Modeling interregional freight flow by distribution systems

Igor Y. Davydenko[1], Lorant A. Tavasszy[2], Chris J. de Blois[3]

[1,2]TNO Sustainable Transport and Logistics, Postbus 49, 2600 AA Delft, the Netherlands

[1,2]Technical University of Delft, Jaffalaan 5, 2628 BX Delft, the Netherlands

[3]Statistics Netherlands, P.O. Box 4481, 6401 CZ Heerlen, the Netherlands

mail for correspondence: igor.davydenko@tno.nl

Submission for track:
B: Freight and Logistics

ABSTRACT

Distribution Centers with a warehousing function have an important influence on the flow of goods from production to consumption, generating substantial goods flow and vehicle movements. This paper extends the classical 4-step freight modeling framework with a logistics chain model, explicitly modeling freight flows via distribution centers and warehouses. Modeling logistics at the regional level establishes a link between trade flow and transport flow, allows determination of the warehouse and distribution centers locations and corresponding freight flows, permits more detailed and accurate policy decision support systems. This paper describes a two-stage logistics model that estimates interregional goods flow via logistics chains. The first stage estimates interregional trade flows by the means of a gravity model application starting from regional production and consumption volumes. The second stage, the logistics chain model, creates logistics chains, which split the production-consumption flow between direct shipments and shipments via warehousing facilities. We use an aggregate multinomial nested logit discrete choice model to determine flow volumes for each of the possible logistics chains. We achieve consistency between the gravity and logistics chain models by a joint estimation of unknown parameters. The proposed two stage model estimates interregional OD freight flows that match the observed real world volumes. We use a new transport flow survey dataset produced by Statistics Netherlands with information on loading and unloading location types. This dataset allows the required model calibration with respect to different types of flow. We present estimated and observed interregional freight flow volumes, estimated parameter values and their interpretation.

Keywords: freight transport modeling, logistics chains, freight modeling framework, model calibration

1. Introduction

Logistics operations play an important role in the functioning of freight systems. In production – consumption relations goods are often transported via distribution centers and warehouses, such as warehouses of wholesalers, distribution centers of supermarket chains, regional distribution systems of manufacturers. Warehouses provide storage for the goods closer to the consumers, thus allowing faster response to the varying demand, decrease inbound transport costs, allowing bigger transport batches to the warehouse facilities, and hence smaller transport costs.

Warehouses and distribution centers generate freight movements. In the Netherlands, distribution centers generate at least 14% of all loaded in Heavy Goods Vehicle (HGV) ton volumes and at least 12% of all HGV trips (based on the extended CBS road transport survey, Davydenko (2011)). Traditional 4-step freight models do not capture explicitly the logistics aspects of freight transport; it is often assumed 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.

The main function of the warehouse and distribution facility location model is related to the estimation of spatial freight flows and generalization of transport costs (Tavasszy 2009). Warehouse (or inventory) model reproduces spatial patterns of freight transport flows that are related to intermediate inventory locations. If external factors such as transport costs, warehouse labor costs, costs of facilities itself change, then the model is able to capture changes in freight flows via these facilities and help predicting future transport flows. Second, the logistics model allows a more realistic estimation of the costs involved in producer-consumer interaction. These costs are important for the estimation of trade flows between regions and countries, allowing a more accurate estimation of distribution models.

Recently there has been a number of review papers published on the state of the art in freight transport modeling. De Jong et al (2012) focus on freight models developed in Europe, Chow et al. (2010) provide an overview of mostly US modeling efforts. Tavasszy et al. (2012) provide a review on particular efforts to include elements of logistics decision making in national and international freight transport models. These review papers pay special attention to the state of the art in logistics modeling in the aggregate freight modeling framework. In this paper, we briefly consider the most relevant for the purpose of this research contributions to the logistics modeling.

Chow et al (2010) point out that capturing of logistics decisions at the aggregate regional level is very data intensive: the models need trade flow, various costs factors as the input; on the other hand a very detailed transport data is necessary for the calibration and validation purposes. The authors state that despite the fact that accurate modeling of logistics structures is very important from the point of view of freight flow and truck trip generation, logistics models are still in the early phases of development. It is probable that logistics models are more developed in Europe and for European application purposes because to date regional logistics models have not been applied by any U.S. state agency due to the unavailability of basic data (Chow et al 2010). Tavasszy et al. (2012) develop various directions into which logistics models can develop. Our focus here is on the particular problem of distribution centers, or as also commonly noted, warehousing.

The SMILE model (Tavasszy 1998, Bovenkerk 2005) has first modeled logistics explicitly as a distinct step in translation of the trade flows into transport flows. This has been done using the concept of logistics chains, representing a discrete choice problem between various ways of shipping goods from the production to consumption points. A multinomial logit was used to determine probability of each choice and consequently transport flows. Due to a lack of data on warehousing activities, the calibration of this part of the model could only be carried out indirectly on transport data. This research builds on the SMILE model lineage, extending it with empirically validated logistics modeling. The SMILE approach to modeling of logistics has also been used in “industrial” models such as TRANS-TOOLS model (Chen (2011) and TRANS-TOOLS Deliverable 6, 2008). SLAM / SCENES models are also based on SMILE philosophy (SCENES deliverable 4, 2000).

In the UK, the EUNET model has been developed. It integrates regional economic and freight logistic model (Jin et al 2005). EUNET2.0 divides the economic trade from the initial producer to the ultimate consumer into a number of logistic stages, as appropriate for each category of commodity. The model estimates a set of OD matrices that are segmented by commodity type and type of distribution stage, including consolidation centres, national and regional distribution centres, and major ports. The model simulates freight demand coming from logistical operations as well as from the wider regional and national economy.

Kim et al (2010) proposes to use physical distribution channel choice models imitating shippers’ choice of distribution chain. The approach is based on the minimization of the overall logistics costs including inventory cost, transportation cost and other cost components, which are not incorporated well in the traditional four-step models. The authors found that a logit model is suitable for the modelling of distribution channels.

Maurer (2008) proposed a logistics model in the context of emissions estimation from freight transport in Great Britain. Maurer used estimated PC flows as an input for supply chain optimization software, which determined location of warehouses. The supply chain optimization resulted in spatial stock allocations and transport needs at the regional level. Given known production-consumption flows for one commodity “Food, drinks and agricultural products”, it has been shown that supply chain cost minimization models can determine allocation of stocks in supply chains. The model considered British supply chains for the drinks commodity as if it were run and optimized by one entity. As the logistics model used is a commercial package, the approach is difficult to validate.

Friedrich (2010a, 2010b) performed a study of food retail sector in Germany and built a model that estimates freight demand of three specific retail chains. Besides company data the author used chamber of commerce data, trade volumes and industry specific data to generate a population of shops or demand points. His model searches a stable-state situation, where locations of DCs serving these shops are determined. The model is based on actor interaction modeling in meso-structures (Liedtke 2006 and Liedtke 2009). The proposed simulation system includes detailed logistic optimization of food retailing companies as well as simplified optimizations of adjacent logistic systems. The resulting simulation model, called SYNTRADE, is able to reproduce logistic structures in food retailing in Germany and has been validated by comparison to the real world transport data. Its application to broader freight demand models would be cumbersome, however, as the data needs cannot be fulfilled.

