1. Dimensions and measurement of freight transport demand

Just as passenger transport, freight transport is a derived demand, in this case derived from the product markets. Knowledge about the situation and the trends on these markets is indispensable for understanding freight transport.

Defining freight transport demand is much more straightforward than measuring it. Freight transport demand refers to the required volume of goods movement between different locations. There are many different ways this volume can be measured, and different operational definitions of freight transport correspond with different drivers of demand, policy levers and impacts on society.

Freight transport demand can be measured in terms of:
- shipments (a shipment is a number of product units that are ordered, transported and delivered together).
- tonnes (t)
- tonne-kilometres (tkm),
- vehicle trips or vehicle tours (a tour is a series of consecutive trips starting and ending at the same location, e.g. for picking up goods from several senders or for delivering to several receivers)
- vehicle-kilometres (vkm)
- money values (transport costs or prices paid for transport services).

These measures can be implemented without distinguishing by transport mode, but one can also look at the mode-specific number of t, tkm vkm, trips or transport costs, etc. (e.g. road vkm).

The amounts of t and tkm are determined largely by international and interregional trade patterns (that in turn are dependent mostly on consumer demand and the economic structure of production).

The amount of vkm (similarly for shipments, trips/tours and transport cost/price) is also dependent on logistics\(^1\) decisions, such as on:
- shipment size
- the location and use of terminals, including:
  - consolidation centres, where several shipments come together to be loaded together into a single larger transport unit

---

\(^1\) For a discussion on the relation between transport and logistics, see de Jong and Ben-Akiva (2010).
• distribution centres, where large collections of goods that arrive are separated into smaller separate loads of goods for further distribution
• depots (where storage takes place)
  • the load factor of the vehicle (the amount of tonnes of cargo relative to vehicle capacity) and the amount of empty driving
  • The efficiency in route planning, given the sender and receiver locations.

For vkm or tkm (or other measures of freight transport) by mode, mode choice enters the picture as well. There can be changes in route and time-of-day choice that do not affect the total number of tkm or vkm (by mode). For the external effects of freight transport (especially the impact on local and global emissions, safety, road damage, congestion) vkm by mode is most important. The societal benefits of freight transport are mostly related to the amount of t and tkm transported, and the contribution of the transport sector to GDP is measured by the price paid for transport services (freight rates).

2. The agents involved in decision-making on freight transport

In freight transport, unlike passenger transport, multiple agents are usually involved in the decision-making about a single shipment. The demand for freight transport services is situated either with the sender or the receiver of the goods. In most cases, especially for raw materials and semi-finished goods, the senders (which are producers or wholesalers) are responsible for having their goods delivered to the receivers (consignees). These senders that have a demand for transport services are then called ‘shippers’. For consumer goods that need to be transported to supermarkets, the retailers (being the receivers) are often in charge of organising the transport.

Sometimes the shippers choose to carry out the transport services themselves (meeting their own demand). This is referred to as ‘own account transport’. The lion’s share of the supply of transport services comes from ‘carriers’ and from intermediaries known as third and fourth party logistics service providers. Here we define carriers to include firms that supply road, rail, maritime, inland waterway and/or air transport services. Third party logistics (3PL) service providers perform logistics activities for a shipper (which can be much broader than just transport, e.g. also stock-keeping), whereas fourth party logistics (4PL) service providers integrate capabilities of several organisations, including their own (e.g. multiple 3PLs for different parts of the logistics chain) to obtain a comprehensive supply chain solution.

Other agents involved in freight transport are terminal operators (these could be linked to a shipper, carrier of logistics service provider, but could also be independent), and port and airport organisations/companies, that perform transhipment operations along the transport chains.

3. Choices determining freight transport demand and their drivers

3.1 Typology of choices

For freight transport, as for passenger transport, it is possible to establish a set of choices made by relevant decision-makers that collectively determine the amount and composition of freight transport demand. These choices include (partly from de Jong and Ben-Akiva, 2010):
  • Choices on production and consumption of goods and on trade and distribution. In most cases the underlying choice here is the sourcing decision; the decision of a
producer, wholesaler or retailer from which supplier to buy the goods – this also
determines the geographical location of the supplier and consequently the trade
relation and transport needs.

- Shipment/inventory choices such as shipment size, frequency, etc. result in shipments
  of commodities with a certain weight, size, and value between the point-of-production
  and the point-of-consumption. The shipment’s size and value are important
  characteristics because they affect the mode choice and the load factor. The load factor
  is the weight of the cargo divided by the capacity of the vehicle or vessel.

- Transport chain choices result in a series of modes and vehicle types used
  consecutively for a transport between the point-of-production and the point-of-
  consumption. This includes information on the transhipment(s) between the modes or
  vehicle types for the same mode. A chain contains a single leg using a single mode in
  the case of direct transport. It can also consist of several legs, each with its own mode
  or vehicle type, as depicted in Figure 1. An example of a multi-modal, multi-leg
  transport chain would be: road transport from the point-of-production to a port,
  followed by sea transport to a second port, and finally road transport to the point-of-
  consumption. Transport chain choices include the choice on the number of legs in the
  chain, the mode choice for each leg and the transhipment location(s). These choices
  result in a modal split and affect the vehicle load factor. Together, the transport
  volumes, the mode shares and the load factor determine the number of vehicle-
  kilometres by mode.

- Finding return loads to avoid empty vehicle/vessel returns.

- Time-of-day choices and other timing issues such as the day of the week that produce
  a distribution of traffic over time periods.

- Route choices that yield the distribution of traffic over the network.

