracts extra distribution volumes in a practically uniform way. The regions to northeast of Randstad will get small extra distribution volumes too. In general, the further a region is from Randstad, the stronger is the scenario effect. This applies for both scenarios, 2a and 2b.

An important lesson from the scenarios 2ab and 3ab is that if the regional distribution volumes are to be changed (decreased or increased), then the distribution attractiveness of the regions to be influenced should be tackled. This can be done through extra taxation or subsidies (these measures are easy to model in the LCM model as subsidies or taxation have a certain monetary value) at the regions that need to be influenced. The model shows that influencing attractiveness of the regions has also an effect on the regions that are not influenced directly, but the effect is approximately an order of magnitude smaller.

An increase in transport costs in scenarios 4a and 4b has led to a decrease in the volumes transported. A 10% increase in transport costs on all segments (scenario 4a) leads to a decrease in transport volumes and the number of vehicle-kilometers driven by -3,1%. Scenario 4b, which assumes that an increase in transport costs would force increased loads on production-distribution segments, has led to a slightly more modest decrease in the number of

ton-kilometers shipped (-2,9%) and a slightly bigger effect on the number of the vehicle-kilometers (-3,4). Scenario 4a also decreases the distribution throughput, but only in the range of -1% to -0,8%, thus having a more limited effect on distribution. An increase in transport costs leads to a spatial trade redistribution, favoring shorter-distance trade flows, which make direct shipments more attractive, even in the case of higher transport prices. This fact should be taken into account: higher transport costs do not automatically lead to more distribution and consolidation, but also impacts spatial redistribution of trade flows. It should be also noticed that the distribution volumes more than 3 times slower than the transport volumes.

Scenario 4b, which assumes that the logistics systems react to an increase in transport prices by a larger degree of consolidation of the flows to the distribution, shows a substantial increase in the regional distribution throughput, by 5,5% for the Randstad region and by 6,2% for the non-Randstad regions. A higher increase in distribution activities of non-Randstad region is due to the fact that the population density is smaller there, and hence larger average distances: consolidation at the distribution centers becomes even more attractive for the non-Randstad regions as shipments to consumption become more expensive and distribution ensures that those shipments are reduced in length.

A 10% decrease in transport costs (scenario 5a) leads to an increase in distribution volumes for both Randstad and non-Randstad regions, however non-Randstad regions get structurally more distribution volumes, an increase of 1,3% compared to 0,8% increase in Randstad region. Essentially the results of scenario 5a are the mirrored results of scenario 4a, where an increase in transport costs leads to a decrease in distribution volumes, with non-Randstad regions reacting somewhat stronger, see TABLE 6.2. If the costs of the deliveries to the customers go less steep down due to an increase in delivery frequency, the distribution gets larger volumes, see scenario 5b outcome. Scenario 5b also shows that non-Randstad regions will get slightly more distribution volumes than the Randstad region. Scenario 5b shows the worst performance of the logistics sector in respect to environmental indicators because it leads to a substantial increase in the number of vehicle-kilometers driven, thus creating an extra load on infrastructure and emitting more pollutants.

It should be noted that while scenarios 4a and 5a are practically mirrored, with the system reacting slightly stronger to a decrease in costs (3,1% decrease in total volumes for scenario 4a and 3,2% increase for scenario 5a), scenarios 4b and 5b are not completely mirrored. In scenario 4b we model a 10% increased transport costs for DC and PC flows, and a 5% increase in transport costs of PD flows. Due to the fact that DC and PC flows represent more than 90% of all ton- and vehicle- kilometers in the modelled system, the aggregate transport price increase is between 9,5% and 10% in both ton-kilometer and vehicle-kilometer measures. In scenario 5b we model a 10% decreased cost for PD flows and a 5% decreased ton-kilometer cost of DC and PC flows. Taking into account that DC and PC flows account for more than 90% of the flow, the system-wide change in transport cost is a decrease of 5% - 5,5%, which is not a mirror situation of scenario 4b. Due to the fact that DC and PC vehicles are carrying less load (95% of the reference load), there is more vehicle movements on PC and DC segments.

It is often assumed that logistics sprawl is driven by decreasing transport costs and difference in the costs of facility, as in the suburban areas the ground prices and some other cost components are smaller than in the historical areas. According to the LCM logistics sprawl scenario analysis, both assumptions are true in the Randstad region. However, it should be noted that transport costs have a smaller impact on logistics sprawl: a decrease in transport cost leads to less sprawl manifestation than a similar percentage change in the warehousing and distribution throughput costs. The sprawl is more sensitive towards region-specific distribution costs than to the system-wide transport costs. This presents a lesson for the decision makers: ground prices and costs of building permits can be an effective tool for dealing with the logistics sprawl.

6.2.3. Discussion on Quantitative Evaluation of Logistics Sprawl

Section 6.2 presented an application case of the LCM, where the effects of a change in model variables lead to a change in the flows that the model estimates. The case of logistics sprawl in the Netherlands proposes scenarios that reflect possible sprawl-related policy measures and quantifies the effects of the measures on transport flows and distribution throughput. Therefore, the model makes possible analysis and understanding of the future freight systems' patterns. This application shows the capacity of the model to help develop effective regional policies concerning spatial development, social policy, transport and infrastructure investments. It also provides a tool for studying the environmental consequences in these densely populated areas through indicators on ton-kilometer shipped and vehicle-kilometer driven.

The LCM model is a neutral tool in a sense that it does not prefer any outcome: the model is capable of finding the values of subsidies or taxation necessary to achieve the desired effect with respect to the regional distribution and warehousing volumes. An important aspect of the LCM model is that it not only looks at the changes of spatial logistics patterns in freight systems, but also includes the impact of the changes in logistics systems on the trade patterns: production and consumption are matched in the gravity model, which uses generalized costs for the deterrence function. In case if only transport and logistics system need to be considered, the gravity model can be "turned off" by "freezing" the trade flows estimated in the base scenario. This would allow studying the expected impacts on spatial organization of trade and spatial organization of transport and logistics separately.

6.3. Elasticity of flows for road transport costs

The combined gravity-logistics chain model provides an opportunity to estimate price elasticity of the road freight systems. The concept of price elasticity is a very useful one: it allows an estimation of the effects of a price change on the demand for a service or a product without a complex model application. The price elasticity for road transport demand has already been discussed in section 5.1.8 on price elasticity implied by the estimation of the gravity model, see also formula 5.1 for a simple definition of it. However, being a very practical concept, there is no one firm and conclusive value or estimation of the price elasticity; the literature overview on this subject provides a wide range of values, see

Significance & CE Delft (2010) and de Jong et al (2010). There is abundant literature on elasticities of mode choice, none around DC's, and very scares for trade.

The price elasticity in respect to ton-kilometer freight volumes and vehicle-kilometer driven are concluded directly from the logistics sprawl case of section 6.2. These results are discussed in relation to the elasticities provided by gravity model applications at the OD level (as opposed to the chain level of the LCM). The price elasticity of the demand for distribution services is also discussed

6.3.1. Elasticity estimated by the LCM in the case of logistics sprawl

There is a wide range of elasticity estimations in the literature. Significance & CE Delft (2010) provide a comprehensive review. An important finding is that elasticities are always calculated around specific degrees of freedom in the system. In the context of specific transport decisions, with only one exception, all elasticity values for road freight are calculated within a context of choice between modes of transport. We add to the literature by considering the choice of DC structure and trade.

Based on the logistics sprawl study of section 6.2, the transport price increase of Scenario 4a and transport price decrease of Scenario 5a lead to a decrease in transport volumes, showing both ton-kilometer price elasticity and vehicle-kilometer price elasticity in the range of -0,32 and -0,31. Note, that if certain assumptions are made on the reaction of transport and logistics systems to the change in vehicle-kilometer prices (Scenarios 4b and 5b), the elasticity values are different and vehicle-kilometer elasticity diverges from the ton-kilometer elasticity.

Considering scenarios 2a,b and 3a,b of the logistics sprawl case, it is possible to draw conclusions on the price elasticity of demand for distribution services. The elasticity is estimated to be in the range -1,45 and -1.18, thus being considerably stronger than the transport demand price elasticity. As the author is not aware of other efforts on estimations of the distribution or warehousing price demand elasticity, these values cannot be compared to other estimations. *This is the first estimation provided on the elasticity of warehousing and distribution demand.*

It is not possible to provide a firm explanation on why demand for distribution is some 4 times more elastic than transport service demand. However, there are some plausible assumptions that may explain this effect. First, the LCM model takes regional production and consumption as exogenous data, thus the model's output satisfies it, meaning that the goods need to be shipped from production to consumption anyway. However, the choice of the chain is an endogenous choice: the flow can be realized as a direct flow (without any distribution) and the flow can be realized via distribution in other region, which is not affected by the price change. Those two possibilities make the model reaction to a change in distribution relatively strong, especially if shifting distribution from region A to region B does not change the number of kilometers driven substantially. It is expected that this elasticity range would also hold in reality, as the shipping companies would have a choice on whether to deal with distribution at all, and if so, then in what region. The main lesson that can be drawn is that price elasticity of distribution facilities is much stronger than that of transport.

6.3.2. Discussion on Gravity Model and LCM Elasticity Estimations

Applications of the Separately Calibrated Gravity Model (SCGM) in section 5.1 and logistics sprawl case both lead to road transport demand price elasticity estimations. Section 5.1.8 provides details on SCGM elasticity estimations and section 6.3.1 provides details on the elasticity estimations done in the combined LCM. These elasticity estimations can be compared.

The SCGM and LCM elasticity estimations cannot be directly compared because SCGM values apply only for OD segments, while LCM estimations apply for substantially more complex logistics chains, including distribution of trade flows. Furthermore, SCGM elasticity estimations depend on the length of the transport segment, reflecting the fact that a longer distance road transport reacts stronger to the structural changes in underlying costs. The LCM elasticity estimations cannot be made distance-dependent, as there is no simple analytical way of doing it (the LCM outcome is the result of flow estimation by a complex chain model).

Direct transport demand price elasticities match if one compares elasticity of PD flows on 25-km segments, which is -0,316, and the LCM estimations of the range -0,32 and -0,31. Note that the gravity model does not make any distinction between ton-kilometers and vehicle-kilometers. Other estimations of the gravity model (other segments and longer stages) are higher, see TABLE 5.3. However, if one takes not only transport price change into consideration, but the whole chain price change, then the range of elasticities for the LCM estimations will be higher, as distribution is more price sensitive than transport. It should also be noted that if a certain response of the transport and logistics sector is assumed, as for instance in Scenario 5b of the logistics sprawl case, the elasticity estimation by LCM is higher and close to the average estimation from the gravity model.

It is agreed that the LCM provides for a better tool for transport demand price elasticity estimations compared to the estimations made using the gravity model. The LCM works at the level of logistics chains and logistics systems, taking a more holistic approach for the freight transport system. The elasticity estimations in the LCM by default include the impact on trade distribution in the coupled gravity model, however, if necessary trade distribution can be excluded from the analysis.

6.4. Discussion

This chapter looked at the application of the combined gravity-LCM model. First, main classes of model variables were discussed, exploring quantitatively how an adjustment of a variable value would impact the resulting flows that the LCM estimates. Subsequently the ideas on useful indicators that can be estimated by the model were explored: the model's output in the form of flow matrices is not a very intuitive way of scenario analysis, thus more aggregate indicators such as vehicle-kilometers travelled, ton-kilometer shipped and regional distribution throughput were discussed. These indicators formed the basis for scenario assessment in the logistics sprawl case. The key results included a demonstration of the LCM model capacity to perform scenario-wise analysis of complex societal issues such as logistics

sprawl, estimate elasticities applicable for road transport in the Netherlands and the choice of distribution regions, and the ability to determine the necessary level of (financial / fiscal / subsidy) intervention to achieve a desired level of change in the transport and logistics patterns.

This chapter has shown that the LCM is the right tool while assessing the problems of spatial organization of logistics systems, such as the question of the sprawl of logistics facilities. The sprawl phenomenon is observed in many large agglomerations, such as Paris, Los Angeles and the Randstad region of the Netherlands. The essence of this phenomenon is that from the 1970s onwards, logistics facilities have been spreading from tightly knit clusters to a much wider area. The scenario-wise LCM application has shown that the sprawl of logistics facilities (or in the context of the LCM of distribution facilities, which is a subclass of logistics facilities) can be modelled and assessed. The modeling exercise confirms that logistics sprawl is driven by the long-term decreasing transport prices and the difference in facility costs between historical areas and the suburban areas, where the facilities tend to be cheaper.

Without making a judgment of the desirability of logistics sprawl, the model illustrates that it can be both facilitated or reduced by influencing the attractiveness of the regions to distribution. As the model provides quantification for this, the main influence mechanisms are clearly in the hands of local or national authorities: it is relatively easy to change the distribution costs by local taxation or provision of subsidies. There are two important lessons that the policy makers can draw from the model application. First is that the regional distribution volumes are sensitive to the distribution costs; and second is that the distribution and warehousing costs should be influenced in the region where the policy should have an effect, the spillover of distribution volumes into other regions is approximately one order of magnitude weaker. Essentially, the case of logistics sprawl is a macro-level illustration of basic logistics trade-offs (FIGURES 2.1, 2.2 and 2.4) that drive design of the logistics systems.

This chapter has also provided an estimation on transport demand price elasticity in respect to ton-kilometers shipped and vehicle-kilometers driven. The demand price elasticity is a very useful and intuitively understandable indicator applicable in almost all fields of human activity where there is trade in products or services. Hence, the enormous interest for the estimations of this indicator. A unique feature of the logistics chain model is that it is capable for provision of the regional distribution and warehousing demand price elasticity. The LCM elasticity estimations are alongside other known values that can be found in the literature. Still, the elasticities depend on many factors, for instance, inclusion or exclusion of the trade distribution in the gravity model.

There many other applications possible of the combined LCM, which are beyond the cases of this chapter. The model could be made a part of a larger (national or continental or global) modeling suites, where trade flow is estimated in “upstream” modules of the model and LCM’s output is used as input for the vehicle choice and / or route choice modeling. An incorporation of the LCM in the comprehensive modeling suites would remove a structural modelling bias and make these models more consistent and structurally sound.