Combes (2010) has focused on the choice of shipment sizes. He proposed a model on the choice of shipment size and transport mode, validated on the French ECHO survey of shippers (Guilbault 2008). Combes (2010) has shown that application of EOQ formulae leads to empirically confirmed estimations of transport batch sizes, stressing it as one of the main drivers for the organization of freight transport systems.

Melendez and Horowitz (2011) developed a transshipment model, which forecasts distribution tour structure. Based on the Ontario Commercial Vehicle Survey, the authors developed a model that determines the routing of goods on their way from production to consumption. The model computes what share of goods would be shipped directly from manufacturing to consumption and what share of goods would be shipped via a logistics facility or a number thereof. The authors have identified eight discrete tour structures, which allowed application of a multinomial logit technique. The selection of probabilities used in the micro simulation process were established using logit models estimated with choice data collected in local surveys.

Individual companies organize their logistics chains by balancing two essential requirements. First requirement is to satisfy customer demand in accordance to customer needs and expectations. These expectations may include certain service levels agreements, for instance, availability of goods on stock at stock points, such that customer demand can be immediately satisfied from those stock points. It can also be that the customers expect fast on-time deliveries, as it is often the case in industries pursuing lean manufacturing practices, such as the automotive industry, Dicken (2003). The second requirement is the cost minimization, such that customer demands can be satisfied at minimum costs. In practice it translates into consolidation of shipments, use of warehouses where the goods are kept on stock or for consolidation and deconsolidation of shipments to optimize unit transport costs. At the company level, cost minimization is subjected to satisfaction of the required service level. These problems are dealt with in the realm of operations research: Baumgartner et al (2012) presents a recent example of a supply chain design with service level requirements.

This paper presents a model that finds the results of logistics activities of Dutch companies at the macro level, i.e. at the level of interregional transport and goods flows. Patterns of the design of supply chains of individual companies can be found at the macro level, where random utility models can be used to find the logistics trade-offs made at the micro level and observed in the macro-level interregional goods flow. We model essentially two choices: the first choice is whether the flow is direct from production to consumption or goes via a distribution center. The second choice is the choice of the region where warehouse is located in case of indirect flow. A nested logit random utility model is sued to model these choices.

This paper makes a step further in respect to the modeling results presented in Davydenko (2013), where it has been shown that a nested logit random utility model can be used for accurate estimation of freight flows generated by warehouses and distribution facilities. In this paper we show that essentially the same conceptual model can also make sufficiently accurate estimations of interregional flows. The main difference between these two estimations is that warehouse throughput is estimated for n values (Davydenko 2013), while interregional transport flows are estimated for n^2 flow OD pairs (n is equal to the number of regions under consideration), the number of estimated model variables is substantially smaller than the number of estimated transport flow OD pairs.

The paper is structured as follows. In section 2 we motivate the presented modeling efforts and describe the problems that are solved by the proposed model. In section 3 on model definition we provide a conceptual overview of the model, describe its two main sub-models: the gravity model (estimation of trade flows) and logistics chain model (estimation of transport flows). We further describe interaction between these two models and provide information over the calibration process. In section 4 on data we describe the main data source used by the model, the annual road transport survey conducted by Statistics Netherlands. Section 5 provides results of the model application case to the Dutch regional freight transport system: we provide indications on observed and estimated regional goods flow volumes at OD level; we also discuss model variables such as estimated vehicles loads, transport costs, warehouse-related costs. Section 6 provides the conclusions and recommendations, policy-related issues and suggestions for further research.

2. The modeling problem

The model presented in this paper extends the traditional 4-step freight modeling approach with a 5th intermediate step, adding a logistics model to the modeling framework. Traditional freight models do not capture explicitly freight flow via warehouses and distribution facilities, and use some factors to account for mismatch between trade flow and transport flow. Logistics can be modeled explicitly by the means of logistics chains, see figure 1. In the 5-step framework, regional production and consumption are estimated on the basis of economic statistics, such as input-output tables (step 1). In step 2, regional production and consumption are converted into trade flows, in the form of production-consumption tables (PC tables, see step 2), often by the means of gravity models. Transport costs are used as resistance factor in these models, possibly including other accessibility components such as trade barriers. The logistics services (step 3) is the main emphasis of this contribution in the form of a possibility to link trading regions directly or via a warehouse. Total costs can be computed for each chain, allowing application of discrete choice methods to determine the flow on each of the chains. Steps 4 and 5 allow for modal split and assignment of the transport flow to the infrastructure networks. Clearly there are many interesting interactions between modal split, routing and warehousing. Our focus in this application, however, is not on detailed routing choices (between distribution centers and consumers, or between producers and distribution centers), but on the routing choice at a higher network level (direct or indirect shipping). Also, multimodality issues in connection to warehousing lie outside the boundaries of the model.

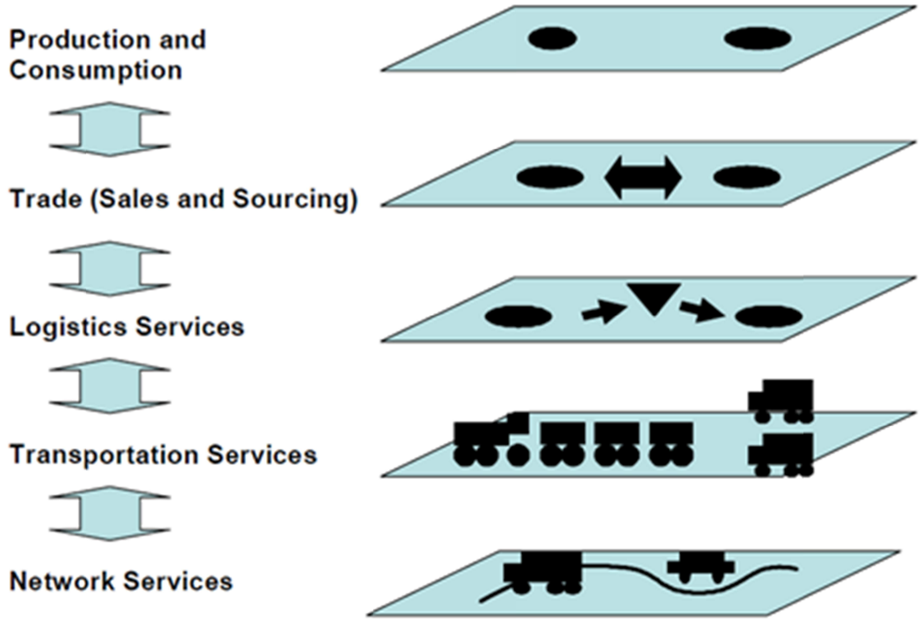


FIGURE 1. Freight modeling framework with explicit emphasis on logistics, adopted from Tavasszy (2006).