Each of these choices can be analysed on either aggregate data, the data at the macro-level of
geographic zones, or disaggregate data, the data at the level of the decision-maker such as
shippers and carriers. Within disaggregate data there is a distinction between data on actual
choices that are referred to as revealed preferences and data on choices in hypothetical
situations, which are referred to as stated preferences. There are partly disaggregate national
freight transport models in countries like Italy and Scandinavia. But most freight transport
models are aggregate models that are sometimes highly segmented by commodity type.

At the macro-level, a distinction can be made between PC (production-consumption) flows
and OD (origin-destination) flows. This is the macro-level equivalent of the distinction at the
micro-level between a transport chain and a leg of a transport chain (de Jong and Kroes, 2014,
also see Figure 1). The transport chain can potentially be a multi-modal transport structure
whereas the leg is a uni-modal structure that is part of a transport chain. By adding the
volumes of the transport chains to and from the same zones, one obtains PC matrices. By
adding the volumes of the legs to and from the same zones, the OD matrices are obtained. PC
matrices contain commodity flows all the way from the production zone to the consumption
zone. These flows may consist of several OD flows. This is because a transport chain may be
used with multiple modes and/or vehicle types as well as one or more transhipments along the
route.

PC flows represent economic relations and transactions within and between different sectors
of the economy. Changes in final demand, international and interregional trade patterns, as
well as in the production structure of the economy, have a direct impact on the PC flow
patterns. Also, the available data on economic linkages and transactions are in terms of PC
flows. They are not in terms of flows between producers and transhipment points, or between transhipment points and consumers. Changes in logistics processes such as the number and location of depots, and in logistics costs, have a direct impact on the composition of the transport chains. This would lead to different OD flows that would only indirectly impact the economic and trade patterns (and hence the PC flows). Assigning PC flows to the transport networks require that they be first transformed into their corresponding OD flows. For instance, the transport chain road-sea-road would lead to road OD legs ending and starting at different ports. This would take the place of a long-haul road transport that would not involve any ports. A similar argument holds for a purely road-based chain that first uses vans to ship goods to a consolidation centre, then consolidates the goods with other flows of goods into a large truck for the main haul, and finally uses a van again to deliver goods from a distribution centre to the C destination. In this scenario, the link assignment depends on the three OD legs. Therefore, converting the PC flows into OD flows, which can be done by a ‘logistics model’, (see de Jong and Ben-Akiva, 2007; and also section 4) is required for a meaningful network assignment. The data available for transport flows from traffic counts, roadside surveys and interviews with carriers are also at the OD level, not at the PC level.

**Figure 1.** Transport chain, transport legs, PC flows and OD flows

<table>
<thead>
<tr>
<th>Mode</th>
<th>truck</th>
<th>ship</th>
<th>truck</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>sender</td>
<td>port</td>
<td>port</td>
</tr>
<tr>
<td>Micro level</td>
<td>transport leg</td>
<td>transport leg</td>
<td>transport leg</td>
</tr>
<tr>
<td>Macro level</td>
<td>OD flow</td>
<td>OD flow</td>
<td>OD flow</td>
</tr>
<tr>
<td>Micro-level</td>
<td>transport chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macro-level</td>
<td>PC flow</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The main reason for consolidation is to share transport costs with other shipments. It allows the use of fuller and larger vehicles and vessels with lower unit cost. Consolidation can take place:

- At the place-of-origin. A shipper can organise batches of products to be transported so that a good match with available transport capacity is obtained.
- In consolidation centres that receive goods from several senders.
- Via vehicles that carry out collection rounds by visiting multiple senders.

Consolidation requires multi-legged transport chains. This would include at least a leg from the sender to the consolidation centre as well as a consolidated leg from the consolidation centre to the consumption zone. Usually, the second leg will be to a distribution centre. There is usually also a third leg from a distribution centre where the consolidated shipment is broken down into multiple shipments for the different receivers. Consolidation can be done with storage at the transhipment location, but can also involve the use of ‘cross-docking’, which has increased substantially in recent years. Cross-docking means that the vehicle from which the goods are unloaded and the vehicle on which the goods are loaded are present at the
terminal at the same time, and that the goods are moved from one vehicle to the other, without any storage.

3.2 Who decides on what?

The relevant agents for production and consumption decisions are of course in first instance the sender and the receiver, but in the end this is driven by consumer demand. Sourcing decisions are the domain of the receivers of the goods.

Managers of the shipper, the carrier and/or the intermediaries may make the transport choices such as the mode choice decision. Interactive decision-making (and especially research methods for dealing with this) is discussed in section 4.5. In general, it is recognised that the shipping firm is the most common decision-maker of mode choice, also for transports carried out by transport suppliers. Many carriers just offer a single mode, and in the case of a multimodal transport chain they may only be involved in a single leg of the chain (e.g. a road haulage firm that provides the first road transport in a road-sea-road transport chain). Logistics service providers on the other hand typically offer door-to-door transport services, and take over responsibility for the entire transport chain.

Even when the receivers are not responsible for organising the transport, they usually are the key decision-maker on the moment in time that the delivery takes place (they typically specify the delivery time window), and also for the shipment size (and thus also the transport frequency) decision, which are determined when they order the goods from the sender. The sender (and its transport suppliers) then have to take the delivery time and shipment size as given.

The firm actually carrying out the transport usually determines route choice. In the case of road transport, truck drivers may have some freedom to choose the route or to change routes as a reaction to unexpected traffic delays.

3.3 Drivers of freight transport choices

In van de Riet et al. (2008) and de Jong and Ben-Akiva (2010) the key drivers of freight transport are identified. The structure of consumer demand and production and the trade patterns are the most important drivers of total freight transport (measured in tonnes or tkm). Logistic developments and attributes of the modes (especially costs, time, reliability, flexibility), on the other hand, are more important drivers of modal split and shipment size.