7. Conclusions

The research presented in this dissertation answers five research questions. The following list repeats the research questions and assesses research results of this thesis.

RQ 1: How can a freight model be designed and implemented with the aim to estimate empirically valid transport flows necessary to ship the goods from production to consumption locations, emphasizing a proper modeling of flows related to warehouses and distribution centers? The LCM model estimates (generates) three types of transport flow: direct production-consumption flow (PC), production-distribution flow (PD) and distribution-consumption flow (DC). The resulting total transport flow (OD) on infrastructure networks is the sum of these three sub-flows. The generated transport flow is structurally not equal to the trade flow and takes into account transport “detours” to visit distribution facilities.

RQ 2: What are the alternative approaches to modeling of distribution structures in a macro level freight model? Literature study shows that distribution structures can be modelled at aggregate and disaggregate levels. Disaggregate level modeling requires generation of a population of actors involved in transport activities and subsequent modeling of the behavior of the actors. Aggregate models operate at the level of aggregate agents or average costs. Disaggregate models are more suited for modeling of specific sectors, such as food retail (Friedrich, 2010) or distribution of drinks (Maurer, 2008). Aggregate models can be broad in scope, such as SMILE (Tavasszy et al. 1998) and EUNET 2.0 (Yin et al. 2005). For both classes of models, the issue of empirical validity is a challenging one.

RQ 3: What are data requirements for logistics model of RQ 3 and what are the data availability, quality and update policies? The LCM model needs three classes of data, of which two belong to the model input: P/C trade flow data and cost data; and one class of data, which is necessary for model calibration, transport flow OD data. In practice, the author of this PhD thesis did not find a reliable data source for the model input data classes; instead, the trade flow and the costs are estimated endogenously within the model. This has been made possible by the third class of data (transport OD flows), which should be, in principle, needed

only for the calibration purposes. However, due to the richness of transport flow database, it is possible to estimate trade flows on it, and in the calibration run estimate the cost parameters of the model. The transport data OD are updated annually by the CBS.

RQ 4: How do changes in transport and warehousing costs interregional freight flows?

This thesis presents a model application case for the question of logistics sprawl in the Netherlands. Possible intervention measures in the form of a monetary based change in regional attractiveness for dealing with the logistics sprawl are presented. The transport services demand price elasticity is estimated; the price elasticity of demand for distribution and warehousing services is also presented, which is a novel estimation.

This chapter further presents out conclusions on the available body of the relevant knowledge, model design choices and available data, quality of the model estimation and discusses the model's applications. It is finalized with an outline of the most promising areas for further research and development.

7.1. Conclusions on available knowledge

Traditional freight transport models, historically derived from the passenger transport models, do not account properly for logistics structures that freight is following on the way from production to consumption. The mismatch between trade flow and transport flow was traditionally reconciled by using a multiplier factor that accounted for transshipments and distribution, thus equaling trade and transport ton units. This modeling practice is not able to capture the operations and complexities of logistics systems, hence leading to incorrect transport flow estimations and conceptually weak modelling. From this fact arises the need to model logistics properly in freight models.

Logistics operations are guided by the business requirements and basic cost trade-offs. The business requirements dictate a certain service level, which is often expressed in product availability rates if one speaks about stocks and points of sale, and delivery frequency if one speaks about stock replenishment and reordering policies. Given business requirements, a company has to make tradeoffs between stock keeping costs and transport costs, which are often translated into a complex spatial optimization problem at the company or supply chain level. Trends, such as decentralization, also play a role in logistics organization. The degree of centralization will be influenced by future changes in the relative costs of logistics inputs and the evolution of management practices in this field (McKinnon, 2009). Trends in increasing international trade and transport, requirements for high quality of logistics, as well as enabling technologies such as IT technology led companies to continuously optimize their distribution networks (Ruijgrok and Tavasszy, 2007).

The subject of supply chain optimization, however complex, is very well studied. The supply chain problem in its most generic form, has to determine production or sourcing locations, warehousing and distribution locations and the paths of goods from production via warehousing or directly to the given consumption locations. This generic supply chain optimization problem is computationally complex and there is a plethora of knowledge on

how to deal with it, for instance, Melo et al. (2009) review 139 literature publications on the subject. However, this wide body of knowledge concerns optimization of individual companies or supply chains, where the central element is that there is an entity which has the full control on the design of the chain.

When it comes to modeling of freight transport and logistics at the regional level, the most knowledge is accumulated in the field of micro models. If one has a complete knowledge on all companies in the region under consideration, the models that reproduce each company would perform reasonably well in respect to representation of processes, and eventually the transport flows that the businesses generate at the regional level. Obviously no one possesses such knowledge over all or nearly all businesses in the region. Therefore, researchers compensate for this deficiency by relaxing the requirements for the models: the businesses' populations are generated together with their trade and transport needs (e.g. Boerkamps, 1999 and Liedtke, 2009a), or concentrating on the modeling of limited population sizes or specific economy sectors (e.g. Friedrich, 2010), or removing the requirement for empirical validity and concentrating on specific commodities (e.g. Maurer, 2008). Overall, validity for broader population using micro level models is not shown / proven.

In aggregate models the choices are based not on the decisions of individual companies or controllers, but mostly deal with generalized utilities of the choices. The SMILE / SMILE+ model (Tavasszy et al., 1998) and the EUNET 2.0 (Williams et al., 2005) are the most interesting and relevant applications of the aggregate choice modeling. Being simple in their essence, the researchers working on this type of models are tempted to make them as rich and detailed as possible, as it is the case for SMILE and, to a lesser degree, for UENET 2.0 too. However, there is no good observational flow data to substantiate these details, which leads to the situation where the models are not calibrated nor validated at the flow O/D level.

The logistics model design discussed in this book relies on two requirements. The first requirement is that it should be empirically valid in its description of flows. The second requirement concerns its usability for strategic policy study applications, which requires that the model is based on (generalized) costs, describes all flows and has a spatial dimension. The scientific literature does not present cases that satisfy these requirements, thus the logistics chain model fills a scientific niche, at the same time presenting its applied value for its target group of users.

7.2. Conclusions on the mathematical model formulation

The logistics chain model takes trade flow as the input and generates transport flow as the output. The required model input in the form of commodity ton trade flows is not observed directly, at least in the Netherlands. Therefore, the model design consists of two coupled models. A gravity model is used to translate regional production and regional consumption into trade flows, and the logistics chain model is used for translation of the trade flows into transport flows.

The gravity model estimates the price sensitivity value for transport and trade flows. Applied to the transport flows, this parameter shows the distance decay. In the case that the model is used for trade flow estimation, the sensitivity parameter determines the spatial trade distribution, namely the impact of the costs and distance on the trade decisions.

The logistics chain model takes the trade flow matrix as the input and translates it into the transport flow matrix. The transport flow matrix is the sum of the three distinct flow types: direct production-consumption flow (PC flow), production-distribution (PD) and distribution-consumption (DC) flows. A two-level nested logit is applied to determine how the trade flow is realized: the top level choice is a binary one, it determines whether the flow will be direct or go via a distribution center. The nested choice is a multinomial one, it determines the share of distribution in each of the n regions. The model captures basis logistics trade-offs between inventory and transport costs, and the trade-off between (centrality) of inventory locations and transport costs. The LCM model is calibrated using a heuristic. The empirical validity of the model is shown by matching the model's output to the observational data by showing the quality of fit.

7.3. Conclusions on data requirements and data availability

The logistics chain model needs the following three classes of data to be available

1. Trade flow matrix or production-consumption matrix as the input
2. Transport flow matrix, which is consistent with the trade flow matrix, for calibration and validation
3. Data on transport and distribution costs for determination of the choices

In practice, only transport flow matrices are available from the empirical sources in Europe from the national statistics bureaus and the Eurostat. These data have rather limited spatial resolution and do not provide any further indication on the purpose of the flow. An empirically based trade flow database at regional level is not available at all. Transport and warehousing costs are not available too: there are known ranges for these costs and 'anecdotal' evidence of some privately paid prices, but not a firm generalization of these costs.

Statistics Netherlands has provided the transport flow data, extended with information on the purpose of transport, which comes from a variable on loading and unloading location types. These Dutch transport data contained not only the total aggregate transport flow data, but also the constituting sub-flows, such as production-distribution, distribution-consumption and production-consumption transport flows. The data were also be used for the estimation of the trade data in a gravity model. Thus, the available transport data lead to a synthetic availability of the trade flow data. The third data class on transport, warehousing and region-specific costs has partly been observed and partly estimated endogenously within the model. The related model variables have been estimated in the calibration procedure and based on the realistic low and upper boundary values for these variables. The model is normalized by a fixed value

of the variable for average production-distribution flow transport costs and shipment sizes. Availability of German and European data allowed model applications for these areas.

Special attention is devoted to the question on the quality of the location variable. Apart from a conclusion on sufficiency of the quality, the work produced an result on a matching of annual regional distribution and warehousing throughput volumes with employment in certain economy sectors. Employment data can be used as the basis for freight trip and flow generation models, as we provide a linear relationship between employment and volume of transport.

7.4. Conclusion on model implementation, estimation and application

The emphasis on the empirical validity of the logistics has led to an extensive analysis of the quality of the model's output measured as the fit between estimated flows and distribution throughput volumes on the one hand, and real world data on the other hand. Two alternative LCM's were specified, which both produced a good result with respect to the fit of the generated interregional transport flows. The main formulation uses transport costs directly as the main decision variable determining disutility of the logistics chain choices, while the alternative formulation determines the shipment sizes and derives relative transport costs from them. The two formulations allow estimations of transport and warehousing costs on the one hand, and shipment sizes (transport batch sizes) on the other hand.

Applications of the separately calibrated gravity model have shown that different types of transport show a different demand price elasticity, for instance, production-distribution (PD) transport flows are less sensitive to the price signals than production-consumption (PC) transport flows. This observation allows confirms the assumption that modeling of logistics is an important endeavor, as lumping all transport flows into one OD matrix hides important complexities and properties that are present in logistics. The logistics chain model is a better way of capturing the divergent transport flow purposes in trade-transport flow reconciliation.

In addition to the gravity model transport price elasticity estimations, the LCM model can also provide estimations on the total transport flow elasticity. The LCM model estimations are conceptually better, since they also include the spatial redistribution of trade, e.g. changes in the sourcing decisions by the companies. To our knowledge, the LCM model is the first empirical quantitative model, which provides an estimation for the warehousing and distribution service demand price elasticity.

The aspect of the policy-related usability of the LCM model demonstrated in the case of the so-called 'logistics sprawl', a phenomenon observed in many large population areas, where logistics facilities tend to spread out from concentrated clusters, dominant in the 70ies to an overall presence in our days. The LCM model has shown an ability to deal with the spatial distribution of the warehouses by assessing quantitatively actions needed for dealing with the phenomenon. Due to the fact that the model is calibrated on empirical data, shows a good

estimation quality and plausible responses in scenarios, the LCM model is a promising scenario analysis tool in the field of spatial location of warehouses and distribution facilities.

7.5. Outline of further research

We see two main areas for further research. The first research area presents a model-technical opportunity. The LCM model of this thesis only represents distribution-related flows of one echelon depth, i.e. a flow visiting a distribution center goes to a consumption location from the distribution. The empirical data suggests that there is a distribution-distribution flow, both in the Netherlands and in Europe. For instance, a company may have a European DC and a number of regional DCs supplied from the European one, which is a form of a multi-echelon distribution system. Incorporating the multi-echelon functionality into the LCM model of this thesis would result in an unsustainable growth of the computational time necessary for model estimation. Therefore, a new class of LCM is necessary to capture the complexities of multi-echelon supply chains, which presents a clear research opportunity. Another research opportunity is in introduction of more details into the logistics engine of the model. It may include an explicit treatment of distribution networks, an introduction of economies of scale in logistics networks (e.g. larger distribution facilities and a more dense spatial customer distribution should lead to smaller costs per unit shipped), take into account continuous deployment of Long and Heavy Vehicles (LHVs). The introduction of these logistics details and functionalities will require proper mathematical formulations and additional data for the model calibration and validation.

The second opportunity for further research is data driven. The LCM was made possible by the road transport data survey provided by Statistics Netherlands, which includes the loading and unloading location type variable collected through road transport surveys, filled in by the businesses operating road trucks registered in the Netherlands. Recently Statistics Netherlands has started encouraging businesses to automate statistics-related reporting (Davydenko 2014c). The statistics bureau provides an XML interface for statistical reporting for automatic preparation and transmission of the reports. There are already a substantial number of companies that report digitally to the CBS using the XML technology; some vendors of Transport Management Software (TMS) and board computers provide built-in functionality for the automated statistical XML reporting, see for instance (Logicway 2014). This development leads to more reliable and rich data, as it is automatically generated and does not contain errors related to human labor of filling in the forms. The XML data contain not a small sample of vehicle movements, but potentially a complete set of all movements. However, the XML data does not automatically contain location type information. Provided known location coordinates or addresses in the XML data, these can be matched with the register of business addresses and data on the type of activity performed at the address by combining data from a number of data sources. The coupling of the addresses visited with the information on business activity would improve high quality data input into the LCM, and probably open up other interesting freight modeling research opportunities.

Summary

The research presented in this dissertation has started with two simple questions. The first question was “how can trade flows be translated into transport flows”, with an understanding that trade is not equal to transport due to the obvious reason that some goods are transshipped at warehouses, distribution and consolidation centers. This question was related to the interregional flows, which occur at the regional, or in the modeling terms, macro level of abstraction. But warehousing and distribution are a business activity, undertaken by individual companies, and thus occur at the micro level of abstraction. Hence the second question was “how the micro-macro divide could be bridged in the realm of logistics modeling”, meaning that the model should be able to incorporate the business logic that the companies follow while establishing and optimizing their logistics designs and solutions.

This thesis provides an answer to the first question, as it shows that the Logistics Chain Model (LCM) successfully translates trade flows into transport flows, estimating locations and throughput of the warehouses and distribution centers with a good precision; the model also determines transport flows with a sufficient fit with respect to the observed transport flows. The second question has also been answered. The LCM shows that a simple logistics-cost driven macro model can sufficiently accurately represent myriads of processes and decisions taken at the level of individual companies, without a need for a very detailed modeling, if one is interested in a descriptive model for the whole regional logistics system. If the primary interest is in how a certain sector works, then perhaps, a detailed modeling is preferred, as for instance, in the model for the German food retailing sector (Friedrich 2010), which deals with micro-level actors to estimate macro-level effects.