There are data-related challenges in the way of implementing of the 5-step modeling framework. In respect to the logistics services step, it requires PC flow data as the input, cost data for the determination of disutility of the logistics choices, and transport flow data for the model calibration and validation. In the paper's application case for the Netherlands, transport flow data is available in the form of transport survey (see section 4 on data). The logistics costs are not directly available, but are partly estimated in the model. The trade flow data is not observed in the Netherlands and has to be estimated in production-consumption and trade steps.

The trade flow data can be estimated on the basis of the observed by the statistics bureaus economy Input-Output tables. To estimate PC flow, four steps need to be taken. First, Input-Output data has to be regionalized (step 1 of the modeling framework). It can be done using regional employment statistics, for instance. Second, the economy sectors have to be converted into commodities (good types). There is generally no 1-to-1 conversion method, however, basic conversion tables are available. Third, monetary production units (Euros) have to be converted into freight-related units (ton). Fourth, production regions must be matched with the consumption regions (step 2 of the modeling framework). It is done by the means of application of constrained gravity model.

We estimate regional production and regional consumption volumes on the basis of road transport survey (see section 4 on data). Production volumes are the volumes coming out from the production location type in the survey. Similarly, consumption volumes are the volumes brought to consumption and production (for further rework) location types. In this way, the production and consumption regional volumes are consistent with road transport data. A two-step model, consisting of a gravity model for trade flow estimation and logistics chain model for warehouse activity estimation is applied. The model is described in the following section.

3. Model Definition

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 flow in a matrix form, namely production-consumption flow or PC flow, essentially representing the physical trade flow between regions. Second, a logistics chain model is used to estimate how PC flow is physically moved between production and consumption locations. The logistics chain model splits PC flow between direct shipments and shipments via distribution centers or warehouses, estimating three types of interregional flow: production-to-distribution (P2D flow), production-to-consumption (P2C flow) and distribution-to-consumption (D2C flow). The model is calibrated on transport survey data, minimizing root-mean-square deviation (RMSD) between observed and estimated OD flows. Figure 2 shows a schematic representation of the model.

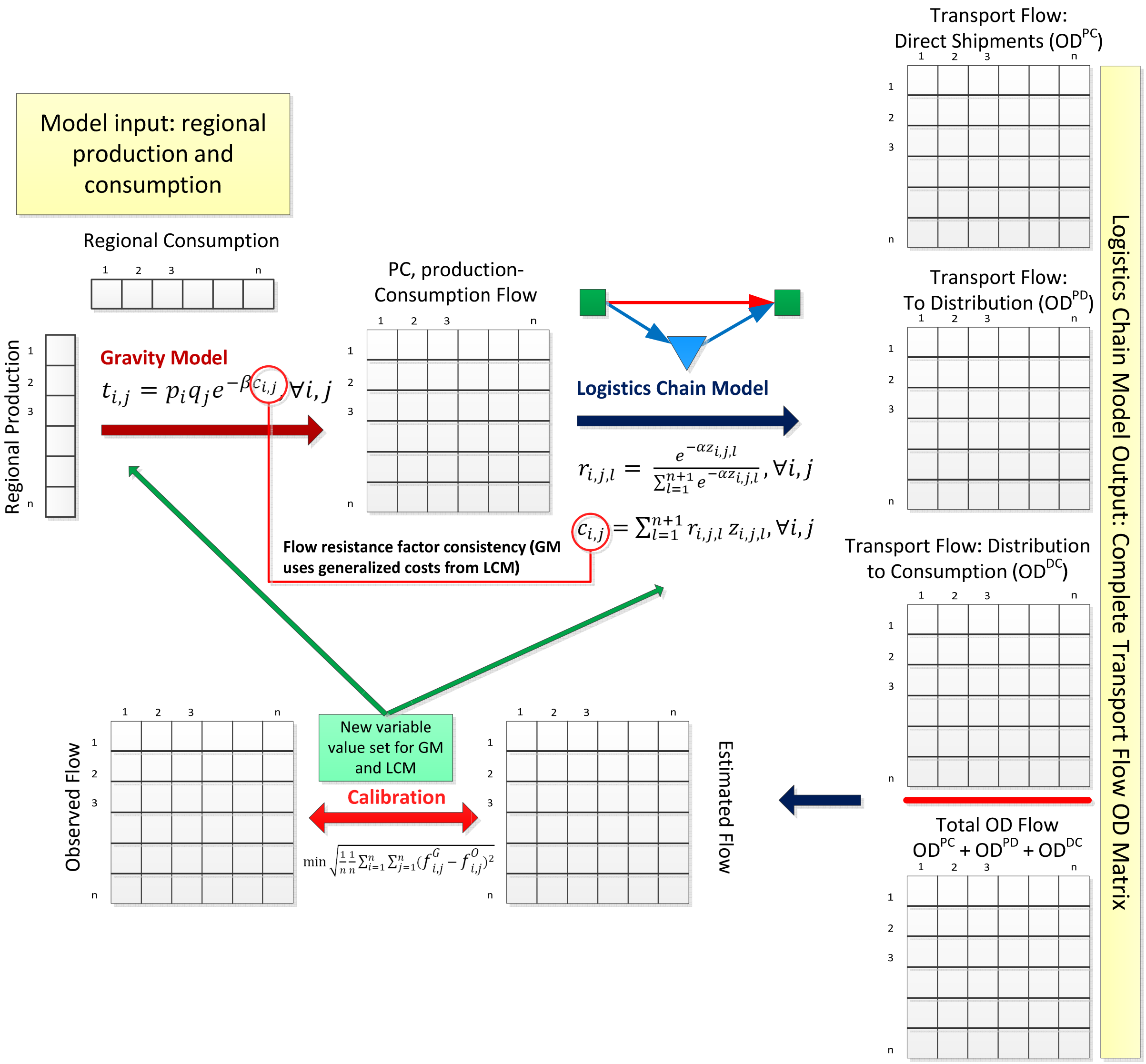


FIGURE 2. Conceptual representation of gravity and logistics chain combined model

3.1. Gravity Model Definition

Let P_i denote regional production volumes loaded into road Heavy Goods Vehicles (HGV) and expressed in ton units, $i = 1,..,n$. In the Dutch model application case, the Netherlands is divided into 40 NUTS3 so-called COROP regions, therefore, n = 40. Let C_i denote regional consumption volumes offloaded from HGV vehicles, $i = 1,..,n$. The Gravity Model (GM) used to obtain the PC flow is defined in the following form.

$$t_{i,j} = p_i q_j e^{-\beta c_{i,j}}, \forall i, j \quad (1)$$

where $t_{i,j}$ represent estimated individual cells of the PC matrix, p_i and q_j are estimated parameters of the gravity model representing regional production and attraction

β is the sensitivity parameter of the gravity model

$c_{i,j}$ is the cost friction factor in the form of generalized logistics cost per ton shipped between production region i and consumption region j. $c_{i,j}$ term in gravity model is a constant, it is computed in the logistics chain model.