The following developments are taking place with respect to these drivers:

- Consumer demand is likely to continue to rise in many if not most parts of the world, which in turn would lead to an increase in the number of freight shipments. Furthermore, consumer demand is also likely to become more dispersed spatially due to further suburbanisation, which would lead to increases in transport distances.
- Increasing trade among countries (due to globalisation) is likely to lead to further increases in transport distances.
- Changes in the logistics systems being used:
  - Unit transport costs have decreased over the last decades while unit inventory costs have increased (only in recent years these trends have halted). This
change in the relationship between storage and transport costs has led to the use of the just-in-time (JIT) concept, which has led to a decrease in inventory levels and shipment sizes and an increase in delivery frequency. Trucks are increasingly used as a means of “mobile inventory”. This has led to an increase in vehicle kilometres and an increased demand for van transport instead of truck transport. The growth of JIT transport increases the service requirements of the transport modes, especially with regard to reliability of the transport time (delivery at the agreed time or within the agreed time window) and flexibility (short reaction time between order and delivery). Road transport modes are considered by the industry to perform considerably better than other modes on these factors, and so the growth of JIT transport has improved the competitive position of road transport.

- Technological developments in production facilities and supply chains are facilitating demand-driven production. This has two components. The first component is lean production – the flexible production of (semi-)manufactured goods, whereby the production facility can be reconfigured within hours (instead of days) to switch between products. This enables manufacturers to produce a wide range of products and a wide diversity of a given product at a single facility. The second component is postponement manufacturing. Semi-manufactured goods are produced according to a demand forecast (BTS: built to store) at a central production facility and are shipped to assembly facilities near the market. At the moment a final product is ordered, it can be assembled at the assembly facility, resulting in very short lead times and quick fulfilment. Due to the manufacture of a variety of components, orders can be customised to match the demands of the customer. This influences what needs to be shipped and where it is shipped. These developments put specific demands on the supply chain. The supply chain must be flexible enough to enable short lead times (time between order and delivery) and reliable delivery of products. ICT tracing and planning systems enable the control of material flows, providing real-time information on the status of the products. This has resulted in a restructuring in the management of the supply chain. The various transport modes differ in the way they can meet the demands for shorter lead times and JIT delivery. The dominant view among shippers is that road transport is the mode that can provide the highest flexibility and reliability.

- Another development in supply chain management is the increased use of distribution centres and hub-and-spoke systems. This helps to reduce the costs of distribution facilities, transportation, warehousing, and inventory. Economies of scale can also be achieved by concentrating production facilities in fewer locations and centralising inventory by reducing the number of stockholding points. Inventory centralisation occurs on a larger geographical scale, which results in longer routes in general, but also to a consolidation of traffic flows. Consolidating freight flows leads to higher load factors, use of larger vehicles, and opportunities for alternative modes on the long haul. Larger vehicles are more economical in terms of cost per tonne than smaller ones, provided they are fully loaded. By consolidating freight flows, it is possible to collect sufficiently large volumes for transport over longer distances by vehicles of a larger size. Furthermore, consolidating freight flows, especially in combination with a trend towards more containerisation and an
increase in global trade volumes, makes non-road transport on the long-haul more attractive.

The options available for modes of freight transport generally are road, rail, inland waterways, sea, air and pipeline. Within these modes, several types of vehicles or vessels, such as articulated trucks, can be distinguished. Road transport is generally the most widely available mode. The availability of inland waterways modes and short sea shipping is the most constrained. The characteristics of the different modes are discussed below.

Rail networks are not as dense as road networks and only a few firms have direct access. Railway operations often require reconfiguration of trains at marshalling yards, which is time-consuming and leads to relatively long door-to-door transport times. In this regard, three different rail products can be distinguished:

- Full train loads between private sidings requiring no remarshalling or transhipment. Such services require very large consignment sizes, but then give low costs at any length of haul.
- Wagonload services which require remarshalling to consolidate traffic and are only viable for large flows over long distances.
- Container or other intermodal services which generally use consolidation by road and require long distances (although traffic to or from ports can serve a terminal at the port, and not require transfer by road, except perhaps within the port, and is thus more favourable for rail than other intermodal traffic).

Rail transport requires heavy investments in tracks, signalling, terminal facilities and equipment, some of which can be shared with passenger transport (but then the capacity also has to be shared with passenger trains, which often get priority). Because of these substantial fixed costs, the unit cost of transport is high at low transport volumes and decreases slowly with increasing transport volumes. Rail transport is often the least cost choice for large quantities of goods transported over long distances. For this reason, rail transport was historically regarded as a natural monopoly and even today; a single entity usually manages the rail tracks of a country. However, in many countries rail deregulation and privatisation have taken place. This has resulted in competition between freight transport rail operators. Labour expenses are the main component of transport cost in road transport. The fixed costs are a considerably less significant portion of the total cost than that of rail transport, partly because it uses all-purpose roads supplied by the state. In most countries, the road transport industry is also highly competitive, served by many firms varying in size from large corporations to one-person owner-operators.