More specifically, Chapter 1 formulated five research questions (see section 1.2 for more detail). The following list repeats the research questions and assesses research results of this thesis.

1. **RQ 1: How do logistics decisions taken by individual companies translate into aggregate flows at macro level?** The LCM model does not look at the decisions made at the firm level. Instead, it works at the level of aggregate decisions, estimating

what share of decision makers have made a certain choice. Subsequently this estimated share is translated into the transport flows: these are related to the probability that a decision maker would make a certain choice. Therefore, the LCM models aggregate transport flows looking at the aggregate decisions and not at the behavior of individual decision makers.

2. **RQ 2: How are supply chains influenced by the costs of transport and warehousing, and how do changes in these costs translate into changes in the interregional freight flow?** The LCM model treats costs explicitly, as the choice of logistics alternatives is determined by the costs of transport and warehousing. Translation of trade flow into transport flow prefers those logistics chains, where the costs are smaller over those chains where the costs are higher. Thus, the costs together with the price sensitivity parameters determine the probability that a certain logistics chain is used, and hence the share of the trade flow that would be shipped through the chain under consideration.
3. **RQ 3: How can a freight generation model be designed and implemented with the aim to estimate empirically valid transport flows necessary to ship the goods from production to consumption locations, emphasizing a proper modeling of flows related to warehouses and distribution centers?** The LCM model estimates (generates) three types of transport flow: direct production-consumption flow (PC), production-distribution flow (PD) and distribution-consumption flow (DC). The resulting total transport flow (OD) on infrastructure networks is the sum of these three sub-flows. The generated transport flow is structurally not equal to the trade flow and takes into account transport “detours” to visit distribution facilities.
4. **RQ 4: What are data requirements for logistics model of RQ 3 and what are the data availability, quality and update policies?** The LCM model needs three classes of data, of which two belong to the model input: P/C trade flow data and cost data; and one class of data, which is necessary for model calibration, transport flow OD data. In practice, the author of this PhD thesis did not find a reliable data source for the model input data classes; instead, the trade flow and the costs are estimated endogenously within the model. This has been made possible by the third class of data (transport OD flows), which should be, in principle, needed only for the calibration purposes. However, due to the richness of transport flow database, it is possible to estimate trade flows on it, and in the calibration run estimate the cost parameters of the model. The transport data OD are updated annually by the CBS.
5. **RQ 5: How can a policy measure be quantified in a model application case with the aim to achieve a specific level of change in spatial organization of distribution facilities?** This thesis presents a model application case for the question of logistics sprawl in the Netherlands. Possible intervention measures in the form of a monetary based change in regional attractiveness for dealing with the logistics sprawl are presented. The transport services demand price elasticity is presented; the price elasticity of demand for distribution and warehousing services is also presented, which is a novel estimation.

This chapter further presents out conclusions on the available body of the relevant knowledge, model design choices and available data, quality of the model estimation and discusses the

model's applications. It is finalized with an outline of the most promising areas for further research and development.

Available knowledge

Traditional freight transport models, historically derived from the passenger transport models, do not account properly for logistics structures that freight is following on the way from production to consumption. The mismatch between trade flow and transport flow was traditionally reconciled by using a multiplier factor that accounted for transshipments and distribution, thus equaling trade and transport ton units. This modeling practice is not able to capture the operations and complexities of logistics systems, hence leading to incorrect transport flow estimations and conceptually weak modelling. From this fact arises the need to model logistics properly in freight models.

Logistics operations are guided by the business requirements and basic cost trade-offs. The business requirements dictate a certain service level, which is often expressed in product availability rates if one speaks about stocks and points of sale, and delivery frequency if one speaks about stock replenishment and reordering policies. Given business requirements, a company has to make tradeoffs between stock keeping costs and transport costs, which are often translated into a complex spatial optimization problem at the company or supply chain level. Trends, such as decentralization, also play a role in logistics organization. The degree of centralization will be influenced by future changes in the relative costs of logistics inputs and the evolution of management practices in this field (McKinnon, 2009). Trends in increasing international trade and transport, requirements for high quality of logistics, as well as enabling technologies such as IT technology led companies to continuously optimize their distribution networks (Ruijgrok and Tavasszy, 2007).

The subject of supply chain optimization, however complex, is very well studied. The supply chain problem in its most generic form, has to determine production or sourcing locations, warehousing and distribution locations and the paths of goods from production via warehousing or directly to the given consumption locations. This generic supply chain optimization problem is computationally complex and there is a plethora of knowledge on how to deal with it, for instance, Melo et al. (2009) review 139 literature publications on the subject. However, this wide body of knowledge concerns optimization of individual companies or supply chains, where the central element is that there is an entity which has the full control on the design of the chain.

When it comes to modeling of freight transport and logistics at the regional level, the most knowledge is accumulated in the field of micro models. Indeed, having such a large body of knowledge on supply chain optimization, it is logical to apply elements of this knowledge at the regional level as well. If one has a complete knowledge on all companies in the region under consideration, the models that reproduce each company would perform reasonably well in respect to representation of processes, and eventually the transport flows that the businesses generate at the regional level. Obviously no one possesses such knowledge over all or nearly all businesses in the region. Therefore, researchers compensate for this deficiency by relaxing

the requirements for the models: the businesses' populations are generated together with their trade and transport needs (e.g. Boerkamps, 1999 and Liedtke, 2009a), or concentrating on the modeling of limited population sizes or specific economy sectors (e.g. Friedrich, 2010), or removing the requirement for empirical validity and concentrating on specific commodities (e.g. Maurer, 2008). Overall, validity for broader population using micro level models is not shown / proven.

The aggregate choices approach is the opposite to the disaggregate choices approach. In aggregate models the choices are based not on the decisions of individual companies or controllers, but mostly deal with generalized utilities of the choices. The SMILE / SMILE+ model (Tavasszy et al., 1998) and the EUNET 2.0 (Williams et al., 2005) are the most interesting and relevant applications of the aggregate choice modeling. Being simple in their essence, the researchers working on this type of models are tempted to make them as rich and detailed as possible, as it is the case for SMILE and, to a lesser degree, for UENET 2.0 too. The richness of these models, for example, is based on the real world knowledge that different commodity groups are subject to different distribution structures (the SMILE model distinguishes 50 logistics families). However, there is no good observational flow data to substantiate these details, which leads to the situation where the models are not calibrated nor validated at the flow level: these models are calibrated on the total number of tons pushed through the logistics system, but no calibration is done at the level of flow (origin and destinations of the goods flows). So there remains a gap in the knowledge: validity of aggregate logistics flow models at the O/D level.

The logistics model design discussed in this book relies on two requirements. The first and apparent requirement is that it should be empirically valid in its description of flows. The second requirement concerns its usability for strategic policy study applications and readiness for embedding into more comprehensive modeling suites such as TRANS-TOOLS (TRANS-TOOLS, Deliverable 6 (2008)). The applicability for the policy studies requires that the model is based on (generalized) costs, describes all flows and has a spatial dimension. The scientific literature does not present cases that satisfy these requirements, thus the logistics chain model fills a scientific niche, at the same time presenting its applied value for its target group of users.

Mathematical model formulation

The logistics chain model takes trade flow as the input and generates transport flow as the output. The required model input in the form of commodity ton trade flows is not observed directly, at least in the Netherlands. Therefore, the model design consists of two coupled models. A gravity model is used to translate regional production and regional consumption into trade flows, and the logistics chain model is used for translation of the trade flows into transport flows.

The gravity model uses a simple formulation (3.4), where the flow is the product of the estimated flow production and flow attraction multiplied by an exponential function of the disutility (costs). The gravity model in this formulation can also be used for transport flow

estimations, which are shown in Chapter 5. The gravity model has only one parameter to estimate, the price sensitivity value. Applied to the transport flows, this parameter shows the distance decay. Also, conclusions can be drawn on transport price sensitivity. In the case that the model is used for trade flow estimation, the sensitivity parameter determines the spatial trade distribution, namely the impact of the costs and distance on the trade decisions.

The logistics chain model takes the trade flow matrix as the input and translates it into the transport flow matrix. The transport flow matrix is the sum of the three distinct flow types: direct production-consumption flow (PC flow), production-distribution (PD) and distribution-consumption (DC) flows. A two-level nested logit is applied to determine how the trade flow is realized: the top level choice is a binary one, it determines whether the flow will be direct or go via a distribution center. The nested choice is a multinomial one, it determines the share of distribution in each of the n regions. The LCM model itself is relatively easy to formulate and to program. However, there is a more difficult part, namely testing its empirical validity. In this thesis the empirical validity of the model is shown by matching the model's output to the observational data, it is mostly done by showing the quality of fit. Therefore, a calibration procedure is required to find the optimal (or rather 'true') values of the model parameters.

A combination of a gravity and logistics chain model is a complex and non-linear system in its behavior in respect to the model parameters. The author is not aware of a method that would find the optimal values of the model parameters in reasonable time. Therefore, a heuristic was used to find the optimal values of the model parameters to ensure the best quality of fit between the generated model output and observed data. The heuristic used in this book is the iterative grid search procedure. An optimal value of a variable is determined by assessing the model output for a range of plausible variable values. The variable value that produces the minimum squared error between the output and observed data is set to be an optimal value. The process is repeated for all model variables. The process needs to be repeated a couple of times for all variables (in practice 5-10 iterations) for the model values to stabilize around their optimal values. The tests have shown that the resulting optimal variable values do not depend on the order in which the variables are estimated, thus providing an indirect evidence that the calibration procedure does not get caught in the local optima.

Data requirements and data availability

The logistics model needs the trade flow (or production-consumption flow) and cost values as the input. If a condition of empirical validity is added, then the model's output has to be tested against empirical data. The model output is in the form of the transport flow matrix, hence, an observation based transport matrix has to be available too. Finally, the choices in the model are cost based, therefore the cost data have to be obtained too. In short, the logistics chain model needs the following three classes of data to be available

1. Trade flow matrix or production-consumption matrix
2. Transport flow matrix, which is consistent with the trade flow matrix
3. Data on transport, distribution and other cost components

In practice, only transport flow matrices are available from the empirical sources in Europe, such as the national statistics bureaus and the Eurostat. These data have rather limited spatial resolution and do not provide any further indication on the purpose of the flow. An empirically based trade flow database at regional level is not available at all. The same is true for transport and warehousing costs: there are known ranges for these costs and ‘anecdotal’ evidence of some privately paid prices, but not a firm generalization of these costs.

The issue of data availability played a large role in the course of this PhD project. Statistics Netherlands has provided the transport flow data, extended with information on the purpose of transport, which comes from a variable on loading and unloading location types. These enhanced road transport data led to a breakthrough with respect to the model formulation, as the Dutch transport data contained not only the total aggregate transport flow data, but also the constituting sub-flows, such as production-distribution, distribution-consumption and production-consumption transport flows. The detailed Dutch transport data have led the author to the realization that the transport data could also be used for the estimation of the absent trade data in a gravity model. Thus, the available transport data lead to a synthetic availability of the trade flow data.

The third data class on transport, warehousing and region-specific costs has partly been observed and partly estimated endogenously within the model. The related model variables have been estimated in the calibration procedure and based on the realistic low and upper boundary values for these variables. The model is normalized by a fixed value of the variable for average production-distribution flow transport costs and shipment sizes. Therefore, the data on transport flows extended by the information on loading and unloading location types has led to the definite model design, where all three classes of data can be made available.

The reliance on the location variable in the Dutch road transport statistics data raised the questions on the data quality. Part of the Chapter 4 is devoted to the question on the quality of the location variable. Apart from a conclusion on sufficiency of the quality, the work produced an interesting result with respect to matching annual regional distribution and warehousing throughput volumes with employment in certain economy sectors. Employment data can be used as the basis for freight trip and flow generation models, as we provide a linear relationship between employment and volume of transport. The analysis does not establish causality, but it might help with distribution modeling for those countries where there is no data available on the loading and unloading location types.

In the course of this thesis, German and European data have become available, making the model applications possible on these data. The German and European model application have been somewhat simpler since the respective data contain both trade and transport flow classes of data. Thus, the model only had to be calibrated to estimate the costs, regional attractiveness and sensitivity variables.

The German data relates to the food distribution in Germany at NUTS2 spatial resolution level, containing information on interregional flows between 41 German regions. The European data described the flows between 256 European regions (at the NUTS2 spatial resolution level), thus making the model application computationally more challenging, due

to a relatively large datasets and the number of logistics chain combinations. Nonetheless, the logistics chain model performed well on these data too, confirming that the model designed for the Dutch data constraints can be applied in an international setting.

Model implementation, estimation and application

The emphasis on the empirical validity of the logistics model presented in this thesis has led to an extensive argumentation in Chapter 5 on the quality of the model's output measured as the fit between estimated flows and distribution throughput volumes on the one hand, and real world data on the other hand. Chapter 5 shows that the model's output fits for different types of flows and experimentation purposes.

Two alternative LCM's were specified, which both produced a good result with respect to the fit of the generated interregional transport flows. The main formulation uses transport costs directly as the main decision variable determining disutility of the logistics chain choices, while the alternative formulation determines the shipment sizes and derives relative transport costs from them. The two formulations allow estimations of transport and warehousing costs on the one hand, and shipment sizes (transport batch sizes) on the other hand. The author nonetheless deems the main model formulation to be more suitable for further implementation and (policy) scenario analysis, as it works directly with the logistics costs. Questions like logistics sprawl, spatial reorganization of logistics facilities and price sensitivities of certain transport and logistics related issues can be dealt with more directly by the main model formulation.

Applications of the separately calibrated gravity model have shown that different types of transport show a different demand price elasticity, for instance, production-distribution (PD) transport flows are less sensitive to the price signals than production-consumption (PC) transport flows. This observation allows a claim that indeed, modeling of logistics is an important endeavor, as transport flows have their logistics-related purpose; and flow properties depend on the flow's purpose. In other words, lumping all transport flows into one OD matrix hides important complexities and properties that are present in logistics. The logistics chain model is a better way of capturing the divergent transport flow purposes in trade-transport flow reconciliation.