Our implementation of the GM is solved iteratively by searching for values p_i and q_j until the following two equations are satisfied:

$$\sum_{j=1}^{n} t_{i,j} - P_i < \varepsilon, \forall i \quad (2)$$

$$\sum_{i=1}^{n} t_{i,j} - C_j < \varepsilon, \forall j \quad (3)$$

Where ε is an arbitrarily small value. The ε value defines how accurate the gravity model solution is. A small values of ε lead to a very precise solution of the model, however, at the expense of extra iterations. Essentially, constraints (2) and (3) guaranty that the regional production outflow and regional consumption inflow in the estimated PC table are equal to the regional production P_i and regional consumption C_j respectively within the error margin ε.

3.2. Logistics Chain Model Definition

The logistics chain model determines how PC 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,..,40. Consult Figure 3 for graphical representation of the choices.

We model logistics chains as two choices. The first choice is whether the shipment is direct or the shipment is carried out via distribution centers. Therefore, this top-level choice has two alternatives, direct shipment from region i to region j, or indirect shipment via a warehouse in the region k, thus following the chain i → k, and k → j.

Let $r_{i,j,l}$ denote the probability that products of region i are shipped to region j via chain l, l = 1,..,n+1. Index i takes the values in the range 1,.., n+1 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. We assume that $l = 1$ value represents direct flow and $l = k$ represents flow via warehouse located in k -1 region. Flow conservation constraint is introduced (4) in order to guaranty that the flow from i to j is carried out (4).

$$\sum_{l=1}^{n+1} r_{i,j,l} = 1, \forall i, j \quad (4)$$

The probability of a direct shipment between *i* and j, $r_{i,j,1}$ is computed according to (5)

$$r_{i,j,1} = \frac{e^{-\alpha\prime\, LgSumDirect_{i,j}}}{e^{-\alpha\prime\, LgSumDirect_{i,j}} + e^{-\alpha\prime\, LgSumIndirect_{i,j}}}, \forall i,j \quad (5)$$

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. In case of direct shipments (6), it is one value; in case of indirect shipments, it is a logsum of 40 alternatives (7).

$$LgSumDirect_{i,j} = \ln e^{-\alpha z_{i,j,1}}\ , \forall i,j \quad (6)$$

$$LgSumIndirect_{i,j} = \ln \sum_{k=2}^{n+1} e^{-\alpha z_{i,j,k}}\ , \forall i,j \quad (7)$$

Therefore, equation (5) computes the probability of direct shipments $r_{i,j,1}$ and equation (8) computes probability for indirect shipments ($l \neq 1$).

$$r_{i,j,l} = \frac{e^{-\alpha z_{i,j,l}}}{\sum_{l=2}^{n+1} e^{-\alpha z_{i,j,l}}}\left(1 - r_{i,j,l}\right), \forall i,j; l \neq 1 \quad (8)$$

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 sensitivity parameter; α is the nested logit cost sensitivity parameter. Smaller values of this parameter make the system less sensitive to the cost signals, higher values of the parameter make the system to react strongly to the cost of a particular chain *l*.

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 (9) and (10)

$$z_{i,j,l} = \frac{d_{i,j} c^{vkm}}{L^{PC}} \qquad \text{if chain } l \text{ is direct } (l=1) \ (9)$$

$$z_{i,j,l} = \frac{d_{i,k} c^{vkm}}{L^{PD}} + \frac{d_{k,j} c^{vkm}}{L^{DC}} + c^{w} + A_k$$

if chain *l* includes warehouse in region *k* ($l \neq 1$) (10)

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

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

$L^{PC}\ L^{PD} L^{DC}$ are the 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. It is a model calibration parameter.

The gravity model described in section 3.1 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 (11) 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 (11) make the gravity and logistics chain models consistent in the terms of costs used.

$$c_{i,j} = \sum_{l=1}^{n+1} r_{i,j,l}\, z_{i,j,l}, \forall i,j \quad (11)$$

The logistics chain model allows estimation of transport Origin-Destination OD table from the trade flow PC 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).

$$f_{i,j}^{G,PC} = t_{i,j}\; r_{i,j,l \in direct\ shipment}, \forall i,j \quad (12)$$

$$f_{i,j}^{G,PD} = \sum_{k=1}^{n} (t_{i,k}\;\; r_{i,k,l \in DC\ in\ j}), \forall i,j \quad (13)$$

$$f_{i,j}^{G,DC} = \sum_{k=1}^{n} (t_{k,j}\;\; r_{i,k,l \in DC\ in\ i}), \forall i,j \quad (14)$$

$$f_{i,j}^{G} = f_{i,j}^{G,PC} + f_{i,j}^{G,PD} + f_{i,j}^{G,DC}, \forall i,j \quad (15)$$

Figure 3 provides a graphical illustration on how PC flow is split into OD flow and the matrix summation. Cell (2, 3) in PC 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. Note that figure 3 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.

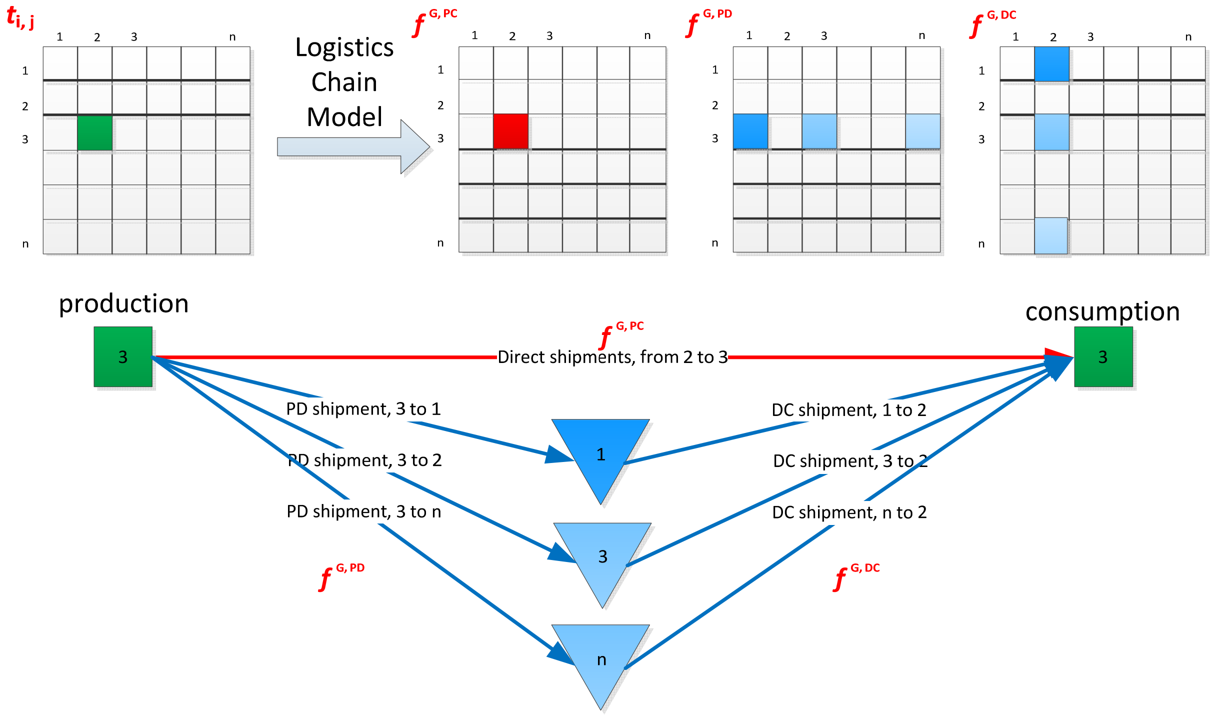


FIGURE 3. Graphical representation of PC flow conversion into transport OD flow

3.3. Model Calibration

A model calibration has been performed in respect to minimization of the mean square error between observed interregional goods flow and estimated by the model good flow in the form (16).