Table 1 (from de Jong and Kroes, 2014) compares the pros and cons of the rail and road systems. The table is based on current practices instead of theoretical characteristics of the modes. In some countries, such as the U.S. where freight transport by rail usually gets priority over passenger transport by rail, the pros and cons of the two systems may be somewhat different. Road transport usually scores best on time, reliability, flexibility and accessibility. Conventional rail and combined road-rail transport (intermodal transportation) have relatively better safety and cost features especially for long distances and/or large volumes. In cases where there is a high load factor, inter-modal transportation also produces less emissions of conventional pollutants and greenhouse gases. Rail freight transport has the potential to be very reliable. But in reality, especially in Europe, shippers consider the timeliness of road transport to be superior. It is difficult for rail to hit tight delivery windows. If road congestion keeps increasing, this may change. Train operations are also less sensitive to weather.
conditions than road transport. International rail transport in Europe is still slow and costly due to the lack of interoperability and responsiveness to market forces dictated by national railroads. It can only remain competitive in long distance transport routes over 350-500 km (Beuthe and Kreutzberger, 2001). Therefore, rail transportation is usually used to transport low value bulk cargo where the most important factor is low rates.

For substantial rail market shares in other cargo, a truly inter-modal system with one logistics service provider that is responsible for the entire transport chain while offering reasonably fast and reliable door-to-door services is required.

Table 1. Strengths and weaknesses of road versus rail in freight transport

<table>
<thead>
<tr>
<th>Mode</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Railway</td>
<td>Adequate service level for bulk</td>
<td>Less innovative (information systems)</td>
</tr>
<tr>
<td></td>
<td>Direct transport between large-volume centres</td>
<td>Compatibility in international transport</td>
</tr>
<tr>
<td></td>
<td>Safety</td>
<td>Time and cost for loading and unloading and marshalling (if needed);</td>
</tr>
<tr>
<td></td>
<td>Low emissions</td>
<td>limited opening hours of facilities</td>
</tr>
<tr>
<td></td>
<td>Price (long distance, large volume)</td>
<td>Bottlenecks on some links due to competition with passenger trains</td>
</tr>
<tr>
<td>Truck</td>
<td>Speed</td>
<td>Higher emissions</td>
</tr>
<tr>
<td></td>
<td>Flexibility, timely available</td>
<td>Capacity bottlenecks, congestion risks</td>
</tr>
<tr>
<td></td>
<td>Spatial coverage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Possibilities for consolidation-en-route</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Small consignments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Point-to-point shipments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality of handling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Information systems</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transport time reliability</td>
<td></td>
</tr>
</tbody>
</table>

Waterways and sea routes generally require no infrastructure investments from the carriers since public authorities usually undertake these costs. Carriers only invest in the equipment (Coyle et al., 1996). In the sea transport market, there are generally several large private carriers competing for international market share while smaller carriers compete for inland waterways transport. Water transport vessels tend to be considerably bigger than that of road and rail transport. They offer low unit rates but also low operational speeds. This makes this mode most suitable for the transport of low value goods over long-distances. Inland waterways transport competes with road, rail and short-sea shipping. Ocean shipping has a large market share when measured by shipping volume in tonnes. But in terms of value, a very high percentage of international transport goes by air.

While passenger transport planes are still used for freight transport in what is known as belly freight, airfreight transport via dedicated freight carriers is increasing. Air transport has high variable costs relative to the fixed costs (Coyle et al., 1996). Compared to other modes, the unit costs are quite high especially for large volumes. For long distances, air transport is considerably faster than all other modes. Given its high costs, it is predominantly used for goods that quickly deteriorate in value or that are needed urgently. Some examples are
flowers, newspapers, spare parts, and express mail. But, more recently, the air transport market has been expanding to other goods.

Pipelines are only used for liquid petroleum and natural gas. The investment in the network is very substantial when transporting small volumes. This cost decreases as the quantity increases. The pipeline networks have low density and low speed.

Both intermodal and multimodal transportation use several modes in the same transport chain between point-of-production and point-of-consumption. The main difference is that intermodal or ‘combined’ transportation is carried out for a single flat rate and uses the same loading unit and volumes on all the modes in the chain. This unity reduces the transhipment costs and time as long as specialised equipment for transporting the containers at intermodal stops are available. The most common containers are the eight feet wide and twenty feet or forty feet long containers. Container movements are often measured in TEUs: twenty-foot equivalent units. The use of containers began in the maritime sector by the Sea-Land Company, and has grown tremendously over the past decades. Containers can be used on sea vessels, trucks, trains both single and double stack, as well as inland vessels. Road-sea-road and road-rail-road are regularly used transport chains that often use containers all the way. Instead of containers, intermodal transportation can use swap bodies that have non-rigid sides. Besides container ships, there are also roll on - roll off ships called RoRos, where the road trailers are driven on board. Trailers can also be loaded onto trains (sometimes known as ‘Rollende Landstrasse’, ‘Iron Highway’).

Overviews of elasticities for the impacts of price changes in roads and rail transport can be found in De Jong et al. (2010) and VTI and Significance (2011).

4. Research methods: freight transport modelling

4.1 Available data

An impediment to detailed freight transport analysis from the macro-perspective is that some of the information, especially on transport cost and logistics cost, is proprietary. Firms in freight transport are usually reluctant to disclose this information to clients, competitors and the public.

Data that might be used for freight transport modelling include:

- Trade statistics and customs data that provide information on import and export of goods, usually in monetary units.
- National accounts, including input-output tables. These give the flow of goods and services between sectors of the economy in monetary units. Multiregional input-output tables provide information on trade between sectors in different regions of a country.
- Transport statistics such as roadside surveys providing information on vehicle origins and destinations, port statistics of incoming and outgoing tonnes by commodity type and truck surveys. In EU countries, information on origin, destination, commodity type, and load is collected from a sample of firms with trucks over 3.5 tonnes.
- Shipper surveys that collect detailed information about the firm’s freight transport and selected individual shipments, (e.g. in France; see Rizet and Guilbault, 2004), and commodity flow surveys, where shippers provide a limited amount of information
about a large sample of shipments, (e.g. in the U.S. and Sweden; see Vanek and Morlok, 1998, and SIKA, 2002).