In addition to the gravity model transport price elasticity estimations, the LCM model can also provide estimations on the total transport flow elasticity. The LCM model estimations are conceptually better, since they also include the spatial redistribution of trade, e.g. changes in the sourcing decisions by the companies. To our knowledge, the LCM model is the first empirical quantitative model, which provides an estimation for the warehousing and distribution service demand price elasticity.

The aspect of the policy-related usability of the LCM model demonstrated in the case of the so-called 'logistics sprawl', a phenomenon observed in many large population areas, where logistics facilities tend to spread out from concentrated clusters, dominant in the 70ies to an overall presence in our days. The LCM model has shown an ability to deal with the spatial

distribution of the warehouses by assessing quantitatively actions needed for dealing with the phenomenon. Due to the fact that the model is calibrated on empirical data, shows a good estimation quality and plausible responses in scenarios, the LCM model is a promising scenario analysis tool in the field of spatial location of warehouses and distribution facilities.

Samenvatting

Het onderzoek dat in dit proefschrift wordt gepresenteerd is gestart met twee simpele vragen. De eerste vraag was: "hoe kunnen handelsstromen worden vertaald naar transportstromen". Hierbij moet worden begrepen dat handel is niet gelijk aan het vervoer, wat het gevolg is van feit dat sommige goederen worden op- en overgeslagen bij magazijnen, distributiecentra en consolidatiecentra. Deze vraag is gerelateerd aan de interregionale stromen, die optreden op regionaal, of in de modellenterminologie, macroniveau van abstractie. Opslag en distributie daarentegen zijn activiteiten die worden uitgevoerd door individuele bedrijven, en dus plaatsvinden op het microniveau van abstractie. Vandaar de tweede vraag van dit onderzoek: "hoe kan de micro-macro kloof worden overbrugd in het domein van logistieke modellen", wat betekent dat het model in staat moet zijn om de 'business logica' die de bedrijven hanteren te volgen in het ontwerp en optimalisatie van hun logistieke ketens.

Dit proefschrift geeft een antwoord op de eerste vraag en laat het zien dat het 'Logistieke Keten Model' (LCM) de handelsstromen naar vervoerstromen vertaalt, de locaties inschat en de doorvoer van de magazijnen en distributiecentra met een goede precisie kan bepalen. Het model bepaalt ook op een passende wijze de transportstromen in relatie tot de waargenomen transportstromen. De tweede vraag wordt ook beantwoord. Het LCM toont aan dat een eenvoudig logistiek kost-gedreven macro-model voldoende nauwkeurig talloze processen en beslissingen genomen op het niveau van individuele bedrijven kan reproduceren. Er is zo geen noodzaak voor een zeer gedetailleerde modellering, als men geïnteresseerd is in een beschrijvend model voor het regionale logistieke systeem. Als de primaire interesse is in hoe een bepaalde sector werkt, dan heeft een gedetailleerde modellering wellicht de voorkeur, zoals bijvoorbeeld in het model voor de Duitse levensmiddelenbranche (Friedrich, 2010), dat zich bezighoudt met microniveau-actoren om macroniveau-effecten te schatten.

In hoofdstuk 1 zijn vijf onderzoeksvragen geformuleerd (zie paragraaf 1.2 voor meer details). De volgende lijst herhaalt de onderzoeksvragen en beoordeelt de onderzoeksresultaten van dit proefschrift.

1. **Onderzoeksvraag 1: Hoe kunnen logistieke beslissingen van individuele bedrijven worden vertaald naar de totale stromen op macroniveau?** Het LCM-model kijkt niet naar de beslissingen die op het bedrijfsniveau plaatsvinden. In plaats daarvan werkt het op het niveau van de geaggregeerde beslissingen: het model schat in welk deel van de beslissers een bepaalde keuze heeft gemaakt. Vervolgens wordt dit geschat aandeel vertaald naar de transportstromen; dit is de kans dat een beslisser een bepaalde keuze zou maken. Daarom bepaalt het LCM-model de geaggregeerde transportstromen en niet het gedrag van individuele beslissers.
2. **Onderzoeksvraag 2: Hoe zijn supply chains beïnvloed door de kosten van vervoer en opslag en hoe kunnen veranderingen in deze kosten worden vertaald naar veranderingen in de interregionale goederenstromen?** Het LCM-model behandelt logistieke kosten op een expliciete wijze, omdat de keuze van de logistieke alternatieven wordt bepaald door de kosten van vervoer en opslag. Vertaling van de handelsstromen naar transportstromen geeft de voorkeur aan logistieke ketens waar de kosten lager zijn. Dus de kosten in combinatie met de prijsgevoeligheidsparameters bepalen de kans dat een bepaalde logistieke keten wordt gebruikt en vervolgens het aandeel van het handelsverkeer dat wordt vervoerd via de keten in kwestie.
3. **Onderzoeksvraag 3: Hoe kan een freight generation model worden ontworpen en geïmplementeerd met als doel het inschatten van empirisch geldige transportstromen die nodig zijn om de goederen te vervoeren van productie- naar consumptielocaties, met nadruk op een correcte modellering van stromen die via magazijnen en distributiecentra gaan?** Het LCM schat de drie soorten van transportstromen in: directe productie-consumptie stroom (PC), productie-distributie stroom (PD) en de distributie-consumptie stroom (DC). De resulterende totale transportstromen (OD) op de infrastructurele netwerken is de som van deze drie substromen. De gegenereerde transportstroom is structureel niet gelijk aan de handelsstroom en houdt rekening met het vervoer via distributiefaciliteiten en magazijnen.
4. **Onderzoeksvraag 4: Wat zijn de benodigde data voor het logistieke model van onderzoeksvraag 3 en wat zijn de beschikbaarheid, kwaliteit en actualiteit van deze data?** Het LCM model heeft de drie klassen van data nodig, waarvan twee klassen behoren tot de model-input: P / C handelsstroomdata en data over de logistieke kosten. De derde klasse van data heeft betrekking op de kalibratie van het model, namelijk de transportstroom OD data. In praktijk heeft de auteur van dit proefschrift geen betrouwbare data voor de model-input gevonden; in plaats daarvan worden de handelsstromen en kostendata endogeen geschat in het model. Dit is mogelijk gemaakt door de derde klasse van data (OD transportstromen), die in principe uitsluitend worden gebruikt voor de kalibratie. Echter, vanwege de rijkdom van de transportstromen data is het mogelijk de handelsstromen in te schatten op basis van deze data. De logistieke kosten (modelparameters) worden endogeen geschat in de kalibratierun. De OD-transportstromen worden jaarlijks geactualiseerd door het CBS.

- Onderzoeksvraag 5: Hoe kan een beleidsmaatregel worden gekwantificeerd in een modeltoepassingscase met het doel om een bepaalde mate van verandering in de ruimtelijke organisatie van distributiefaciliteiten te bereiken?** Dit proefschrift presenteert een modeltoepassingscase voor de kwestie van de *logistics sprawl* in Nederland. Er worden mogelijke interventie maatregelen gepresenteerd in de vorm van een geld-gebaseerde verandering in de aantrekkelijkheid van een regio voor het omgaan met de logistieke wildgroei. De prijselasticiteit van de vervoervraag wordt gepresenteerd, alsmede de prijselasticiteit van de vraag naar de distributie- en warehousingdiensten, bij wijze van nieuwe schatting.

Dit hoofdstuk trekt verder conclusies over de beschikbare en relevante kennis en onderzoeksresultaten, model design keuzes en beschikbare data, de kwaliteit van de modelschatting en bespreekt de toepassingen van het model. Het wordt geconcludeerd met een overzicht van de meest veelbelovende gebieden voor verder onderzoek en ontwikkeling.

Beschikbare kennis

De traditionele goederenvervoermodellen, historisch afgeleid van de personenvervoermodellen, kunnen niet goed de logistieke structuren - de weg die vracht volgt van productie tot consumptie - modelleren. De discrepantie tussen handels- en transportstromen werd traditioneel verholpen met behulp van een vermenigvuldigingsfactor, wat zorgde voor een overschot aan transport wegens de overslag- en distributieactiviteiten, dus transport was structureel gelijk aan handel. Deze praktijk van modellering is niet in staat om de activiteiten en de complexiteit van de logistieke systemen vast te leggen; dit leidt tot onjuiste schattingen van de transportstroom en conceptueel zwakke modellering. Hieruit ontstaat de noodzaak om logistiek goed in goederenvervoermodellen op te nemen.

Logistieke activiteiten worden beïnvloed door de eisen vanuit de bedrijfsomgeving en basis kostenafwegingen. De bedrijfseisen dicteren een bepaald serviceniveau, dat vaak wordt uitgedrukt in de kosten van de beschikbaarheid van producten als men spreekt over de voorraden en de punten van verkoop; en de leveringsfrequentie als men spreekt over bevoorrading- en orderbeleid. Vanwege de bedrijfseisen moet een bedrijf een afweging maken tussen voorraadkosten en de kosten van vervoer, die vaak wordt vertaald naar een complex ruimtelijk optimalisatieprobleem bij het bedrijf of op het niveau van de supply chain. Trends zoals decentralisatie spelen ook een rol in de logistieke organisatie. De mate van centralisatie zal worden beïnvloed door de toekomstige veranderingen in de relatieve kosten van logistieke processen en de evolutie van de managementpraktijken op dit gebied (McKinnon, 2009). Toenemend volume van internationale handel en transport, eisen voor hoge kwaliteit van de logistieke operaties, evenals ondersteunende technologieën zoals IT-technologie hebben geleid tot de situatie waarin de bedrijven continu hun distributienetwerken optimaliseren (Ruijgrok en Tavasszy, 2007).

Het onderwerp van supply chain optimalisatie is uitvoerig bestudeerd. Het probleem van de supply chain optimalisatie, in de meest algemene vorm, moet een antwoord geven voor de optimale locaties van productie of inkoop, opslag en distributie en de stromen van goederen

van productie, eventueel via opslag of distributie, naar de opgegeven consumptielocaties. Deze generieke ketenoptimalisatie is een wiskundig complex probleem en er is een overvloed aan kennis over hoe om te gaan met dit probleem. Bijvoorbeeld, Melo et al. (2009) bestudeert 139 literatuurpublicaties over het onderwerp. Echter, deze hoeveelheid aan kennis betreft de optimalisatie van individuele bedrijven of ketens, waarbij de centrale gedachte is dat er een entiteit aanwezig die in staat is om de volledige controle over de logistieke keten uit te oefenen.

Als het gaat om het modelleren van het goederenvervoer en logistiek op regionaal niveau is de meeste kennis opgebouwd op het gebied van micromodellen. Inderdaad, is het logisch om met deze grote hoeveelheid aan kennis van supply chain optimalisatie de delen van deze kennis ook toe te passen op het regionaal niveau. Als men volledige data en kennis over alle bedrijven in de regio in kwestie heeft, zouden de modellen die elk bedrijf reproduceren redelijk goed kunnen presteren met betrekking tot de representatie van de processen, en uiteindelijk de vervoersstromen die de bedrijven op het regionaal niveau genereren. Uiteraard beschikt niemand over die kennis over alle of bijna alle bedrijven in de regio. Daarom compenseren onderzoekers dit tekort door een versoepeling van de eisen voor de modellen: de samenstelling van de bedrijven wordt samen gegenereerd met hun handel en transportbehoeften (bijvoorbeeld Boerkamps, 1999 en Liedtke, 2009a), of de modellen worden gericht op de modellering van de beperkte populatie van bedrijven of specifieke economische sectoren (bijvoorbeeld Friedrich, 2010), of de eis van empirische validiteit wordt opgegeven (bijvoorbeeld Maurer, 2008). Al met al is de geldigheid van microniveau modellen voor bredere toepassing niet aangetoond of bewezen.

De aanpak middels geaggregeerde keuzes is het tegenovergestelde van de gedesaggregeerde keuzes aanpak. In geaggregeerde modellen zijn de keuzes niet gebaseerd op de beslissingen van individuele bedrijven of entiteiten, maar worden meestal bepaald door algemene utiliteiten van de keuzes. Het SMILE / SMILE+ model (Tavasszy et al., 1998) en het EUNET 2.0 (Williams et al., 2005) zijn de meest interessante en relevante voorbeelden van de geaggregeerde keuze- modellering. Omdat deze modellen eenvoudig in hun essentie zijn, worden onderzoekers, die aan dit soort van modellen werken verleid om ze zo rijk en gedetailleerd mogelijk te maken, zoals het geval is voor het SMILE en, in mindere mate, voor het EUNET 2.0. De rijkdom van deze modellen, bijvoorbeeld, is gebaseerd op de observatie dat verschillende goederensoorten onderworpen zijn aan verschillende logistieke structuren (het SMILE model onderscheidt 50 logistieke families). Er zijn echter geen goede waargenomen stroomdata op dit detailniveau: de modellen zijn niet gekalibreerd of geverifieerd op het herkomst-bestemmingsniveau. Deze modellen zijn maximaal gekalibreerd op het totale aantal ton dat is doorgezet door het logistieke systeem, maar er is geen kalibratie gedaan op het niveau van de stroming (herkomst en bestemming van de goederenstromen). Zo blijft er een gat in de kennis, namelijk de ontbrekende geldigheid van de geaggregeerde keuze-modellen op het transport O/D niveau.

Het ontwerp van het logistieke model zoals in dit boek wordt besproken is gebaseerd op twee vereisten. De eerste vereiste is dat het logistiek model empirisch geldig is qua beschrijving van transportstromen. De tweede eis heeft betrekking op de bruikbaarheid voor de strategische beleidsstudies. Het model moet ook geïntegreerd kunnen worden in de grotere

modellen zoals TRANS-TOOLS (TRANS-TOOLS, Deliverable 6 (2008)). De toepasbaarheid voor de beleidsstudies vereist dat het model is gebaseerd op de (totale) kosten, het alle stromen beschrijft en een ruimtelijke dimensie heeft. De wetenschappelijke literatuur kent geen modellen die aan al deze eisen voldoen, waardoor dit logistieke ketenmodel een wetenschappelijke niche vult.