$$\min \sqrt{\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(f_{i,j}^{G} - f_{i,j}^{O})^2} \quad (16)$$

The following variables has been used as the model calibration parameters: c^w (cost per ton of warehouse or distribution center ton throughput); vehicle loads for the three transport stages, L^{PC} $L^{PD} L^{DC}$; regional attractiveness A_k for the distribution or warehousing activities; logit sensitivity parameters α and α' in the chain model; gravity model sensitivity parameter β. The A_k can be interpreted as the centrality factor: if distribution volumes in a region cannot be explained only by the costs of logistics chains, but there are some other factors at play such as historical (legacy) industries, availability of labor and infrastructure, and other factors.

It should be noted that vehicle loads L^{PC} $L^{PD} L^{DC}$ for the three transport stages can be seen as the proxy for transport costs on these transport stages. The model uses the constant vehicle-kilometer cost factor c^{vkm}; ton-kilometer costs are follow directly from the vehicle loads as L^* / c^{vkm}. Therefore, the model estimates vehicle loads and ton-kilometer costs at the same time. It should be further noticed that flows $f_{i,j}^{G}$ and $f_{i,j}^{O}$ in (16) are composite flows consisting of PC, PD and DC sub-flows (formulas (12)-(15)). The model can be calibrated in respect to these constituting flows as well. As we discuss in the results section, an adjustment of the composition of the total flow $f_{i,j}$ can lead to a generally better model results.

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 variable values stabilize after 4-6 iterations. The comparison of the observed and estimated regional warehouse throughput volumes is given in section 5 on results.

4. Data

The Dutch statistics bureau, Statistics Netherlands (CBS), performs annual transport surveys of HGV operators. Commercial entities operating HGVs in the Netherlands are obliged to describe vehicle operations for one week in a year. The week for which the data is collected is chosen randomly; some operators are not required to complete annual surveys, but approached less frequently in order to lighten the administrative burden. Depending on sampling frequency and other factors, CBS scales up the transported ton volumes and the number of vehicle trips to the annual level. Table 1 describes the structure of CBS transport survey dataset used for the modeling purposes described in this paper.

TABLE 1. Transport survey dataset structure

Variable (field)	***Description***
Year	Year identifier of the annual flow
Loading Region	NL NUTS3 region where the goods are loaded. The Netherlands is divided into 40 NUTS3 regions.
Loading Location Type	Specifies loading location type. Variable can take the following values: Production Consumption Sea Terminal Rail Terminal Airport Inland Waterways Terminal Entrepot Distribution Facility (Warehouse) Other / Unknown
Unloading Region	NL NUTS3 region where the goods are unloaded
Unloading Location type	Specifies unloading location type. See Loading Location Type description for variable values
Commodity Transported	Specifies type of goods transported according to the NSTR-2 commodity classification. NSTR-2 classification distinguishes 52 types of goods grouped into 10 classes
Weight Transported	Annual ton volume transported between the pair of loading and offloading regions, loaded and unloaded at the specified loading and offloading location types and of the specified commodity type.

The dataset represents transport observations for the period 2007-2009. The dataset contains approximately 250 000 records, which represents the flow for 6 739 200 possible relations (9 loading location types x 9 unloading location types x 52 commodity types x 40 loading regions x 40 unloading regions). For the basic model runs, we aggregated all commodities together, making no distinction between good types. This measure effectively reduces the number of relations to 129 600.

The uniqueness of the dataset used for the modeling purpose is that it provides data on loading and unloading location types. A loading-unloading location type pair specifies the purpose of the flow. For instance, goods loaded at production location type and unloaded at distribution location type represent PD (production-distribution) flow. Similarly the regional production P_i and regional consumption C_j vectors have been constructed. P_i values are obtained as the sum of outgoing ton volumes from production location type in region *i*; C_j is the sum of incoming ton volumes into region *j* to the consumption and production (for further rework) location types.

The distribution location type is associated with warehousing and distribution activities. Previous research has shown a strong correlation between freight volumes passing through distribution locations, as defined in the dataset, and employment in wholesale and warehousing economy sectors (Davydenko et al 2012). Employment data can also be used to estimate throughput of the logistics distribution centers (Davydenko et al 2011). The link with employment in wholesale and warehousing sectors suggests that the distribution location type is related to freight volumes coming in and out of facilities where stocks are actually kept.

5. Results

Model calibration was performed on interregional goods flows, according to the calibration procedure described in section 3.3. The model is calibrated by minimizing the difference between observed flow and generated flow, according to formulae (16). We consider flows between 40 Dutch regions, which implies the flow matrix size of 40x40, each cell representing a flow from region *i* to region *j*. For calibration we do not consider intraregional flows, setting $f_{i,i} = 0, \forall i$. There is no good estimation of the transport costs within the region; intraregional flows are generally large: these factors would substantially influence the modeling outcome if these flows are considered within the model.

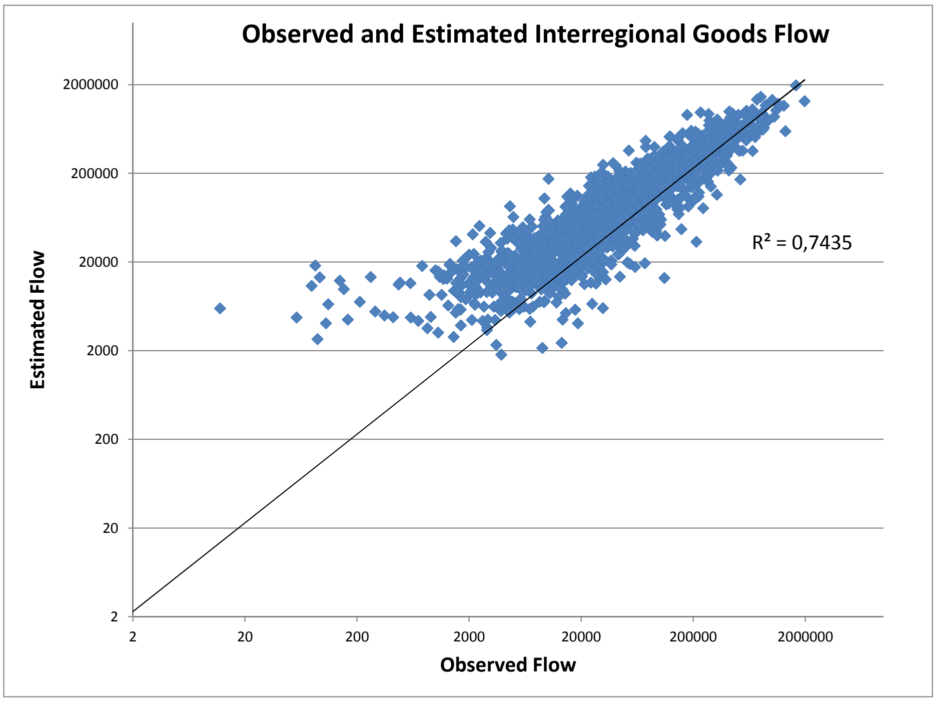


FIGURE 4. Observed and estimated total interregional goods flow, annual ton per relation