- Consignment bills and RFIDs, which are electronic tags for tracking and tracing, contain detailed information. But this information is not publicly available.
- Traffic counts containing both manual and automated counts that usually distinguish between trucks and cars.
- Network data from road, rail, waterways, etc., and timetables for transport services that operate at fixed times, such as liner services.
- Data on terminals such as location, types of goods, throughput, and costs.
- Cost functions that are usually based on a sample of firms, or engineering assumptions.

4.2 Model structure

Freight transport models are used to assess the impacts of different types of autonomous developments and policy measures, such as changes in national regulations and taxes or infrastructure investments in specific links, nodes and corridors. A wide range of models and model systems are applied by public agencies. Furthermore, a lot of freight transport modelling takes place at universities and at the individual firm level (see Tavasszy and de Jong, 2014, for a reference book). Models to optimise transport and logistics within a specific firm or supply chain are not discussed in this chapter. Nevertheless, there are many things that models for government agencies or models in scientific research can learn from models for the private sector (as will be discussed below).

An overview of national and international freight transport models carried out in 2004 (de Jong et al., 2004), which was updated in 2013 (de Jong et al., 2013), found that:

- Important differences between the freight and passenger transport markets are the diversity of decision-makers in freight (shippers, carriers, intermediaries, drivers, operators), the diversity of items being transported and the limited availability of data, especially disaggregate data (partly due to confidentiality reasons). Most freight transport models apply simplifying assumptions on the first two issues and only use aggregate data.

- At the national and international level, freight transport models usually start from models of economic processes (production, consumption, trade), for instance in the form of input-output models. Urban and regional models, in as far as they are available, often overlook the link to the economy and focus on building up and assigning truck matrices to a network.

- The four-step modelling structure from passenger transport (see Figure 2 below) has been adopted in freight transport modelling with some success:
  - Generation models for production and attraction per sector (e.g. mining) or commodity group (e.g. petroleum)

---

2 There are also freight transport models which do not fit into the four-step model structure and which have no base in passenger transport modelling, such as models that explain the share of the monetary expenditure on a certain transport mode in total production cost, which are based on the economic theory of the firm.
o Distribution models, sometimes with a dependence of the distribution on the transport resistance between zones from the modal split model

o Modal split models

o Network assignment,

However, additional steps are often needed to transform trade flows in money units to physical flows of goods in tonnes and further into vehicle flows with specific vehicle utilisation factors. These additional processes can be modelled as fixed rates, but also by explicit representation of logistics choices. Also other logistics aspects that are related to the trade-off between transport and inventory costs are usually not included in freight transport models, even though the logistics solutions of firms influence the mode split.

**Figure 2.** The conventional four-step model in freight transport

A better representation of the freight transport model system might be the following (Ben-Akiva, 2011).

**Figure 3.** A revised representation of the structure of the freight model system
This structure has fewer steps (only three models: economic activity choices, logistics choices and route choice), but each of the first two steps includes several related choices. This model representation is consistent with the distinction between PC flows and OD flows discussed in section 3.

An important question then is whether each of the submodels (the three groups of choices) in Figure 3 should be modelled at the aggregate or the disaggregate level. We think that, given the current data situation and computer technology, much can be said for modelling the logistics choices at the disaggregate level and the other choices at the aggregate level. This leads to the ADA model (Ben-Akiva and de Jong, 2013), as presented in Figure 4.

The general structure of Figure 3 is worked out in the more specific ADA model structure (see Figure 4). But there are more ways to work out the general structure. The boxes in Figure 4 indicate model components. The top level of Figure 4 displays the aggregate models. Disaggregate models are at the bottom level. So in the ADA model system, we first have an aggregate model for the determination of PC flows, then a disaggregate “logistics” model, and finally another aggregate model for network assignment.

The model system starts with the determination of flows of goods between production (P) zones and consumption (C) zones (being retail for final consumption; and further processing of goods for intermediate consumption). These models are commonly based on economic statistics (production and consumption statistics, input-output tables, trade statistics) that are only available at the aggregate level (with zones and zones pairs as the observational units). Indeed, to our knowledge, no models have been developed to date that explain the generation and distribution of PC flows at a truly disaggregate level.

The aggregation of OD flows between firms to OD flows between zones provides the input to a network assignment model where, in the final step of the ADA model, the zone-to-zone OD flows are allocated to the networks for the various modes. Assignment can, in principle, be done at the level of individual vehicles (microscopic or mesoscopic models for simulating route choice). Most model systems perform an assignment of aggregate zone-to-zone flows (possibly with several user classes) to the networks in order to use available software, to keep the model tractable and to keep the run time manageable. This approach is also used because the network level is the level at which validation/calibration data are usually available (e.g., traffic counts at various locations/screenlines).

Figure 4. ADA structure of the (inter)national/regional freight transport model system
Most existing freight transport model systems include submodels for generating PC or OD matrices (possibly by mode) and routines for assigning either one of these matrices to the networks (unimodal or multimodal). Assignment of PC flow to the networks would not be correct (see section 3). A relatively new phenomenon of the ADA model is the inclusion of a logistics model that on the basis of PC flows produces OD flows for network assignment. The logistics model consists of three steps:

A. Disaggregation to allocate the flows to individual firms at the P and C end;
B. Models for the logistics decisions by the firms (e.g., shipment size, use of consolidation and distribution centres, modes, loading units, such as containers);
C. Aggregation of the information per shipment to origin-destination (OD) flows for network assignment.

This model structure allows for logistics choices to be modelled at the level of the actual decision-maker, along with the inclusion of decision-maker attributes.