Wiskundige model formulering

Het logistieke ketenmodel gebruikt handelsstromen als input en genereert transportstromen als output. De vereiste modelinput in de vorm van ton handelsstromen wordt niet direct waargenomen, althans niet in Nederland. Daarom bestaat het modelontwerp uit twee gekoppelde sub-modellen. Een zwaartekrachtmodel wordt gebruikt om de regionale productie en regionale consumptie te vertalen naar de handelsstromen, het logistieke ketenmodel wordt vervolgens gebruikt voor de vertaling van de handelsstromen naar de transportstromen.

Het zwaartekrachtmodel gebruikt een eenvoudige formulering (3.4), waarbij de stroom het product is van de geschatte stroom productie en stroom attractie, vermenigvuldigd met een exponentiële functie van de *disutility* (kosten). Het zwaartekrachtmodel in deze formulering kan ook worden gebruikt voor de transportstroomschattingen die weergegeven zijn in hoofdstuk 5. Het zwaartekrachtmodel heeft slechts één parameter te schatten, namelijk de waarde van de prijsgevoeligheid. Toegepast op de transportstromen geeft deze parameter het afstandsverval. Ook kunnen conclusies worden getrokken over de prijsgevoeligheid van het wegvervoer. Als het model wordt gebruikt voor de schatting van de handelsstroom bepaalt deze parameter de ruimtelijke distributie van handel, namelijk het effect van de kosten en afstand op de handelsbeslissingen.

Het logistieke ketenmodel gebruikt de handelsstroommatrix als de input en vertaalt deze naar de transportstroommatrix. De transportstroommatrix is de som van de drie aparte soorten stromen: die van de directe productie-consumptie (PC-stroom), productie-distributie (PD) en de distributie-consumptie (DC). Een genest logit op twee niveaus wordt gebruikt om te bepalen hoe de handelsstroom wordt gerealiseerd: de hoogste niveau keuze is binair en bepaalt of de stroom direct of via een distributiecentrum gaat. De geneste keuze is multinomiaal en bepaalt het aandeel van de distributieactiviteiten in elk van de n regio's. Het LCM model zelf is relatief eenvoudig geformuleerd en geprogrammeerd. Er is echter een moeilijker deel, namelijk het testen van de empirische geldigheid. In dit proefschrift wordt de empirische geldigheid van het model getoond door de matching van de model output met de waarnemingsgegevens, meestal in de vorm van de kwaliteit van de fit. Daarom wordt een kalibratieprocedure vereist om de optimale (of liever "true") waarden van de modelparameters te vinden.

Een combinatie van de zwaartekracht- en logistieke ketenmodellen is een complex en niet-lineair systeem ten opzichte van de modelparameters. De auteur kent geen methode die de optimale waarden van de modelparameters in redelijke tijd kan vinden. Daarom werd een heuristische methode gebruikt om de optimale waarden van de modelparameters te vinden voor de beste kwaliteit van de fit tussen de gegenereerde model-output en de waargenomen

data. De heuristische methode die in dit boek is gebruikt betreft de iteratieve *grid search* procedure. Een optimale waarde van een variabele wordt bepaald door het beoordelen van de model output voor een reeks van plausibele variabelwaarden. De variabelwaarde die de minimale kwadratische fout tussen de model-output en de waargenomen data produceert, is beschouwd als een optimale waarde. Het proces wordt herhaald voor alle modelvariabelen. Het proces moet een aantal keer worden herhaald voor alle modelvariabelen (in de praktijk 5-10 herhalingen) voordat de variabelen rond hun optimale waarden stabiliseren. De testen hebben aangetoond dat de resulterende optimale variabelwaarden niet afhankelijk zijn van de volgorde waarin de variabelen worden geschat, waardoor een indirect bewijs wordt verkregen dat de kalibratie niet in de lokale optima belandt.

Vereisten en beschikbaarheid data

Het logistieke model heeft de handelsstroom (of productie-consumptie stroom) en de kostendata nodig als input. Als een eis van empirische geldigheid van het model wordt toegevoegd moet de output van het model worden getoetst aan empirische data. De model output is in de vorm van de transportstroommatrix, waarvoor een op waarneming gebaseerde transportmatrix ook beschikbaar moet zijn. Tenslotte zijn de keuzes in het model gebaseerd op de kosten, daarom moet de kostendata ook verkregen worden. Kortom, het logistieke ketenmodel heeft de volgende drie klassen van data nodig:

1. Handelsstroommatrix of productie-consumptiematrix
2. Transportstroommatrix, die in overeenstemming is met de handelsstroommatrix
3. Gegevens over transport-, distributie- en andere kostencomponenten

In de praktijk zijn alleen transportstroom matrices verkrijgbaar bij de empirische bronnen in Europa, zoals de nationale statistische bureaus en Eurostat. Deze data hebben een vrij beperkte ruimtelijke resolutie en geven geen enkele verdere indicatie over het doel van het transport. Een empirisch gebaseerde handelsstroom-database op regionaal niveau is niet beschikbaar. Hetzelfde geldt voor de transport- en distributiekosten: er zijn bekende ranges voor deze kosten en enkele 'anekdotische' waarden, maar geen een betrouwbare generalisatie van deze kosten.

De kwestie van databeschikbaarheid heeft een grote rol gespeeld in de loop van dit promotieonderzoek. Het CBS heeft de transportstroomdata aangevuld met informatie over het doel van het vervoer, wat afkomstig is van de locatietypevariabele voor laad- en loslocaties. Deze aangevulde transportstroomdata hebben tot een doorbraak met betrekking tot de modelformulering geleid, omdat de Nederlandse transportdata niet alleen de totale stroomdata bevatten, maar ook de deelstromen, zoals productie-distributie, distributie-consumptie en productie-consumptie vervoersstromen. De gedetailleerde Nederlandse transportdata hebben de auteur tot het besef gebracht dat de transportdata ook kunnen worden gebruikt voor de schatting van de ontbrekende handelsdata in een zwaartekrachtmodel. Dus de beschikbare vervoerdata leiden tot de synthetische data over de handelsstromen.

De derde dataklasse over transport-, distributie- en regio-specifieke kosten is gedeeltelijk waargenomen en gedeeltelijk endogeen geschat binnen het model. De gerelateerde

modelvariabelen zijn bij de kalibratieprocedure geschat in de realistische range van beneden- en bovengrenswaarden voor deze variabelen. Het model wordt genormaliseerd door een vaste waarde van de variabele over de gemiddelde productie-distributie transportkosten en zendinggrootte. Op deze wijze hebben de data over transportstromen aangevuld met de informatie over de laad- en loslocaties geleid tot het definitieve modelontwerp, waarbij alle drie klassen van gegevens beschikbaar kunnen worden gemaakt.

De afhankelijkheid van de locatievariabele in de Nederlandse wegvervoerdata leidde tot de vragen over de kwaliteit van de locatiedata. Een deel van hoofdstuk 4 is gewijd aan de vraag over de kwaliteit van de locatievariabele. Afgezien van een conclusie over het voldoen van de datakwaliteit heeft het werk een interessant resultaat opgeleverd met betrekking tot de match tussen de jaarlijkse regionale distributie-doorvoervolumes en de werkgelegenheid in bepaalde economische sectoren. Werkgelegenheidscijfers kunnen worden gebruikt als een basis voor *freight trip and flow* generatie modellen, omdat er is een lineair verband tussen de arbeid en het transportvolume. De analyse kan niet een causaliteit vaststellen, maar het zou kunnen helpen met de distributie-modellering voor die landen waar er geen data beschikbaar zijn over de laad- en los-locatietypes.

In de loop van dit onderzoeksproject zijn Duitse en Europese data beschikbaar gekomen, waardoor modeltoepassingen met deze data mogelijk werden. De Duitse en Europese modeltoepassingen zijn iets eenvoudiger geweest omdat de desbetreffende data beide handels- en transportstroomklassen van data bevatten. Het model hoefde slechts gekalibreerd worden om de kosten, regionale aantrekkelijkheid en de gevoeligheidsvariabelen te schatten.

De Duitse data hebben betrekking op de voedingsmiddelendistributie in Duitsland op het NUTS -2 ruimtelijke resolutie niveau, en omvat informatie over interregionale stromen tussen 41 Duitse regio's. De Europese data beschrijven de stromen tussen 256 Europese regio's (in de NUTS-2 classificatie), waardoor de modeltoepassing rekentechnisch meer uitdagend werd vanwege de relatief grote datasets en het aantal logistieke ketencombinaties. Het logistieke ketenmodel heeft ook goed gepresteerd op deze data. Dit bevestigt dat het model, dat voor de Nederlandse databeperkingen werd ontworpen, ook in een internationale setting kan worden toegepast.

Modelimplementatie, schatting en toepassing

De nadruk op de empirische geldigheid van het logistiek model in dit proefschrift heeft geleid tot een uitgebreide argumentatie in hoofdstuk 5 over de kwaliteit van de output van het model, gemeten als de fit tussen de geschatte stromen en distributie-doorzetvolumes enerzijds en de waargenomen datasets aan de andere kant. Hoofdstuk 5 bespreekt de output van het model in de termen van fit kwaliteit voor verschillende stromen en experimenten.

Twee alternatieve logistieke ketenmodellen (LCM's) werden gespecificeerd, die beide een goed resultaat met betrekking tot de fit van de gegenereerde interregionale transportstromen opleveren. De formulering van het hoofdmodel gebruikt transportkosten als de belangrijkste beslissingsvariabele die de *disutility* van de logistieke keten bepaalt, terwijl de alternatieve

formulering de zendinggrootte bepaalt en daaraan de relatieve transportkosten ontleent. De twee formuleringen schatten de transport- en distributiekosten aan de ene kant en de zendinggrootte (transport seriegroottes) aan de andere kant. De auteur is van mening dat de hoofd modelformulering meer geschikt is voor een verdere implementatie en (beleid) scenarioanalyse, omdat het direct werkt met de logistieke kosten. Vraagstukken zoals *logistics sprawl*, ruimtelijke reorganisatie van logistieke faciliteiten en de prijsgevoeligheden kunnen direct worden behandeld door het toepassen van de hoofdmodelformulering.

Toepassingen van het afzonderlijk gekalibreerd zwaartekrachtmodel hebben aangetoond dat de verschillende vormen van vervoer verschillende vraagprijselasticiteiten laten zien. Bijvoorbeeld, de productie-distributie (PD) transportstromen zijn minder gevoelig voor de prijssignalen dan de productie-consumptie (PC) transportstromen. Deze observatie leidt tot een conclusie dat modellering van logistiek inderdaad een belangrijk streven is, omdat de vervoersstromen logistiek gerelateerde doeleinden hebben en transportstroomeigenschappen afhankelijk zijn van het doel van de stroom. Met andere woorden: opsomming van alle transportstromen tot één HB-matrix verbergt belangrijke complexiteiten en eigenschappen die aanwezig zijn in de logistiek. Het logistieke ketenmodel is een betere manier van het vastleggen van de uiteenlopende doeleinden van transportstromen in de vertaling van handelsnaar transportstromen.

In aanvulling op het zwaartekrachtmodel en de transportprijselasticiteitsschattingen kan het LCM -model ook de schattingen van de totale transportstroomelasticiteit maken. De schattingen van het LCM-model zijn conceptueel beter dan die van zwaartekrachtmodel, omdat ze ook de ruimtelijke herverdeling van de handel meenemen, bijvoorbeeld veranderingen in de inkoopbeslissingen van de ondernemingen. Voor zover bekend, is het LCM-model het eerste empirische kwantitatieve model dat een schatting van vraagprijselasticiteit voor opslag- en distributiediensten biedt.

Het aspect van de bruikbaarheid van het LCM voor de beleidsstudies is aangetoond in de case van *logistics sprawl*, een fenomeen waargenomen in veel grote bevolkingscentra, waar logistieke faciliteiten een neiging hebben om zich te verspreiden uit geconcentreerde clusters (dominant in de jaren 70) tot een overal aanwezigheid tegenwoordig. Het LCM-model heeft een functionaliteit om de ruimtelijke verdeling van de distributiecentra te bepalen door kwantitatieve beoordeling van acties die nodig zijn om met dit fenomeen om te gaan. Omdat het model wordt gekalibreerd op empirische data geeft het een kwalitatief goede schatting en toont het plausibele reacties in scenario's. Hiermee is het LCM-model een veelbelovend instrument voor scenario-analyse op het gebied van ruimtelijke locatiekeuzes van logistieke ketens (warehouses, crossdock- en distributiecentra).