As it can be observed in Figure 4, the model reproduces generally well the interregional OD flows. These are composed as the sum of production-consumption, production-distribution and distribution-consumption OD flows. These 3 types of constituting flows can be considered separately. Figures 5-7 show model accuracy in respect to these flows.

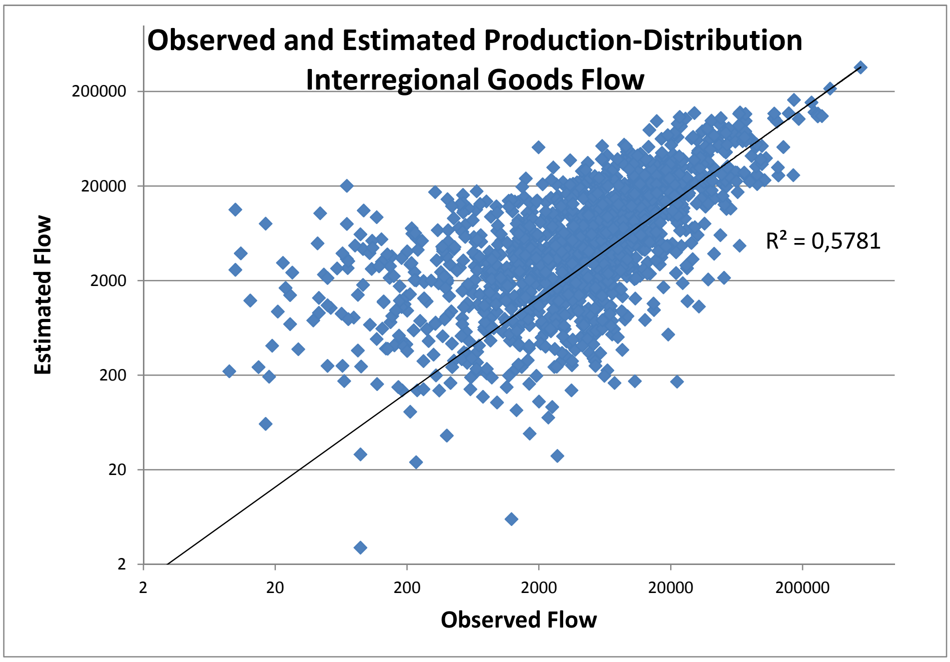


FIGURE 5. Observed and estimated production-distribution interregional goods flow, annual ton per relation

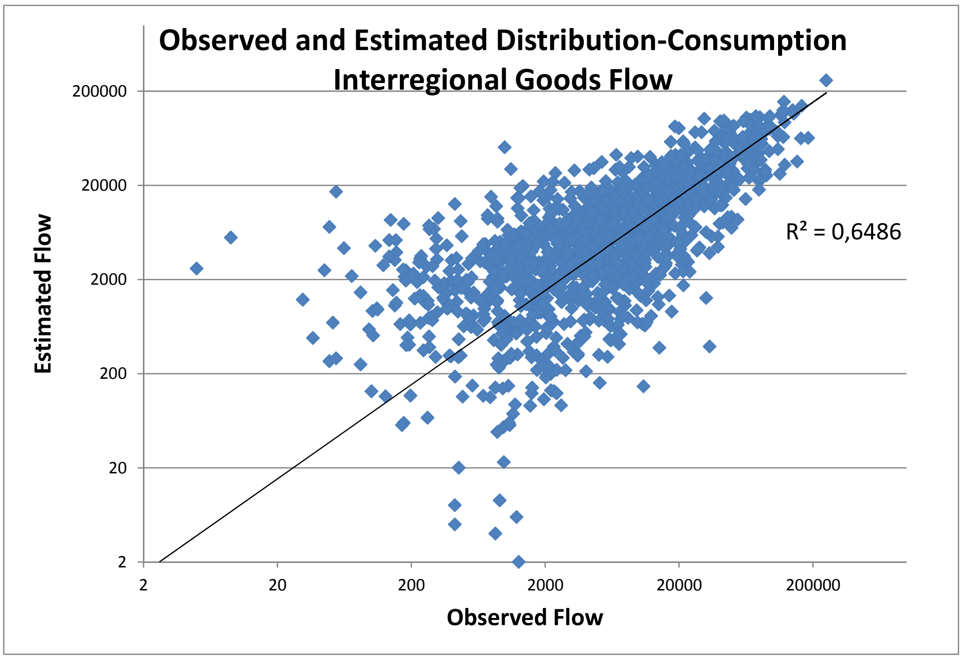


FIGURE 6. Observed and estimated distribution-consumption interregional goods flow, annual ton per relation