4.3 Model components

In this subsection we discuss the model components that were distinguished in Figures 2, 3 and 4. This discussion is partly based on de Jong et al. (2004), de Jong and Ben-Akiva (2010) and de Jong et al. (2013).

Models for economic activity choices

These are models that (by themselves, or together with other models) explain and forecast PC matrices.

The simplest method is the application of growth factors on a PC or OD matrix for a base year. This requires that a base year PC or OD matrix has been established. A base PC matrix could be determined on the basis of international and/or regional trade data, or on the basis of a Commodity Flow Survey. A base OD matrix can be based on transport statistics (by mode) and traffic counts. Matrix expansion is usually done by putting growth factors on the row totals (the amount of goods sent by the zones, by commodity type) and on the column totals (the amount of goods received by the zones, by commodity type) and then using numerical procedures (e.g. Furness) to obtain an internally consistent matrix again. The forecasts for the row and column growth factors (production and attraction) can come from time series analysis of trade or transport data and their underlying drivers (such as GDP, possibly by sector) or from the application of GDP elasticities based on the literature.

In a distribution module of a freight transport system, the trade flows (in tonnes) between origin zones and destination zones are determined based on measures of production and attraction and a measure of transport resistance. The latter is expressed as transport cost or generalised transport cost. The most commonly used method is the gravity model. In such models, the flow between zone i and zone j is a function of the product of production and attraction measures of zone i and zone j respectively, divided by some measure of the (generalised) transport cost between the zones. Gravity models for distribution in freight are for instance included in:

- The Dutch SMILE model (Tavasszy et al., 1998);
- The Great Belt traffic model (Fosgerau, 1996);
- The new BasGoed model for The Netherlands (de Jong et al., 2011)
The main advantages of gravity models are their limited data requirements (only aggregate zone to zone flows and aggregate explanatory variables, including transport cost, are required). Inclusion of (generalised) transport cost also implies that the distribution becomes policy sensitive (policy measures that change the costs or time between the zones for one or more modes then can have an effect). Disadvantages are the limited scope for including explanatory factors and policy effects and the limited number of calibration parameters (a complicated spatial pattern is represented with just a few degrees of freedom; this often leads to a rather poor fit).

Input–output models are basically macro-economic models that start from input-output tables. These are tables that describe, in money units, what each sector of the economy (e.g. textile manufacturing) delivers to the other sectors, also including the final demand by consumers, import and export. National input-output tables have been developed for many countries, usually by a central statistical office. A special form of input-output table, which for many countries does not exist, is a multi-regional or spatial input-output table. This not only includes deliveries between sectors, but also between regions (trade flows). Most multi-regional input-output tables distinguish only a few, large regions within a country. The input-output model assumes that for forecasting, the multi-regional input-output table can be scaled up on the basis of predicted sectoral growth. The new input-output table can then give the future trade flows between regions, using either:

- Fixed technical and trade coefficients: the present production and trade patterns are extrapolated into the future.
- Elastic technical and trade coefficients: functions are estimated (e.g. multinomial logit) in which the fraction that is consumed in region i of the production of sector s in region j depends on the total production of region j in sector s and the (generalised) transport cost, in relation to other regions. This makes generation and distribution sensitive to changes in transport cost and time (a form of induced demand).

Examples of multi-regional input-output models in freight transport are:

- The Italian national model system for passengers and freight (Cascetta, 1997; Marzano and Papola, 2004), which uses 17 sectors and 20 regions and also has elastic coefficients.
- The SCENES European model system for passengers and freight and its predecessor STREAMS, Strategic Transport Research for European Member States (Leitham et al, 1999), with 33 sectors and more than 200 zones in Europe and elastic coefficients (SCENES Consortium, 2001).

These models have a one-way dependency between land use and transport: the spatial distribution influences transport. A feedback to land use (e.g. in the form of a logsum variable, which gives the expected utility from the choices in the transport model) is possible in theory, but it would add considerably to model complexity and run times. An advantage of input-output modelling is that it provides a clear link between economics and transport (and freight transport is a derived demand). Disadvantages (also see de Jong et al., 2004) are the assumption of fixed production technologies, the need to convert from money units to weight units and the heavy data requirements: for many countries, up-to-date input-output tables, certainly multi-regional ones, are not available.
Computable general equilibrium (CGE) models are models establishing equilibrium in several related markets (e.g. commodity market, labour market). CGE models in economics (not focussing on transport) often include economic issues that are not handled in transport models, such as type of competition and economies of scale. Spatial CGE or SCGE models (e.g. Bröcker, 1998) have been developed recently, and are now becoming operational models for (inter)national and regional forecasting and evaluation. Such spatial models also include transport and the influence of the economy on transport as well as the impact of transport on the economy (which is an indirect effect of transport projects). Examples of SCGE models are:

- The CGEurope model (Bröcker, 1998), which is part of the transport model for Europe developed for the European Commission (Transtools).
- The PINGO model is a static Spatial Computable General Equilibrium (SCGE) model of the Norwegian economy (Vold and Jean-Hansen, 2007). It is based on a 2002 base year Social Accounting Matrix. The model includes an explicit production function for the transport sector. The cost data that act as input come from the Norwegian transport logistics model.
- The Dutch SCGE model RAEM (Ivanova et al, 2007), is a recursive dynamic model of production, consumption and trade for the Netherlands, including international trade, federal government as a macroeconomic agent, labour markets and migration. The model uses cost data as input from the Dutch and European freight and passenger transport network models.

The main advantage of SCGE models is also their main disadvantage: these models are very broad (transport, economy, land use) and can give interactions between markets that other models are not capable of, but the way in which transport is depicted in these models is often highly simplified.