References

- Anand N., Quak H., Duin van R., Tavasszy L. (2012), City logistics modeling efforts: Trends and gaps-A review, *Procedia-Social and Behavioral Sciences*, 2012
- Anderson, J. E., A Theoretical Foundation for the Gravity Equation, *The American Economic Review*, Vol. 69, No. 1 (Mar., 1979), pp. 106-116
- Annema J.A., Hoen A., Turton H., Schrattenholzer L., Mudgal S., Fergusson M. (2006), Scientific review of TREMOVE – a European transport policy assessment model, Netherlands Environmental Assessment Agency (MNP) Report
- Bard, J. F., & Nananukul, N. (2009), Heuristics for a multiperiod inventory routing problem with production decisions. *Computers & Industrial Engineering*, doi:10.1016/j.cie.2009.01.020
- BCG (2007), *Surviving the China Rip Tide: How to Profit from the Supply Chain Bottleneck*, the Boston Consulting Group
- Beamon M.B.(1998), Supply chain design and analysis: Models and methods. *International Journal of Production Economics*, 55,28 1-294
- Ben-Akiva M., Meersman H., Voorde van de E. (2013), *Freight Transport Modelling*, ISBN: 9781781902851
- Bergkvist E., Westin L. (1998), Forecasting interregional freight flows by gravity models. Utilising OLS-, NLS- estimations and Poisson-, Neural Network- specifications, 38th Congress of the European Regional Science Association Vienna, Austria, August 28 - September 1, 1998. Available online at <http://www.diva-portal.org/smash/get/diva2:329548/FULLTEXT01.pdf> , accessed on 04-12-2014
- Bergkvist E. (2000), Forecasting interregional freight flows by gravity models. *Jahrbuch für Regionalwissenschaft*, 20(2), 133-148.
- Bergstrand, J. H. (1985), The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence, *The Review of Economics and Statistics*, Vol. 67, No. 3 (Aug., 1985), pp. 474-481

- Black W.R. (1972), Interregional Commodity Flows: Some Experiments With The Gravity Model, *Journal of Regional Science*, Volume 12, Issue 1, pages 107–118, April 1972
- Blumenfeld D.E., Burns L.D., Diltz D.J. (1985), Analyzing trade-offs between transportation, inventory and production costs on freight networks, *Transportation Research Part B: Methodological*, Volume 19, Issue 5, October 1985, Pages 361–380
- Boerkamps, J., van Binsbergen, A.,(1999), GoodTrip -- A New Approach for Modelling and Evaluation of Urban Goods Distribution, Urban Transport Conference, 2nd KFB Re-search Conference, Lund, Sweden.
- Bovenkerk, M., 2005. SMILE+, the new and improved Dutch national freight model system. Paper Presented at the European Transport Conference 2005, Strasbourg
- Burger, M., Van Oort, F., & Linders, G. J. (2009). On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4(2), 167-190.
- Cairncross, F., *The death of distance*, 1997, The Orion Publishing Group, ISBN 0-87584-806-0
- Campbell J.B., McKinnon A.C. (1997), Trends in UK Road Freight Transport. Draft Report prepared under the EU 4th Framework Project REDEFINE. Heriot-Watt University, Edinburgh
- CBS Population and population dynamics, 2013. CBS Online statistics <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLEN&PA=37943eng&LA=EN>
- CBS StatLine, statline.cbs.nl (23.03.2011)
- Chow J. Y. J., Yang C.H., Regan A.C. (2010), State-of-the art of freight forecast modeling: lessons learned and the road ahead, *Transportation*, November 2010, Volume 37, Issue 6, pp 1011-1030
- Chen, M. (2011), ETIS and TRANS-TOOLS v1 freight demand. Presented at the CTS-seminar on European and national freight demand models in Stockholm
- Coyle J.J, Langley C.H., Gibson B. and Novack R.A. (2008), *Supply Chain Management: A Logistics Perspective*, 8th edition, p.299, ISBN-13: 978-0324376920
- Combes, F. (2010) *The shipment size in interurban freight transport*. PhD dissertation, LVMT/ENPC, Paris.
- Combes, F. (2013), *Inventory Theory and Freight Transport Modelling* in Tavasszy, L. & de Jong, G. (2013), *Modelling Freight Transport*. Elsevier, ISBN: 978-0-12-410400-6
- ComExt intra- and extra-EU trade data, European University Institute, <http://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/ComExt.aspx>, Retrieved on 23-08-2013
- CROW (2007), Toepassing kengetallen goederenvervoer van en naar bedrijventerreinen, CROW-rapport 05-07, available online at <http://www.meob-omwonenden.nl/documenten/Rapport.pdf> , retrieved on 09-12-2014

- Dablanc, L., Rakotonarivo, D. (2010). The impacts of logistics sprawl: How does the location of parcel transport terminals affect the energy efficiency of goods' movements in Paris and what can we do about it? *Procedia - Social and Behavioral Sciences*, 2 (3), pp. 6087-6096.
- Dablanc, L., Warehousing development and logistics sprawl, comparing Paris and Atlanta, 4th METRANS National Urban Freight Conference Long Beach, Oct 12-16, 2011
- Dablanc L., Ogilvie S., Goodchild A. (2014) Logistics Sprawl: Differential Warehousing Development Patterns in Los Angeles and Seattle, TRB 2014 Annual Meeting, January 2014 Washington DC.
- Davydenko, I. (2006), Logistics network design for Access Business Group, Technische Universiteit Eindhoven, 2006, ISBN 978-90-444-0596-5
- Davydenko, I., Tavasszy, L. (2010), Approaches to logistics choice modeling in aggregate goods flow, TRAIL Congress, 2010 Rotterdam
- Davydenko, IY, Tavasszy, LA, Meijeren, JJ & Ivanova, O (2011). Estimation of freight flow and trip generation according to transport purpose in urban and regional logistics networks. METRANS Urban Freight Conference Proceedings (pp. 1-17). Long Beach: METRANS.
- Davydenko, Igor Y.; Tavasszy, Lóránt A.; Smeets, Peter S.G.M. (2012). Commodity Freight and Trip Generation by Logistics Distribution Centers Based on Sectorial Employment Data, 91st Annual TRB Meeting, Washington DC
- Davydenko I. Y., Tavasszy L.A., Quak H. (2013a), A quantitative model for exploration of logistics sprawl of the future, METRANS I-NUF conference paper, Oct. 2013 Long Beach CA
- Davydenko, I.Y., Tavasszy, L.A. (2013b), Estimation of Warehouse Throughput in a Freight Transport Demand Model for the Netherlands, presented at annual TRB meeting, January 2013, Washington DC.
- Davydenko I.Y., Tavasszy L.A., de Blois C.J. (2013c), Modeling interregional freight flow by distribution systems, 13th WCTR, July 15-18, 2013 – Rio de Janeiro, Brasil
- Davydenko, I.Y., Thissen, M., Tavasszy, L.A. (2014a), Translation of Trade Flows Into Transport Flows in European Logistics Facility Choice Model, Transportation Research Board (TRB) 93rd Annual Meeting (2014), Washington, D.C.
- Davydenko I., Ehrler V., Ree de D., Lewis A., Tavasszy L. (2014b), Towards a global CO2 calculation standard for supply chains: suggestions for methodological improvements, 93rd Annual Meeting of the Transportation Research Board (2014), Washington D.C.
- Davydenko I.Y., Zhang M., Tavasszy L.A. (2014c), CBS transport data analytics, TNO Report TNO 2014 R10747
- Davydenko I., Ehrler V., Ree de D., Lewis A., Tavasszy L. (2014d), Towards a global CO2 calculation standard for supply chains: Suggestions for methodological improvements, Transportation Research Part D: Transport and Environment, V.32, P. 362-372
- DBR 2013, Sustainable Accessibility of the Randstad Programme, <http://www.nwo.nl/en/research-and-results/programmes/magw/sustainable-accessibility-of-the-randstad/index.html> retrieved on 06-08-2014

- De Ceuster, G., L. Franckx, B. van Herbrugge n, S. Logghe, B. van Zeebroeck, S. Tastenhoye, S. Proost, J. Knockaert, I. Williams, G. Deane, A. Martino, D. Firello (2005), TREMOVE 2.30 Model and Baseline Description. Final report, 18 February 2005, Leuven: Transport and Mobility, www.tremove.org
- DHL (2012), Delivering Tomorrow: Logistics 2050, A Scenario Study, Deutsche Post AG, Bonn, Germany, ISBN 978-3-920269-54-2
- Disdier, A. C., & Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and statistics*, 90(1), 37-48.
- Drezner Z., Hamacher H.W., Facility Location: Applications and Theory, Springer, New York (2004)
- Duin van, J.H.R., Tavasszy, L.A., Quak, H.J. (2013). Towards E(lectric)-urban freight: First promising steps in the electric vehicle revolution. *European Transport - Trasporti Europei*, (54)
- Ekström, M., Sjödin, Å., & Andreasson, K. (2004). Evaluation of the COPERT III emission model with on-road optical remote sensing measurements. *Atmospheric Environment*, 38(38), 6631-6641.
- Erlander S., Steward N.F. (1990), *The Gravity Model in Transportation Analysis: Theory and Extensions*, ISBN 90-6764-089-1
- European Commission (2012), *Road Transport - A change of gear*, Luxembourg: Publications Office of the European Union, ISBN 978-92-79-22827-8, also accessible online http://ec.europa.eu/transport/modes/road/doc/broch-road-transport_en.pdf , 26-11-2014
- European Commission (2013), *EU ENERGY, TRANSPORT AND GHG EMISSIONS TRENDS TO 2050 REFERENCE SCENARIO 2013*
- European Commission Impact Assessment http://ec.europa.eu/smart-regulation/impact/index_en.htm, last retrieved on 28-08-2014
- European Union (2013), *EU transport in figures, statistical pocketbook 2013*, <http://ec.europa.eu/transport/facts-fundings/statistics/doc/2013/pocketbook2013.pdf> last retrieved 17-12-2014
- Eurostat (2008), *Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008) (NACE Rev. 2)*, accessible online at http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NACE_REV2&StrLanguageCode=EN&StrLayoutCode=HIERARCHIC , last accessed on 24-12-2014
- Eurostat (2011), *Comext Database*, epp.eurostat.ec.europa.eu/newxtweb/ , 11-03-2011
- Feldman E., Lehrer F. A., and Ray T. L. (1966), *Warehouse Location Under Continuous Economies of Scale*, *Management Science* 196612:9 , 670-684
- Fowkes, A. S., Whiteing, A. E., Johnson, D. H. (2007), *LEFT: The Right Approach to Aggregate Strategic Freight Transport Modelling?* Proceedings of 39th UTSG conference, 03.-05.01.07. Leeds.

- Friedrich, H. (2010), Simulation of logistics in food retailing for freight transportation analysis, PhD Dissertation, Karlsruhe Institute of Technology
- Friedrich H., Tavasszy L. and Igor Davydenko (2013), Distribution Structures in Tavasszy, L. & de Jong, G. (2013), Modelling Freight Transport. Elsevier, ISBN: 978-0-12-410400-6
- Gils van O.M., Huys O.M., Jong de B. (2009), De Nederlandse mainports onder druk, ISBN13 9789049100230
- Groenewout 2005, CAST,
http://www.groenewout.com/Definition/Details/499095/521690/1869_cast.aspx, accessed on 08-08-2014
- Harris F.W. (1913), How many parts to make at once, The Magazine of Management, Volume 10, Number 2, February 1913
- Higgins C.D., Ferguson M., Kanaroglou P.S., Varieties of logistics centres: developing a standardized typology and hierarchy, TRB 2012 Annual Meeting Paper revised, Washington DC.
- Holmblad P.M. (2004), A stylized PCOD-model for relations between zonal trade flows, transports, and distribution centers, ECTRI/FERSI Young Researchers Seminar, Lyon, France
- Huber S., Luft D., Klauenberg J., Thaller C. (2014), Integration of transport logistics hubs in freight transport demand modeling, European Transport Conference 2014
- Kim E., Park D., Kim C., Lee J. (2010), A New Paradigm for Freight Demand Modeling: The Physical Distribution Channel Choice Approach, International Journal of Urban Sciences. 14(3). 2010, 240-253
- Ivanova, O. (2014), Modelling Inter-Regional Freight Demand with Input-Output, Gravity and SCGE Methodologies, in: Tavasszy, L.A. and G. de Jong (eds.), Modelling Freight Transport, London: Elsevier
- Jong G. de, Gunn H., Walker W. (2004), National and International Freight Transport Models: An Overview and Ideas for Future Development, Transport Reviews: A Transnational Transdisciplinary Journal, Volume 24, Issue 1, 2004
- Jong G. de, Ben-Akiva M. (2007), A micro-simulation model of shipment size and transport chain choice. Transportation Research Part B 41, p. 950–965.
- Jong de, G. C., Schrotten, A., van Essen, H., Otten, M., & Bucci, P. (2010). The price sensitivity of road freight transport _ A review of elasticities. In E. van de Voorde, & Th. Vanelslander (Eds.), Applied transport economics, a management and policy perspective. Antwerpen: De Boeck.
- Jong de G.C.; Vierth I.; Tavasszy L.; Ben-Akiva M.E. (2012) Recent developments in national and international freight transport models within Europe, Transportation, . doi: 10.1007/s11116-012-9422-9
- Holguín-Veras, J. , E. Thorson, J.C. Zorilla (2008), Commercial Vehicle Empty Trip Models With Variable Zero Order Empty Trip Probabilities. Networks and Spatial Economics, p. 19.
- Holguín-Veras J., Jaller M., Sánchez-Díaz I., Wojtowicz J., Campbell S., Levinson H., Lawson C., Powers E.L., Tavasszy L. (2012), Freight Trip Generation And Land Use,

- NCHRP report 739 and NCFRP report 19. Available online at http://onlinepubs.trb.org/onlinepubs/ncfrp/ncfrp_rpt_019.pdf retrieved on 09-12-2014
- Lankhuizen, M.B.M., Th. de Graaff and H.L.F. de Groot (2012), Product heterogeneity in distance decay: the effect of multiple dimensions of distance across different product categories, mimeo VU University Amsterdam.
- Liedtke G., Ott A., Schmedding D. (2004), A Network-economical Analysis of the Freight Transport Market, Universität Karlsruhe, Institut für Wirtschaftspolitik und Wirtschaftsforschung (IWW)
- Liedtke G., Babani J., Ott A. (2006), Data organization pattern for microscopic freight demand models, 85th Annual Meeting of the Transport Research Board, Washington DC 2006
- Liedtke G. (2009a), Principles of micro-behavior commodity transport modeling, Transportation Research Part E 45 (2009) 795–809
- Liedtke G.T., Tavasszy L.A., Wisetjindawat W. (2009), A Comparative Analysis of Behavior-Oriented Commodity Transport Models, 88th Annual Meeting of the Transportation Research Board, January 11-15, 2009. Washington D.C.
- Ligterink, NE, Tavasszy, LA & Lange, R de (2012). A velocity and payload dependent emission model for heavy-duty road freight transportation. Transportation Research. Part D: Transport & Environment, 17(6), 487-491.
- Linders G.J., Blois de C., Boonstra H.J., Groot de H., Exel J. (2008), Integration of international trade and transport flow statistics for the Netherlands, Statistics Netherlands Voorburg / Heerlen, April 2008.
- Logicway 2014, Uitvoering CBS-enquete, <http://www.logicway.nl/nl/component/content/article/2-ongecategoriseerd/39-nieuwe-modulen> , retrieved on 07-11-2014
- Mallik S. (2010), "Customer Service in Supply Chain Management". In Hossein Bidgoil. The Handbook of Technology Management: Supply Chain Management, Marketing and Advertising, and Global Management, vol 2 (1 ed.). Hoboken, New Jersey: John Wiley & Sons. p. 104. ISBN 978-0-470-24948-2.
- Maurer, H. H.(2008), Development of an Integrated Model for Estimating Emissions from Freight Transport, PhD dissertation, The University of Leeds Institute for Transport Studies
- Mayer-Schönberger V., Cukier K. (2013), Big Data: A Revolution That Will Transform How We Live, Work, and Think.
- Meixell M.J., Gargeya V.B. (2005), Global supply chain design: a literature review and critique, Transportation Research Part E 41 (2005), 531-550
- Melachrinoudis E., Min H.(2007), Redesigning a warehouse network, European Journal of Operational Research, 176 (2007), pp. 210–229
- Melo M.T., Nickel S., Saldanha-da-Gama F. (2009), Facility location and supply chain management – A review, European Journal of Operational Research, Volume 196, Issue 2, 16 July 2009, Pages 401–412