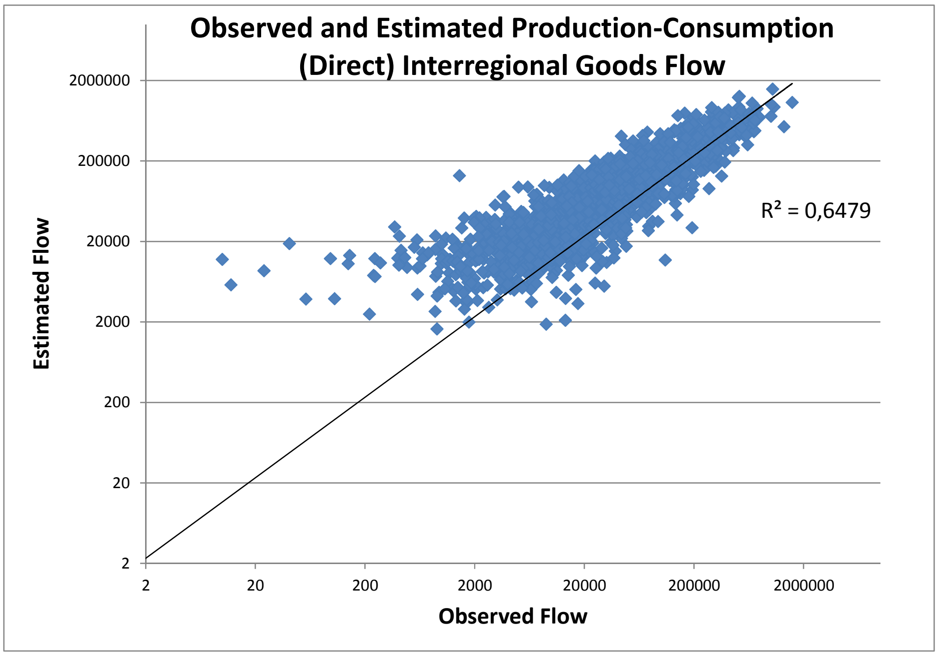
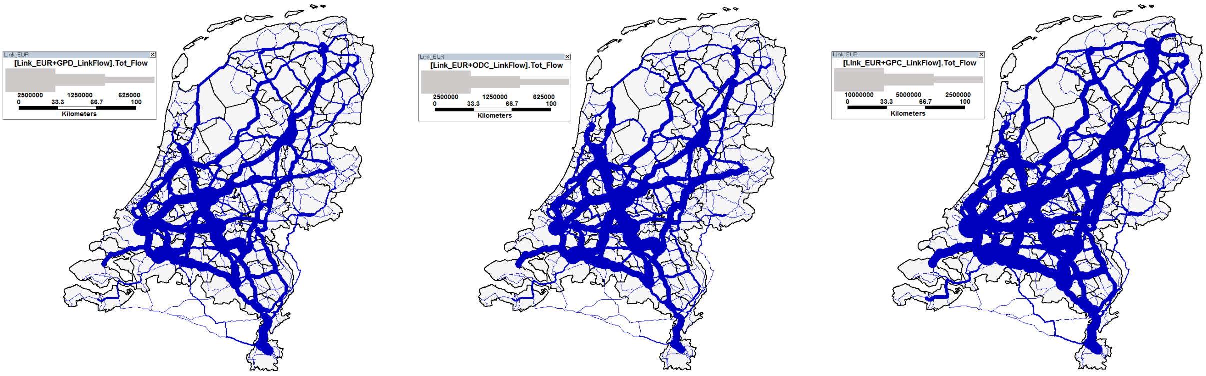


FIGURE 7. Observed and estimated production-consumption (direct) interregional goods flow, annual ton per relation

The estimated flows can be also visualized in the form of network flows. Figure 8 shows 3 types of flow, namely production-distribution, distribution-consumption and production-consumption as the flow intensity on the Dutch road network. The thickness of the line corresponds to the flow intensity. The network flows are estimated by Transcad software as the shortest path between origin region centroid and destination region centroid, without accounting for access and regress flows.



Production-Distribution Distribution-Consumption Production-Consumption

FIGURE 8. Estimated network flows

We have also observed that model calibration on the partial flows provides generally better estimations of OD flows. The calibration function (16) has been substituted with the function that explicitly takes the constituting sub-flows into account, formulae (17).

$$\min \quad a\sqrt{\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(f_{i,j}^{G,PD}-f_{i,j}^{O,PD}\right)^{2}} + b\sqrt{\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(f_{i,j}^{G,PC}-f_{i,j}^{O,PC})^{2}} + c\sqrt{\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(f_{i,j}^{G,DC}-f_{i,j}^{O,DC})^{2}} \quad (17)$$

We attribute better calibration results based on the sum of MSE to the fact that only some 16% of the ton flow is shipped indirectly via the distribution centers, thus influence of direct flow estimation prevails in the overall model accuracy. When constituting sub-flows are given weights *a*, *b* and *c* (17), it is possible to find a balance between optimization of these different flows. For the shown calibration run, we used *a = c = 1* and *b = 0.1*.

The regional warehouse throughput w_i can also be estimated by the model. It can be obtained by the summation of incoming or outgoing flows from distribution centers, as stipulated by the formulae (18). Figure 9 shows estimated and observed regional warehouse throughput based on outgoing transport flow.

$$w_i = \sum_{i=1}^{n}\left(f_{i,j}^{DC}\right), \forall i \quad (18)$$

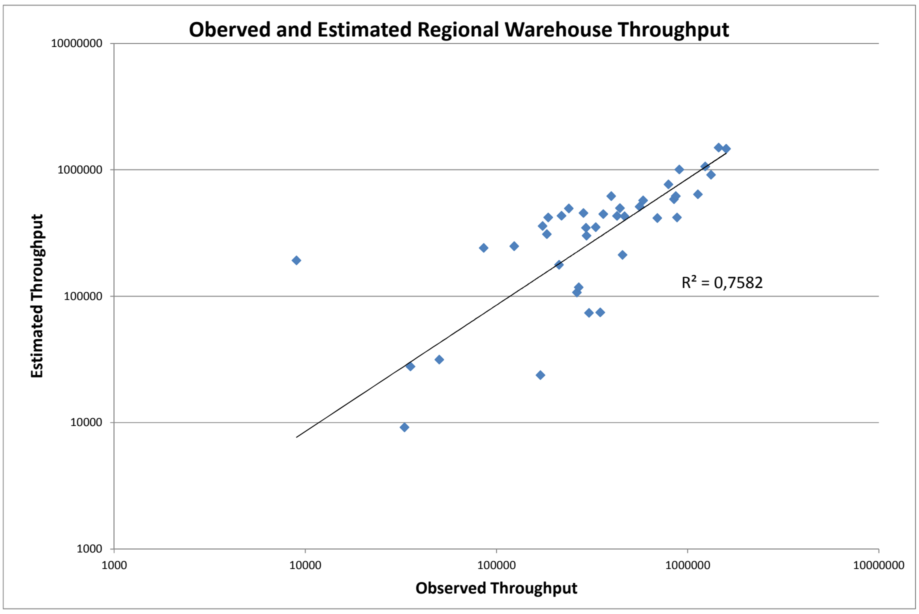


FIGURE 9. Observed and estimated regional warehouse throughput, annual ton per region

The model estimates transport costs and average vehicle loads for the three types of transport movements. It also estimates regional attractiveness for the distribution. Table 2 shows estimated transport costs and vehicle loads; Table 3 presents estimated regional attractiveness A_k for distribution. Note that the regional distribution attractiveness, specified in formulae (10), presents an extra cost of distribution activities in region *k*. Negative A_k values decrease the warehousing costs, increasing attractiveness of the region.