Models for logistics choices

This group of models includes standard models for modal split at the OD level, both aggregate and disaggregate, but also models that determine transport chains, given the zones of production and consumption, as well as models determining shipment size.

Mode choice in freight transport is usually studied in isolation (or in combination with network assignment, as multi-modal assignment). However, mode and shipment size are closely linked decisions (Johnson and de Jong, 2010). Large shipment sizes usually coincide with higher market shares for non-road transport, whereas there is a high correlation between road transport and small shipment sizes. Previous models that combined both choices in a single model include McFadden et al. (1985), Abdelwahab and Sargious (1992), Abdelwahab (1998) and Holguín-Veras (2002). These authors estimated joint discrete-continuous models, where mode choice is the discrete and shipment size the continuous choice. Other authors (e.g. Chiang et al., 1981; de Jong, 2007) have studied this joint decision-making problem as two simultaneous discrete choices, classifying shipment size into discrete categories.

In the context of the ADA model, the logistics choices should include shipment size and the whole transport chain between P and C. The choice alternatives then are different available chains (e.g. road-rail-road, road-sea-road). Some firms within specific zones may have direct access (and egress) to rail, inland waterways or sea transport because they have their own rail sidings or quays. If these chains are not available for a sending/receiving firm, then rail and water transport will only be available by adding access/egress transport to/from a railway.
station/port, which usually takes place by road transport. Transport chains may also include distinctions between vehicle or vessel types within modes (so that for instance the sequence small truck-large truck- small truck also becomes a choice alternative), and one might also include a distinction between containerised or not. Determination of the transport chain together with shipment size takes place in the Norwegian and Swedish national freight models (see de Jong et al., 2013), using deterministic minimisation of total logistics costs. A model of transport chain and shipment size choice, estimated on disaggregate data is Windisch et al. (2011). These models take the locations of transhipment points, where consolidation, distribution and storage can take place, as given (these might be changed manually in specific scenarios). The SMILE and SMILE+ (Tavasszy et al., 1998; Bovenkerk, 2005) included aggregate models for the determination of the optimal locations for national and European distribution centres.

Freight transportation can shift to other time periods of the day if satisfactory alternatives offering better transport time or cost advantages are available. Shifts to time periods that avoid congestion and result in shorter travel times may take place in freight transport. Shifts toward night-time transport have only occurred on a small scale so far. Most long-distance freight transport is already avoiding rush hours. But city distribution often cannot do this because of the delivery time window. Holguín-Veras et al. (2007) argue that the effectiveness of tolls in shifting freight traffic to off-hours outside normal working hours might be limited. Even when the carriers could pass toll surcharges to their customers, the price signal is of no consequence when compared to the receivers’ incremental costs due to off-hour deliveries (staying open longer). To overcome this, Holguín-Veras et al. (2007) propose the provision of tax incentives to receivers willing to do off-hour deliveries. Other objections for off-hour deliveries are overtime wages and community objections.

**Route choice/network assignment modelling**

See Ortuzar and Willumsen (1994) or Hensher and Button (2008) for a treatment of assignment methods. In some transport models, passenger cars and trucks are assigned simultaneously to the same networks since they compete for the same network capacity. In several models, a ‘congestion feedback’ effect is included. This means that after the initial transport demand has been loaded on to the networks, new transport times, which include time losses to congestion, are calculated. These new times are then re-inserted into the demand models. This can create a change in demand such as a shift to less congested modes. The new demand is assigned again to the networks, and so on in a demand-supply iteration cycle until equilibrium is reached.

4.4 Valuing attributes in freight transport

Monetary valuations of quality attributes in freight transport, especially the value of transport time savings (VTTS), are used regularly in the appraisal of infrastructure projects and sometimes also in generalised cost functions in transport forecasting models. Recently there is a tendency to try to include reliability of transport time in project cost-benefit analysis as well.

The valuation methods used can first of all be classified into factor-cost methods and modelling studies (de Jong, 2008).

The factor-cost method is used in most countries that have official freight VTTS for cost-benefit analysis. It tries to find the cost of all input factors that will be saved in case of travel
time savings. A decrease in travel time could release production factors (e.g. labour, vehicles) to be used in other shipments. These items can be calculated using data on wages and vehicles. There is no consensus on the issue whether fuel cost, fixed cost of transport equipment, overheads and non-transport inventory and logistic cost should be included.

The modelling studies for valuation can be classified into revealed preference (RP) studies and stated preference (SP) studies. Joint RP/SP models are also possible in freight, but have been very few so far. RP studies in freight use data on the actual behaviour of shippers, carriers, intermediaries or drivers. A number of situations can be thought of in which these decision-makers have to trade off time against cost:

- mode choice between a fast and expensive mode and a slower and cheaper mode;
- choice of carrier, or between own account transport and contracting out;
- choice between a fast toll route and a congested toll-free route; and choice of supplier.

In an SP freight VTTS study, decision-makers (in practice: shippers or carriers) are asked to elicit their preferences for hypothetical alternatives constructed by the researcher. These hypothetical alternatives refer to transports that have actually been carried out and will have different attribute levels for transport time and cost, and possibly also for other attributes of the shipment. The setting of the SP experiment can be that of mode choice (e.g. repeated pairwise choices between a road and a rail alternative for the same shipment: between-mode experiment) or route choice, as in the RP. Good experience in freight VTTS research however has been obtained in abstract time versus cost experiments in which all alternatives that are presented refer to the same mode and the same route. In an abstract time versus cost experiment the alternatives have different scores on travel time, travel cost and possibly other attributes, but the alternatives are not given a mode or route label, such as “rail transport” or “motorway with toll”.

