Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 11 (2015) 515 – 530

Transportation Research Procedia

www.elsevier.com/locate/procedia

10th International Conference on Transport Survey Methods

# An innovative integration methodology of independent data sources to improve the quality of freight transport surveys

Oliver Roider [a]*, Gerd Sammer [a], Christoph Link [a], Rudolf Bauer [b], Werner Schachinger [b]

[a] *Institute for Transport Studies, University for Natural Resources and Life Sciences, Peter-Jordan-Strasse 82, 1190 Vienna, Austria*
[b] *Paradigma Consulting, Mariahilfer Strasse 47/1/3, 1060 Vienna, Austria*

**Abstract**

Past experiences show that data of the official Austrian freight transport statistics are often underestimated. Therefore, a methodology was developed, merging existing independent road freight transport data to a consistent and valid road freight matrix. The methodology comprises four steps, using data of the Austrian and European freight transport statistics, data of roadside interviews of truck drivers, and data of counting stations and toll gantries. The methodology was applied to data from the year 2009. Results show the reliability and plausibility of the methodology, indicated by a high correlation with high quality roadside traffic counts.


Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC

*Keywords:* freight transport statistics, freight transport modelling, data reliability, data validation

## 1. State-of-the art and state-of-practice data collection and data processing for freight statistics

Freight data are mostly based on sample data collections at a national level. To render meaningful results, the data from the sample need grossing up to the relevant population, and a weighting procedure is required. This can be done by "traditional" grossing up procedures which do not use origin-destination (OD) or route-information. The

* Corresponding author. Tel.: +43-1-47654-5307; fax: +43-1-47654-5344.
*E-mail address:* oliver.roider@boku.ac.at


Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC
doi:10.1016/j.trpro.2015.12.043

results are only based on information drawn from the sampling variables, e.g. the ownership rate of heavy goods vehicles (Eurostat 2011). An alternative is a weighting and grossing up procedure which uses information about origin-destinations (OD) or route-information of trips. The required information is commonly supplied by freight transport models; however, this procedure might be associated with a modelling error. Validating the result by using independent counting data instead is an alternative, but this approach is rarely used.

Most countries use (freight) transport models for infrastructure planning as well as for any impact assessment of possible legal and fiscal transport policies. Freight transport models require an origin-destination-matrix of transport flows. These matrices can be derived from economic data showing the locations of production and consumption of goods and the commodity flows between these locations. In this context, the OD-matrix is a result mainly of three steps of a multi-step approach, including trip generation, trip distribution, and modal split, but it does not take any network assignment into account. Transport models using economic data are more suitable for predictions since forecasts of economic developments are more accurate than forecasts of transport flows. However, these models have to model logistic choices with regard to several aspects, such as size of inventory, shipment size, loading unit, or trip chains. This makes them quite complex, synthetic, and often empirically not well founded (Tavasszy et al. 2012). Alternatively, other transport models use OD-matrices gathered from commodity flow surveys. Data are processed by statistical means and fed into transport models (de Jong et al. 2012). Hereafter, a literature review of selected examples is documented, which uses freight transport models for grossing-up and weighting of freight survey data as well as for a validation by independent traffic counts. Admittedly, it is very difficult to get detailed information on the grossing-up, weighting, and validation procedures for freight transport surveys and statistics. The result of the literature review indicates a lack of published methodology for freight transport issues.

The freight transport model of the Netherlands, which is part of the national transport model, follows the traditional four-step-model approach. A particular feature of this model is the consideration of logistics activities: an econometric model describes the spatial and functional relationship between production and distribution of goods based on a gravity model (Tavasszy et al. 2010). SLAM (Spatial Logistics Appended Module) is an additional tool which enables an optimization of logistic centres in Europe. Main task of this module is the transformation of economic OD-relations between different regions (based on economic models) into freight transport volumes. Results of the SLAM module are used as input for the transport model in order to calculate a modal split and transport routes (Combes and Leurent 2007).

In France, the freight data are based on the national commodity flow survey and data provided by Eurostat. The French national transport model (MODEV) includes passenger mobility as well as freight transport and is used for weighting and validation. It uses a 4-step approach. Trip generation is the result of a regression model, using cross-sectional data for estimating purposes. A gravity model is used for the distribution, and an aggregate logit model for the modal split, taking road, rail, combined road-rail, and inland waterway transport into account. The calibration of the model is mainly based on the national commodity flow survey and data provided by Eurostat but also on freight traffic counting (de Jong et al. 2012).