- McKinnon A. (2009), The present and future land requirements of logistical activities, Land Use Policy 26S (2009) S293–S301
- NTG Beleid goederenvervoer,
<http://www.nationaletransportgids.nl/goederentransport/paginas/wegtransport/beleid-goederenvervoer.html> , accessed on 26-11-2014
- Quak, H.J., De Koster, M.B.M. (2009). Delivering goods in urban areas: How to deal with urban policy restrictions and the environment. *Transportation Science*, 43 (2), pp. 211-227.
- Orsini, D., O.Gavaud, and P.L. Bourhis. Logistics Facilities impacts on the territory. in 2009 European Transport Conference. 2009. London: AET
- Pattanamekar P., Park D., Lee K., Kim C. (2009), Estimating Multimodal OD Matrix from Mode Specific OD Matrices, 88th TRB Annual Meeting, January 11-15, 2009, Washington D.C.
- Port Statistics, Port of Rotterdam, 2010, http://www.portofrotterdam.com/en/Port/port-statistics/Documents/Port_Statistics_2009_tcm26-64785.pdf (23.03.2011)
- Pride W.M., Hughes R.J., Kapoor J.R. (2008), *Business*, p. 417, ISBN-13: 978-0324829556
- Rich P., Holmblad P.M., Hansen C.O. (2009), A weighted logit freight mode-choice model, *Transportation Research Part E* 45 (2009) 1006–1019
- Rijkswaterstaat (2012), *Kwaliteitskader Strategische Verkeers en Vervoermodellen, Rijkswaterstaat, Dienst Verkeer en Scheepvaart (DVS)*
- Rudberg M., Wikner J. (2004), Mass customization in terms of the customer order decoupling point, *Production Planning & Control: The Management of Operations* Volume 15, Issue 4, 2004, Special Issue: Special Issue Mass customization
- Ruijgrok C. J., Tavasszy L.A. (2007), The development of International Freight Transport in Europe as a result of developments in International trade and Logistics, paper prepared for the Canada's Asia-Pacific Gateway and Corridor Research Consortium, Vancouver, BC, May 2-4, 2007, http://www.gateway-corridor.com/roundconfpapers/documents/Ruijgrok_Tavasszy.pdf last accessed on 19-12-2014
- Sahraoui A., Jayakrishnan R. (2005), Microscopic-Macroscopic Models Systems Integration: A Simulation Case Study for ATMIS, *SIMULATION*, Vol. 81, Issue 5, May 2005 353-363
- Samimi A., Mohammadian A., Kawamura K. (2009), Behavioral freight movement modeling, Resource Paper, The 12th International Conference on Travel Behavior Research Jaipur, Rajasthan, India, December 13-18, 2009
- SCENES Consortium, 2000. SCENES European transport forecasting model and appended module, technical description. SCENES deliverable 4 for the European Commission DG-TREN, ME&P, Cambridge
- Significance and CE Delft. (2010). Price sensitivity of European road transport: Towards a better understanding of existing results. A Report for Transport & Environment, Significance, The Hague.
- SMILE+ Handboek (2004), Algemeen, 25 november 2004

- Statistisches Bundesamt (2006a), Außenhandel nach Ländern und Güterabteilungen der Produktionsstatistiken (Spezialhandel). March 2006
- Statistisches Bundesamt (2006b), Produktion im produzierenden Gewerbe Jahr 2005. Fachserie 4, Reihe 3.1. 2006. – URL www.destatis.de
- Taylor H.D. (1998), Economic Tradeoffs of Substituting Transportation for Inventory in the Department of Defense: A Case Study of Pipeline Reduction, Master Thesis, Virginia Polytechnic Institute and State University
- Tavasszy, L.A., Smeenk, B., Ruijgrok, C.J.(1998): A DSS for modelling logistics chains in freight transport systems analysis, *International Transactions in Operational Research* 50(6), pp. 447–459
- Tavasszy, L.A., E. Huijsman, C. Cornelissen (2001), Forecasting the impacts of changing patterns of physical distribution on freight transport flows in Europe, *Proceedings of the 9th World Conference for Transport Research*, Seoul
- Tavasszy, L. A. (2006), Freight Modelling – An Overview of International Experiences, paper prepared for the TRB Conference on ‘Freight Demand Modelling: Tools for Public Sector Decision Making’, September, pp. 25-27, Washington DC.
- Tavasszy, Lóri; Ruijgrok, Kees; Davydenko, Igor (2010), Incorporating logistics in freight transportation models, 12th WCTR, July 11-15, 2010 – Lisbon, Portugal
- Tavasszy, L.A., Minderhoud, M., Perrin, J.-F., Notteboom, T. (2011): A strategic network choice model for global container flows: specification, estimation and application. *J. Transp. Geogr.* 19(6), 1163–1173
- Tavasszy L.A., Ruijgrok K., Davydenko I. (2012), Incorporating Logistics in Freight Transport Demand Models: State-of-the-Art and Research Opportunities, *Transport Reviews*, Vol. 32, No. 2, 203 – 219, March 2012
- Tavasszy, L. & de Jong, G. (2013), *Modelling Freight Transport*. Elsevier, ISBN: 978-0-12-410400-6
- Tavasszy L. (2014), Towards a sustainable multimodal freight transport system for the Randstad, VERDUS Conference Rotterdam, 16-17 June, 2014
- Tetraplan (2009), Traffic flow: Scenario, Traffic Forecast and Analysis of Traffic on the TEN-T, Taking into Consideration the External Dimension of the Union, Final Report. TENCONNECT. Copenhagen: Tetraplan A/S.
- Thissen M, Van Oort F, Diodato D and Ruijs A (2013a). European regional competitiveness and Smart Specialization; Regional place-based development perspectives in international economic networks, Cheltenham: Edward Elgar.
- Thissen M, Diodato D and Van Oort F. (2013b). Integrated regional Europe: European regional trade flows in 2000. PBL Netherlands Environmental Assessment Agency, The Hague.
- TNO RESPONSE (2004): Op zoek naar lagere kosten en hogere service, 2004, www.tno.nl/downloads/RESPONSEtm.pdf, 21-10-2014

- TNO RESPONSE (2009): Design, evaluation and optimization of Logistics Networks, <https://www.tno.nl/downloads/Presentation%20RESPONSE1.pdf> retrieved on 21-10-2014
- TRANS-TOOLS, Deliverable 3 (2006), Report on model specification and calibration results, WP 3, TNO Inro (The Netherlands, Delft)
- TRANS-TOOLS, Deliverable 6 (2008), TRANS-TOOLS Final Report, WP 6. TNO (The Netherlands, Delft)
- United Nations Framework Convention on Climate Change (UNFCCC) (2007). UNFCCC fact sheet on emission trends 1990-2005. Part 2. 1990-2005 trends for Annex I. Parties to Kyoto. http://unfccc.int/files/essential_background/background_publications_htmlpdf/climate_change_information_kit/application/pdf/fact_sheet_2007_ghg_2_0.pdf, accessed on 02-09-2014
- US DoT (2013), Overview of Logistics and Industry Re-Organization, http://ops.fhwa.dot.gov/freight/freight_analysis/econ_methods/microecon_frmwk/sec_2.htm , accessed on 20-10-2014
- Wang, Q., Holguín Veras J. (2008), Investigation of Attributes Determining Trip Chaining Behavior in Hybrid Microsimulation Urban Freight Models, Transportation Research Record: Journal of the Transportation Research Board, Volume 2066 / 2008 Freight Systems 2008
- Wang, Q., Holguín Veras J. (2009), Tour-Based Entropy Maximization Formulations of Urban Freight Demand, Paper presented at 88th Annual Transportation Research Board Conference, Washington: TRB.
- Williams I., Jin Y., Deane G. , Zhu Y. (2005), The EUNET2.0 Freight and Logistics Model - Final Report / WSP. 2005. – Research report. Project for Department for Transport, England (DFT) Contract PPAD 8/134/24
- Yang ChoongHeon, Regan A. C., Son Y.T., (2010), Another View of Freight Forecasting Modeling Trends, KSCE Journal of Civil Engineering 14(2): 237-242.
- Yin, Y., Williams, I., Shahkarami, M., 2005. Integrated regional economic and freight logistics modelling: results from a model for the Trans-Pennine Corridor, UK. Paper Presented at the European Transport Conference 2005, Strasbourg
- Zhang, M. (2013) Freight Transport Model for Integrated Network, Service, and Policy Design, PhD thesis, TU Delft

Appendix 1: NSTR goods classification

0	LIVE ANIMALS
1	CEREALS
2	POTATOES
3	OTHERS VEGETABLES, FRESH OR FROZEN, FRESH FRUIT
4	TEXTILES, TEXTILE ARTICLES AND MAN-MADE FIBRES
5	WOOD AND CORK
6	SUGAR BEETS
9	OTHER RAW ANIMAL AND VEGETABLE MATERIALS
11	SUGARS
12	BEVERAGES
13	STIMULANTS AND SPICES
14	PERISHABLE FOODSTUFFS
16	NON-PERISHABLE FOODSTUFFS AND HOPS
17	ANIMAL FEEDINGSTUFFS AND FOODSTUFF WASTE
18	OIL SEEDS AND OLEAGINOUS FRUIT AND FATS
21	COAL
22	LIGNITE AND PEAT
23	COKE
31	CRUDE PETROLEUM
32	FUEL DERIVATIVES
33	GASEOUS HYDROCARBONS, LIQUID OR COMPRESSED
34	NON-FUEL DERIVATIVES
41	IRON ORE
45	NON-FERROUS ORES AND WASTE
46	IRON AND STEEL WASTE AND BLAST-FURNACE DUST
51	PIG IRON AND CRUDE STEEL; FERRO-ALLOYS
52	SEMI-FINISHED ROLLED STEEL PRODUCTS
53	MATERIAL OF IRON OR STEEL
54	STEEL SHEETS, PLATES, HOOP AND STRIP
55	TUBES, PIPES, IRON AND STEEL CASTINGS AND FORGINGS
56	NON-FERROUS METALS
61	SAND, GRAVEL, CLAY AND SLAG
62	SALT, IRON PYRITES, SULPHUR
63	OTHER STONE, EARTHS AND MINERALS
64	CEMENT, LIME
65	PLASTERS
69	OTHER MANUFACTURED BUILDING MATERIALS
71	NATURAL FERTILIZERS
72	CHEMICAL FERTILIZERS
81	BASIC CHEMICALS
82	ALUMINIUM OXIDE AND HYDROXIDE
83	COAL CHEMICALS
84	PAPER PULP AND WASTE PAPER
89	OTHER CHEMICAL PRODUCTS
91	TRANSPORT EQUIPMENT
92	TRACTORS, AGRICULTURAL MACHINERY AND EQUIPMENT
93	OTHER MACHINERY, APPLIANCES, ENGINES, AND PARTS
94	MANUFACTURES OF MATERIAL
95	GLASS, GLASSWARE, CERAMIC PRODUCTS
96	LEATHER, TEXTILES AND CLOTHING
97	OTHER MANUFACTURED ARTICLES
99	MISCELLANEOUS ARTICLES

Appendix 2

Regional distribution volumes (ton) of scenarios 1, 2 and 3 of the logistics sprawl application case of section 6.2.1. Regions in red color belong to the Randstad region of the Netherlands.

Reg. ID	Regions	Base (1)	2a	2b	3a	3b
1	NL111 Oost-Groningen	311.764	310.351	268.644	313.295	361.375
2	NL112 Delfzijl en omgeving	77.620	77.317	66.855	77.948	90.012
3	NL113 Overig Groningen	323.240	321.894	278.672	324.747	374.569
4	NL121 Noord-Friesland	355.823	354.562	307.454	357.480	411.723
5	NL122 Zuidwest-Friesland	349.044	348.392	302.385	350.285	403.163
6	NL123 Zuidoost-Friesland	499.605	497.821	431.856	501.957	577.882
7	NL131 Noord-Drenthe	616.112	613.250	531.340	619.349	713.753
8	NL132 Zuidoost-Drenthe	160.955	160.137	138.742	161.866	186.544
9	NL133 Zuidwest-Drenthe	473.348	471.008	408.469	476.159	548.298
10	NL211 Noord-Overijssel	1.377.651	1.372.480	1.191.339	1.384.843	1.593.405
11	NL212 Zuidwest-Overijssel	564.521	562.012	487.804	567.968	653.573
12	NL213 Twente	744.230	739.883	641.949	749.483	862.627
13	NL221 Veluwe	1.279.196	1.279.419	1.111.959	1.282.197	1.474.119
14	NL222 Achterhoek	479.585	476.373	413.530	483.538	556.282
15	NL223 Arnhem/Nijmegen	747.370	742.805	645.403	753.468	866.190
16	NL224 Zuidwest-Gelderland	3.149.032	3.140.974	2.732.619	3.165.211	3.635.441
17	NL230 Flevoland	869.699	872.464	758.668	869.274	999.046
18	NL310 Utrecht	2.432.384	2.815.826	2.450.734	2.107.291	2.420.593
19	NL321 Kop van Noord-Holland	527.918	531.914	462.982	525.511	603.625
20	NL322 Alkmaar en omgeving	550.081	555.071	483.384	546.776	627.858
21	NL323 IJmond	267.141	270.185	235.329	264.904	304.199
22	NL324 Agglomeratie Haarlem	263	306	267	226	259
23	NL325 Zaanstreek	1.505.282	1.521.432	1.324.777	1.493.790	1.715.736
24	NL326 Groot-Amsterdam	2.018.263	2.349.278	2.045.800	1.738.954	1.997.154
25	NL327 Het Gooi en Vechtstreek	106.704	123.752	107.664	92.286	106.051
26	NL331 Agglomeratie Leiden en Bollenstreek	986.687	1.152.245	1.003.943	846.997	972.588
27	NL332 Agglomeratie s-Gravenhage	686.601	801.338	698.341	589.641	676.929
28	NL333 Delft en Westland	1.163.876	1.357.105	1.182.943	1.000.403	1.148.211
29	NL334 Oost-Zuid-Holland	2.031.702	2.364.611	2.060.748	1.750.219	2.008.832
30	NL335 Groot-Rijnmond	2.068.494	2.395.570	2.089.410	1.790.770	2.053.206
31	NL336 Zuidoost-Zuid-Holland	1.354.963	1.564.165	1.362.671	1.176.994	1.350.674
32	NL341 Zeeuws-Vlaanderen	131.599	129.695	113.275	133.756	152.930
33	NL342 Overig Zeeland	497.352	493.605	430.889	502.312	574.939
34	NL411 West-Noord-Brabant	3.060.859	3.049.482	2.659.714	3.080.157	3.529.171
35	NL412 Midden-Noord-Brabant	1.311.312	1.301.556	1.132.824	1.324.148	1.519.880
36	NL413 Noordoost-Noord-Brabant	1.864.263	1.847.635	1.606.882	1.884.480	2.164.158
37	NL414 Zuidoost-Noord-Brabant	1.524.723	1.506.686	1.309.911	1.544.898	1.774.269
38	NL421 Noord-Limburg	768.066	758.292	658.266	778.481	895.212
39	NL422 Midden-Limburg	597.005	589.384	511.581	604.904	695.581
40	NL423 Zuid-Limburg	328.290	324.073	281.346	332.629	382.393