SP data has some advantages in the case of freight modelling, in particular as it may be possible to obtain data (e.g., on costs and rates) which would be difficult to acquire by other methods (also see Fowkes et al., 1991). Practically all SP surveys in freight transport have been carried out as computerised interviews, which can provide the highest degree of customisation.

Carriers are in the best position to give the VTTS that is related to the costs of providing transport services. Modelling results in the Netherlands and other countries so far (see de Jong, 2008) indicate that the VTTS that is related to the transport services is more or less equal to the vehicle and labour cost per hour (the ‘factor cost’), at least for road transport.

Shippers that contract out are most interested in another VTTS, that is the VTTS that is related to the goods themselves. This includes the interest costs on the capital invested in the goods during the time that the transport takes (only important for high-value goods, but we did not impose a definition of high value on the respondents), a reduction in the value of perishable goods during transit, but also the possibility that the production process is disrupted by missing inputs or that customers cannot be supplied due to lack of stock. The latter two arguments are also (possibly even more so) important for the value of transport time reliability.

If both VTTS components are properly distinguished, the carrier VTTS and shipper (contract out) VTTS can be added to obtain the overall VTTS for use in societal cost-benefit analysis.
4.5 Interactive decision-making

Even though for many decisions it is possible to identify a most common single decision-maker, multi-agent interactive decision-making is relevant for a lot of choices in freight transport. It therefore makes sense to represent such decisions (certainly if several choices are modelled simultaneously, such as mode and shipment size) in a model of behavioural interactions between different agents in freight transport. This is a relatively new field. Some contributions are (also see the freight modelling review of de Jong et al., 2013):

- Hensher (2002) and Hensher and Puckett (2005) developed interactive agency choice experiments (IACE). These include sequential choices, where agents are informed about the previous choices of other agents, and correlation over alternatives and choice sets within and between agents. This is done for pairs of agents (e.g. shipper-carrier) and the process of feeding back information continues for all pairs where agreements have not been reached. A disadvantage of IACE is that the survey costs of interviewing shippers on the responses of the carriers etc. can be quite high (compared to more standard stated preference surveys) and that the resulting samples are small. This led to the development of minimum information group inference, MIGI (Puckett and Hensher, 2006), where agents are interviewed only once, but in the analysis each shipper is matched with a carrier and their group decision making is inferred.

- Friesz et al. (2008) presented a dynamic game-theoretic model with sellers and receivers of products and transporters in the context of an urban transport network, extending earlier models with shippers and carriers only.

- In Holguín-Veras et al. (2006) interactions between carriers and receivers were studied in the context of deliveries outside the peak period (decision making on delivery times).

Holguín-Veras et al. (2011) used experimental economics (with students as participants) to study interactions between shippers and carriers in mode and shipment size choice.

5. Scope for transport policy.

The public authorities’ main objectives for transportation and land use policies with respect to freight transport demand are to promote accessibility and decrease the negative external and environmental effects of transportation. Apart from this, security in transportation is also an important reason for government intervention. Among the key drivers of freight transport demand, the most important driver are growth in the gross domestic product (GDP), growth in trade and the growing distance between suppliers and receivers. International trade has been growing considerably faster than global GDP. In Europe, the creation of the single market has contributed to the growth of freight transport. Logistic developments such as the centralisation of inventories, emergence of hub ports and airports, reduction of lead time (as part of just-in-time supply chains) and increase of business-to-business electronic information exchange (see Ruijgrok, 2001) and the performance of the modes (in terms of the attributes costs, time, reliability and flexibility) are important drivers for modal split and shipment size.

Transport and spatial planning policies do not focus on the important drivers of total freight transport as were mentioned above. Hence they can only have a limited impact on the volume of freight transport. Transportation and land use policies can significantly impact on freight
transport in the following way (also see van de Riet et al., 2008; De Jong and Ben-Akiva, 2010):

- These policies can affect residential and business location choices. For example, regulations or financial incentives to businesses can stimulate mixed land use or spatial clustering.
- Governments can promote certain logistic developments. For example, they can encourage the emergence of information brokers and other means of providing data on consignments.
- Transport policy can influence some of the characteristics of the modes such as transport time by investing in network links and public or public/private terminals as part of city logistics. They can also affect cost by implementing taxes, tolls and subsidies. Note however that several logistics costs components have already been increasing considerably in recent years, due to fuel price increases, driver shortages and environmental concerns. In some segments of the freight transport market such as low value goods and long distance shipments, the modal split is rather responsive to time and cost changes by mode. Cost increases might also lead to increases in the efficiency of transport (higher load factors, less empty driving) and be an incentive to choose more nearby suppliers. The scheduling of freight transport and of distribution in particular may be influenced by delivery time windows, time-of-day specific vehicle bans and by time-of-day specific road pricing. These restrictions are more effective when applied in combination with incentives for the receivers to change delivery times to periods outside the normal business hours. The choice of vehicle type can be influenced by vehicle weight and height restrictions, emission standards or load factor requirements in specific areas such as city centres.
- Introducing more competition into the rail freight sector through government policies might lead to improved service characteristics such as innovativeness, reliability, interoperability, and flexibility of the rail freight sector vis-à-vis the road sector.

In summary, the most effective policies for affecting freight transport demand would appear to be policies focused on changing transport mode characteristics and spatial planning policies. Such policies have the potential for producing significant changes in freight transport demand. In particular, they can shorten transport distances and improve the way the transport system is used (by affecting the choice of mode, time-of-day, and route). Some of these policies will take a long time to implement and for their results to be realised (such as land use planning and infrastructure expansion), while others can be implemented more quickly and will produce their desired effects in the short term (such as road pricing). Because large shifts in demand will be needed to produce a sustainable freight transport system and time is of the essence, both short-term and long-term actions should be considered.

References