The U.S. Freight Analysis Framework published by the U.S. Department of Transportation provides OD-matrices of freight transport volumes by traffic types (domestic, transit, origin, and destination traffic), transported goods, and transport modes (road, rail, ship, air, multimodal transport). The main data source is a commodity flow survey among enterprises in selected sectors of trade, conducted every 5 years (U.S. Census Bureau 2011, U.S. Department of Transportation 2015). However, certain sectors, such as the construction industry, retail, and foreign trade are not covered. These "out-of-scope" flows are estimated by including further data sources (data of industrial output, the employment rate, regional indicators etc.). For example, the freight transport volume of the retail sector is estimated based on input/output information of the national accounts. The commodity flow survey is the most important data source, but additional ones are used as well (Krieger 2011):

- roadside interviews of truck drivers,
- license plate matches at ramps of highways,
- weigh-in-motion measuring by slap detection on highways,
- GPS data of trucks, transmitted automatically and anonymously.

Missing values of the OD-matrix are calculated by using an iterative log-linear model to compensate for deviations from the total flow of goods reported in the commodity flow survey. Data of waterborne commerce, waybills (consignment notes) as well as data of the previous survey are taken into account. The assignment of trips on the road is done by means of a freight transport model.

In South Korea, the commodity flow survey conducted every 5 years consists of various sub-surveys which use different techniques, locations of data acquisition, and address specific companies. The survey includes a shipper survey (mining, manufacturing, and wholesale), truck diary survey, traffic counts at logistic hubs and toll gates, as well as a warehouse and a hazardous material survey. Based on the resulting data, figures for freight volumes at origin and destination (trip generation) are generated by using a statistical weighting procedure, while the freight distribution is calculated by using gravity models calibrated for each commodity type. For the assignment of flows in the highway network, transport volumes are converted into vehicle trips (Park and Hahn 2014).

## 2. Problems with current approaches used to generate freight statistics

The problems which occur when generating valid freight statistics for Austria are representative for most other countries. To compile the national freight transport statistics in Austria, commodity surveys are conducted on an annual basis. Randomly selected freight forwarders and logistic enterprises have to fill-in questionnaires, indicating the daily trip patterns of some of their vehicles including information about goods transported. The answers are weighted and grossed up to provide national transport statistics including an OD-matrix for freight transports. Several authors argue that this procedure results in insufficient data quality: Freight flows would regularly be underestimated in official freight statistics since these statistics often suffer from random as well as systematic errors due to the methodology used for the collection and processing of the data (Sammer et al. 2009, Sammer et al. 2008, Grafl et al. 2007, Spiegel 2003 and 2007, Pischinger et al. 1993). This underrepresentation can be proven by using independent information of traffic counts and energy consumption data. The example of the UK shows a difference of 29% of total HGV kilometrage in 2006 depending on whether the data are obtained by road side traffic counts or based on the continuing survey data (McKinnon & Leonardi, 2009). A main reason is that the task of filling-in the questionnaires requires a considerable amount of time. Thus, it is reasonable to assume that companies do not dedicate enough time to filling in answer to all questions manually and thoroughly, which leads to surveys with incomplete datasets. This seems especially true for the retail sector, where goods are delivered via subsequent short trips and are often bunched together and then recorded. Shortcomings could be reduced by using available and highly accurate data of the freight transport on motorways and express roads provided by counting stations or toll gantries as well as OD-information derived from toll gantry data. In particular, toll gantries make it possible to estimate further trip information, such as routes and distances travelled, and they provide vehicle features and information about all trucks independent of the country where the transport company is based. However, the currently used method to gross up survey data of the official statistics in Austria and other European countries does not include these data – contrary to the approaches used in the U.S. or in South Korea. More accuracy could be achieved by combining different and independent data sources (Wermuth et al., 2004). However, according to Kockelman et al. (2009), research is needed regarding how different freight traffic data sources can be combined in order to enrich existing statistics (Kockelmann, Browne, & Leonardi, 2009). Such an approach is explained in the following section.

## 3. Data availability

The research project IMoVe-Gueter (2014), commissioned by the Austrian Ministry of Transport, Innovation and Technology aimed at (1) develop a methodology by combining official freight statistics based on a commodity flow survey with other independent data and (2) to apply the methodology in order to generate a consistent OD-matrix for freight transports. The following types of data were considered for the methodology:

Official Road Freight Transport Statistics

Official freight transport statistics for Austria are mainly based on a periodic survey among Austrian freight forwarders. They report trip patterns of selected heavy goods vehicles (HGV), such as origins and destinations of the vehicle, loaded weight and kind of good(s) transported as well as vehicle characteristics, such as the number of axles and the maximum permissible weight. The cells of the model considered are based on the 99 so-called “political districts" (administrative districts) in Austria and on NUTS[1]-3-level abroad. In total, the sample comprises trips of roughly 26,000 vehicles per year with a payload of more than 2 metric tons (Karner et al. 2012, Eurostat 2011). The data – as those of all other European countries– are passed on to the Statistical Office of the European Commission (Eurostat). These data from the 28 EU member states and Switzerland, Lichtenstein, and Norway are merged on NUTS-3-level (European Commission 2014). Survey methods, aggregation level, and types of vehicle considered differ, because it is left to the individual countries how they handle this.