TABLE 2. Estimated costs and average vehicle loads

Parameter	**Value**
Production – Distribution load, ton	7,533
Distribution – Consumption load, ton	7,250
Production – Consumption load, ton	4,417
Production – Distribution cost, Eurocent per ton-km	17,26
Distribution – Consumption cost, Eurocent per ton-km	17,93
Production – Consumption cost, Eurocent per ton-km	29,43
Warehouse costs, Euro per ton throughput	4,50

TABLE 3. Estimated regional warehousing attractiveness, Euro per ton throughput

Region	**Attractiveness**
NL111, Oost-Groningen	-21
NL112, Delfzijl en omgeving	-9
NL113, Overig Groningen	-23
NL121, Noord-Friesland	-17
NL122, Zuidwest-Friesland	-1
NL123, Zuidoost-Friesland	-15
NL131, Noord-Drenthe	-21
NL132, Zuidoost-Drenthe	-9
NL133, Zuidwest-Drenthe	-9
NL211, Noord-Overijssel	-21
NL212, Zuidwest-Overijssel	-1
NL213, Twente	-21
NL221, Veluwe	-5
NL222, Achterhoek	1
NL223, Arnhem/Nijmegen	1
NL224, Zuidwest-Gelderland	-23
NL230, Flevoland	3
NL310, Utrecht	-11
NL321, Kop van Noord-Holland	23
NL322, Alkmaar en omgeving	29
NL323, IJmond	53
NL324, Agglomeratie Haarlem	53
NL325, Zaanstreek	-1
NL326, Groot-Amsterdam	-7
NL327, Het Gooi en Vechtstreek	53
NL331, Agglomeratie Leiden en Bollenstreek	13
NL332, Agglomeratie s-Gravenhage	27
NL333, Delft en Westland	3
NL334, Oost-Zuid-Holland	-5
NL335, Groot-Rijnmond	-3
NL336, Zuidoost-Zuid-Holland	5
NL341, Zeeuws-Vlaanderen	53
NL342, Overig Zeeland	11
NL411, West-Noord-Brabant	-25
NL412, Midden-Noord-Brabant	-5
NL413, Noordoost-Noord-Brabant	-17
NL414, Zuidoost-Noord-Brabant	-21
NL421, Noord-Limburg	-15
NL422, Midden-Limburg	-9
NL423, Zuid-Limburg	-25

The calibrated model can be used for the policy advice studies. The model variables can be adjusted to accommodate possible future scenarios. For instance, transport costs might be adjusted to take into account possible introduction of kilometer levies, changes in fuel costs, introduction of Long and Heavy Vehicles (LHV). The regional warehousing attractiveness variable can be used to estimate the impact of new infrastructure, such as new highways and better connection options, as well as changes in labor composition and the effect of the graying of workforce. The changes in model parameters will lead to a new spatial allocation of distribution and warehousing activities, correspondingly changing spatial distribution of transport flows.

6. Conclusions

This paper 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 described in the paper estimates interregional goods flow carried out by road transport within the Netherlands. We have used road transport flow data from annual surveys conducted by Statistics Netherlands for the model calibration.

We overcome the trade flow data availability problem by estimating 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 we found values for these parameters such that MSE between estimated and observed interregional goods flows is minimized.

We present and discuss in detail the model estimation results. For the total OD flows, R^2 between estimated and observed flows is 0,74. For the constituting PC, PD and DC flows R^2 values are between 0,58 and 0,65. The model also performs well in respect to estimation of regional warehouse throughput.

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. Future efforts will be directed at inclusion in the model of the multi-echelon distribution structures, i.e. logistics chains that include distribution-to-distribution flows. There is some evidence in the data that multi-echelon distribution systems generate noticeable amounts of goods flow.

7. Acknowledgments

The authors are grateful to Statistics Netherlands for the provision of road transport survey data described and used in this paper. This work is financed by the Dutch research agency NWO in the framework of the Sustainable and Accessible Randstad program. The authors thank Mo Zhan from TNO Sustainable Transport and Logistics for constructing flow visualizations.

8. References

Baumgartner, K., Fuetterer, A., Thonemann, U.W., Supply chain design considering economies of scale and transport frequencies, European Journal of Operational Research 218 (2012) 789–800

Bovenkerk, M.: SMILE+: the new and improved Dutch national freight model system. In: Paper presented at the European transport conference, Strasbourg (2005)

Chen, M. (2011), ETIS and TRANS-TOOLS v1 freight demand. Presented at the CTS-seminar on European and national freight demand models in Stockholm

Chow, J.Y.J., Yang, C.H.Y., Regan, A.C.: State-of-the art of freight forecast modeling:

lessons learned and the road ahead. Transportation 37(6), 1011–1030 (2010)

Combes, P.F. (2010), The choice of shipment size in freight transport. PhD thesis, Universite´ Paris-Est, Paris

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, Igor Y.;Tavasszy, Lóránt A. Antal (2013). Estimation of Warehouse Throughput in a Freight Transport Demand Model for the Netherlands. 92nd Annual TRB Meeting, Washington DC.

de Jong, G., Vierth, I., Tavasszy, L. and Ben-Akiva, M. (2012): Recent developments in national and international freight transport models within Europe, Transportation, Doi: 10.1007/s11116-012-9422-9

Dicken, P. (2003), ‗Global Production Networks in Europe and East Asia: The Automobile Components Industries', GPN Working Paper 7, University of Manchester

Jin, Y., Williams, I., Shahkarami, M., Integrated regional economic and freight logistics modelling: results from a model for the Trans-Pennine Corridor, UK, European Transport Conference 2005

Guilbault, M. (2008), Enqueˆte ECHO « Envois-Chargeurs-Ope´rateurs de transport ». Re´sultats de re´fe´rence (Synthese no 58). INRETS.

Friedrich, H. (2010a), Simulation of logistics in food retailing for freight transportation analysis, PhD Dissertation, Karlsruhe Institute of Technology

Friedrich, H. (2010b), Simulation of logistics in food retailing for freight transportation analysis. In: World Conference on Transport Research (WCTR), Lisbon

Kim, E., Park, D., Kim, C., Lee, J.-Y., A New Paradigm for Freight Demand Modeling: The Physical Distribution Channel Choice Approach, International Journal of Urban Sciences , 14(3), 2010, 240-253

Maurer, H. H.(2008), Development of an Integrated Model for Estimating Emis-sions from Freight Transport, PhD dissertation, The University of Leeds Institute for Transport Studies
Melendez, W. J, Horowitz, A. J., The Transshipment Problem in Travel Forecasting: Tour Structures Model, TRB 2011 Annual Meeting

Liedtke, G. (2006): An Actor-Based Approach to Commodity Transport Modelling. Nomos, Baden-Baden

Liedtke, G. (2009): Principles of micro-behavior commodity transport modeling. Transp. Res. Part E 45, 795–809

SCENES deliverable 4 (2000), SCENES European transport forecasting model and appended module. Technical description, Cambridge (2000)

Tavasszy, L.A., Smeenk, B., Ruijgrok, C.J.: A DSS for modelling logistics chains in freight transport systems analysis. Int. Trans. Oper. Res. 50(6), 447–459 (1998)

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, L A, Davydenko, I Y & Ruijgrok, C (2009). The extended generalized cost concept and its application in freight transport and general equilibrium modeling. In s.n. (Ed.), Proceedings of the 1st conference on the Integration of Spatial Computable General Equilibrium and Transport Modelling (pp. 1-20). Tokyo: SANJO-Hall.

Tavasszy L.A., Ruijgrok K., Davydenko I., (2012). Incorporating Logistics in Freight Transport Demand Models: State-of-the-Art and Research Opportunities, Transport Reviews: A Transnational Transdisciplinary Journal, 32:2, 203-219

TRANS-TOOLS, Deliverable 6 (2008), TRANS-TOOLS Final Report, WP 6. TNO (The Netherlands, Delft)