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 100 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Davydenko, I.Y., *Logistics Chains in Freight Transport Modeling*, T2015/10, May 2015, TRAIL Thesis Series, the Netherlands

Schakel, W., *Development, Simulation and Evaluation of In-car Advice on Headway, Speed and Lane*, T2015/9, May 2015, TRAIL Thesis Series, the Netherlands

Dorsser, J.C.M. van, *Very Long Term Development of the Dutch Inland Waterway Transport System: Policy analysis, transport projections, shipping scenarios, and a new perspective on economic growth and future discounting*, T2015/8, May 2015, TRAIL Thesis Series, the Netherlands

Hajiahmadi, M., *Optimal and Robust Switching Control Strategies: Theory, and applications in traffic management*, T2015/7, April 2015, TRAIL Thesis Series, the Netherlands

Wang, Y., *On-line Distributed Prediction and Control for a Large-scale Traffic Network*, T2015/6, March 2015, TRAIL Thesis Series, the Netherlands

Vreeswijk, J.D., *The Dynamics of User Perception, Decision Making and Route Choice*, T2015/5, February 2015, TRAIL Thesis Series, the Netherlands

Lu, R., *The Effects of Information and Communication Technologies on Accessibility*, T2015/4, February 2015, TRAIL Thesis Series, the Netherlands

Ramos, G. de, *Dynamic Route Choice Modelling of the Effects of Travel Information using RP Data*, T2015/3, February 2015, TRAIL Thesis Series, the Netherlands

Sierzchula, W.S., *Development and Early Adoption of Electric Vehicles: Understanding the tempest*, T2015/2, January 2015, TRAIL Thesis Series, the Netherlands

Vianen, T. van, *Simulation-integrated Design of Dry Bulk Terminals*, T2015/1, January 2015, TRAIL Thesis Series, the Netherlands

Risto, M., *Cooperative In-Vehicle Advice: A study into drivers' ability and willingness to follow tactical driver advice*, T2014/10, December 2014, TRAIL Thesis Series, the Netherlands

Djukic, T., *Dynamic OD Demand Estimation and Prediction for Dynamic Traffic Management*, T2014/9, November 2014, TRAIL Thesis Series, the Netherlands

Chen, C., *Task Complexity and Time Pressure: Impacts on activity-travel choices*, T2014/8,

- November 2014, TRAIL Thesis Series, the Netherlands
- Wang, Y., *Optimal Trajectory Planning and Train Scheduling for Railway Systems*, T2014/7, November 2014, TRAIL Thesis Series, the Netherlands
- Wang, M., *Generic Model Predictive Control Framework for Advanced Driver Assistance Systems*, T2014/6, October 2014, TRAIL Thesis Series, the Netherlands
- Kecman, P., *Models for Predictive Railway Traffic Management*, T2014/5, October 2014, TRAIL Thesis Series, the Netherlands
- Davarynejad, M., *Deploying Evolutionary Metaheuristics for Global Optimization*, T2014/4, June 2014, TRAIL Thesis Series, the Netherlands
- Li, J., *Characteristics of Chinese Driver Behavior*, T2014/3, June 2014, TRAIL Thesis Series, the Netherlands
- Mouter, N., *Cost-Benefit Analysis in Practice: A study of the way Cost-Benefit Analysis is perceived by key actors in the Dutch appraisal practice for spatial-infrastructure projects*, T2014/2, June 2014, TRAIL Thesis Series, the Netherlands
- Ohazulike, A., *Road Pricing mechanism: A game theoretic and multi-level approach*, T2014/1, January 2014, TRAIL Thesis Series, the Netherlands
- Cranenburgh, S. van, *Vacation Travel Behaviour in a Very Different Future*, T2013/12, November 2013, TRAIL Thesis Series, the Netherlands
- Samsura, D.A.A., *Games and the City: Applying game-theoretical approaches to land and property development analysis*, T2013/11, November 2013, TRAIL Thesis Series, the Netherlands
- Huijts, N., *Sustainable Energy Technology Acceptance: A psychological perspective*, T2013/10, September 2013, TRAIL Thesis Series, the Netherlands
- Zhang, Mo, *A Freight Transport Model for Integrated Network, Service, and Policy Design*, T2013/9, August 2013, TRAIL Thesis Series, the Netherlands
- Wijnen, R., *Decision Support for Collaborative Airport Planning*, T2013/8, April 2013, TRAIL Thesis Series, the Netherlands
- Wageningen-Kessels, F.L.M. van, *Multi-Class Continuum Traffic Flow Models: Analysis and simulation methods*, T2013/7, March 2013, TRAIL Thesis Series, the Netherlands
- Taneja, P., *The Flexible Port*, T2013/6, March 2013, TRAIL Thesis Series, the Netherlands
- Yuan, Y., *Lagrangian Multi-Class Traffic State Estimation*, T2013/5, March 2013, TRAIL Thesis Series, the Netherlands
- Schreiter, Th., *Vehicle-Class Specific Control of Freeway Traffic*, T2013/4, March 2013, TRAIL Thesis Series, the Netherlands

About the author

Igor Davydenko (1974) holds the degree of Professional Doctorate in Engineering from the Technical University of Eindhoven (2006), having specialized in Logistics Management Systems. Before moving to the Netherlands in 2003, he had obtained Master Degree in applied mathematics (1996) and worked on various projects related to IT applications in industrial and logistics processes. Since 2006 Igor has been working at TNO in Delft as research scientist and consultant. He conducted part-time PhD research at TU Delft in the period 2009-2015. Igor's research interests lie in the choice modeling of logistics chains and transport modalities; data for strategic freight models; environmental aspects of transport, especially shipment-level emissions and energy input in complex transport chains; model applications for the choice of international transport and logistics chains.

Journal publications

Davydenko I., Ehrler V., Ree de D., Lewis A., Tavasszy L. (2014), Towards a global CO2 calculation standard for supply chains: Suggestions for methodological improvements, *Transportation Research Part D* 32 (2014) 362–372

Igor Y. Davydenko and Lóránt A. Tavasszy, Estimation of Warehouse Throughput in Freight Transport Demand Model for the Netherlands, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2379, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 9–17. DOI: 10.3141/2379-02

Auvinen H., Clausen U., Davydenko I., Diekmann D., Ehrler V., Lewis A. (2014), Calculating emissions along supply chains — Towards the global methodological harmonisation, *Research in Transportation Business & Management* (2014), <http://dx.doi.org/10.1016/j.rtbm.2014.06.008>

L. A. Tavasszy, K. Ruijgrok, Igor Davydenko, Incorporating Logistics in Freight Transport Demand Models: State-of-the-Art and Research Opportunities, *Transport Reviews: A Transnational Transdisciplinary Journal*, Volume 32, Issue 2, 2012

Peer Reviewed Congress Papers

Igor Davydenko, Mark Thissen, Lorant Tavasszy, Translation of Trade Flows Into Transport Flows in European Logistics Facility Choice Model, presented at Transportation Research Board (TRB) 93th Annual Meeting (January 2014), Washington, D.C.

Igor Y. Davydenko, Lorant A. Tavasszy, A quantitative model for exploration of logistics sprawl of the future, Presented at METRANS Urban Freight Conference, Long Beach, Ca, October 2013

Igor Y. Davydenko, Lorant A. Tavasszy, Chris J. de Blois, Modeling interregional freight flow by distribution systems, 13th WCTR, July 15-18, 2013 – Rio de Janeiro, Brasil

Verena Ehrler, Igor Davydenko, Diederik de Ree, Jaurieke Ton, Heidi Auvinen, Alan Lewis, Saskia Seidel, Andreas Lischke, Hedi Maurer, Standardized Emission Calculations along Supply Chains as a Basis for Smart Global Transport Solutions in a “Reality of Less” – Is an Approach within Reach?

AUVINEN, Heidi; CLAUSEN, Uwe; DAVYDENKO, Igor; DE REE, Diederik; DIEKMANN, Daniel; EHRLER, Verena; LEWIS, Alan; TON, Jaurieke, Calculating Emissions Along Supply Chains – Towards the Development of a Harmonised Methodology, 13th WCTR, July 15-18, 2013 – Rio de Janeiro, Brazil

I. Davydenko, M. Zhang, L. Tavasszy, Assessment of SMART-CM platform benefits as the result of shorter transport time, E-Freight Congress 2012, Delft

I. Davydenko, L.A. Tavasszy, J. J. van Meijeren, O. Ivanova, Estimation Of Freight Flow And Trip Generation According To Transport Purpose In Urban And Regional Logistics Networks, Presented at METRANS Urban Freight Conference, Long Beach, Ca, October 14, 2011

I.Y. Davydenko, L. A. Tavasszy, P.S.G.M. Smeets, Commodity Freight And Trip Generation By Logistics Distribution Centers Based On Sectorial Employment Data, Transportation Research Board (TRB) 91th Annual Meeting (2012), Washington, D.C.

Tavasszy, L., Ruijgrok, K, Davydenko, I., Incorporating Logistics in Freight Transportation Models: State of the Art and Research Opportunities (2010), 12th WCTR, July 11-15, 2010 – Lisbon, Portugal

Verweij, K., Davydenko, I., Zommer, G., Estimation of the Impact of Long and Heavy Vehicles on Future European Transport Demand and Modal Shift (2010), 14th – 17th March 2010, Melbourne, Australia.

Tavasszy, L., Davydenko, I., Ruijgrok, K, The Extended Generalized Cost Concept and its application in Freight Transport and General Equilibrium Modeling (2009), Integration of Spatial Computable General Equilibrium and Transport Modelling Bilateral Joint Seminar under agreement between NWO and JSPS, 2009 August 19-20 The University of Tokyo, SANJO-Hall

Zomer, G., Davydenko, I., Krupe, S., Need for advanced collaborative business models in intermodal transport (2008), Vervoerslogistieke werkdagen 2007

Guis, G.W., Davydenko, I, Benchmarking as promotional and Transparency Tool for Intermodal Transportation (2007), Vervoerslogistieke werkdagen 2007

Selected Reports

I.Y. Davydenko, M. Zhang, L.A. Tavasszy (2014), CBS transport data analytics, TNO Report TNO 2014 R10747

Igor Davydenko, Lori Tavasszy, Verena Ehrler, Alan Lewis (2014), COFRET D3.3 – Suggestions and recommendations towards global harmonisation of carbon footprint calculation principles and comparable reporting, COFRET Deliverable

D. de Ree, J. Ton, I. Davydenko, M. Chen, J. Kiel, H. Auvinen, K. Mäkelä, Assessment and typology of existing CO2 calculation tools and methodologies, COFRET D3.1, 2012

M. Zhang, I. Davydenko, M. Chen, T. van Rooijen, A. Roest Crolius and Y. Bai, Report on inventory and assessment of rail freight strategies and developments in China and Russia, RETRACK project deliverable D13.1, 2012

Davydenko I., Landa Maxta I., Martens R., Nesterova N., Wark T., Potential for Eurasia land bridge corridors & logistics developments along the corridors, RETRACK project deliverable D13.2, 2012

De Ceuster, G., Breemersch, T., Van Herbruggen, B., Verweij, K., Davydenko, I., Klingender, M., Jacob, B., Arki, H., Bereni, M., Effects of adapting the rules on weights and dimensions of heavy commercial vehicles as established within Directive 96/53/EC (2008), TREN/G3/318/2007

Loes Aarts, Marieke Honer, Igor Davydenko, Hans Quak, Jannette van Staalduinen, Kees Verweij, Langere en Zwaardere Vrachtauto's in Nederland: Kennis en ervaringen over de periode 1995-2010, Rijkswaterstaat - Dienst Verkeer en Scheepvaart, Maart 2010

Aarts, L., Honer, M., Davydenko, I., Quak, H., Van Staalduinen, J., Verweij, K., Kennis en ervaring Langere en Zwaardere Vrachtauto's (2010)

Davydenko, I., Rothkrantz, C., Evaluation of intermodal transportation within Reorient corridor (2007)

Lammers, B., de Kievit, M., Davydenko, I., Maatregelenmatrix CO2 reductie (2008)

Rijswijck, van, H., Davydenko, I., Intelligent Agents in Transportation Planning (2007)

Zommer, G., Davydenko, I., Israel Transport Master Plan - Plan 2 Description of EU Transport policies and regulations (2008)

Davydenko, I., Logistics network design for Access Business Group, Eindhoven : Stan Ackermans Instituut, 2006. – (Eindverslagen Stan Ackermans Instituut ; 2006/023), ISBN 90-444-0596-9, ISBN 978-90-444-0596-5