Cross-Alpine and Cross-border Freight Transport on roads (CAFT)

The CAFT-survey (cross Alpine freight data) aims at identifying international road traffic starting in, ending in, and passing through Austria. Data about the cross-Alpine and cross-border freight transport in Austria are collected every 5 years. The last survey took place in 2009 and was organized as roadside survey of drivers of HGV at 9 Alpine crossings and 9 border crossings. It is commissioned by the Federal Ministry for Transport, Innovation and Technology. The content is comparable to the official national statistics. In 2009, the sample covered about 22,000 vehicles (Herry et al. 2011).

Data of the road toll system for HGVs on the higher class road network and from automatic vehicle counting stations

In Austria, vehicles with a total weight above 3.5 metric tons have to pay a toll when using motorways and express roads. Toll gantries between two interchanges (toll sections) register vehicles passing underneath and record the number of axles and the environmental performance indicator of the HGV. In 2009, there were 886 toll sections in the Austrian motorway and express road network recording automatically 580 million passages of HGVs per year (Asfinag 2012). 932 permanent HGV-counting stations were located in the Austrian road network in 2009. Since the devices at the counting stations differ from a technical point of view, the recording of the HGV volume is not done in a standardized way. However, the number of vehicles passing the counting stations (sometimes differentiated by the number of axels) can be used (DTV-Verkehrsconsult GmbH 2011). The stamp recording the passing time in combination with a HGV-identification code of each vehicle makes it possible to estimate the route and the OD-relation of the entrance and exit of the vehicle to the motorway and express road network. However, this information was not available for this research project.

The Austrian Transport Model

The Austrian Transport Model allows trip assignments of passenger and freight transport matrices. The aggregation level is municipalities in Austria and NUTS-3-level or beyond in the other European countries. It includes the transport infrastructure of all roads of regional importance. All toll gantries and automatic counting stations are incorporated in the transport model (TRAFICO et al. 2009).

## 4. Methodological Approach

The combination of the different data sources into a consistent matrix takes four methodological steps, each resulting in a (preliminary) freight transport OD-matrix. Each matrix is subjected to a plausibility check.

### *4.1. Step-1: Spatial adaptation of the freight matrices derived from the official Austrian freight statistics*

The objective of Step-1 is a disaggregation of the official Austrian freight statistics to the level of municipalities. The basis are figures from the official Eurostat statistics (European Road Freight Transport Statistics, ESGVS) on NUTS-3-level as well as a special analysis of the national statistics of HGVs registered in Austria (Austrian Road Freight Transport Statistics, SGVS), broken down to each region of Austria. Both sources show the amount of HGV

[1] Nomenclature of Territorial Units for Statistics

traffic for 2009 measured in total trips per year, but differ in their spatial references. ESGVS contain trips on NUTS-3-level, i.e. all trips between NUTS-3-regions in Austria, or trips between a NUTS-3-region in Austria and one abroad. The special analysis of the national statistics of HGVs registered in Austria is based on political districts. Each political district can be clearly assigned to a particular NUTS-3-region. Districts $i \in K$ or $j \in L$ are indexed with $i_K = 1, \ldots, N_K$ or $j_L = 1, \ldots, N_L$. The sum of the political districts assigned to NUTS-3-region ($N_K$, $N_L$) is equal to the sum of all political districts in Austria (N):

$$\mathrm{N} = \sum_{\mathrm{K}=1}^{M} N_K = \sum_{\mathrm{L}=1}^{M} N_L \tag{1}$$

Step-1 comprises the consolidation and harmonization with regard to the spatial aggregation. Firstly, the principle of maximum entropy is applied in order to estimate the missing disaggregate OD information of the ESGVS data at the level of political districts in Austria. Thus it attempts to fulfil Shannon's entropy function (King et al. 2004):

$$-\left(\sum_{ij} f_{ij} \ln f_{ij} + \sum_{iC} f_{iC} \ln f_{iC} + \sum_{Cj} f_{Cj} \ln f_{Cj}\right) \tag{2}$$

$f_{ij}$ … number of HGV trips from political district i to political district j within Austria (domestic traffic) [trips / year]

$f_{iC}$ … number of HGV trips from political district i in Austria to district C abroad (origin traffic regarding to Austria) [trips / year]

$f_{Cj}$ … number of HGV trips from district C abroad to political district j in Austria (destination regarding to Austria) [trips / year]

The maximisation of the entropy is an iterative process which meets the following requirements of an OD matrix:

I. Domestic trips $F_{KL}$ recorded in the ESGVS between two NUTS-3-regions K, L must be equal to the sum of calculated trips between districts of these regions (trips within a NUTS-3-region included). M represents all NUTS-3-regions.

$$\mathrm{F_{KL}} = \sum_{i=1}^{N_K} \sum_{j=1}^{N_L} f_{ij} \qquad \mathrm{K, L} = 1, \ldots, \mathrm{M} \tag{3}$$

II. The sum of all trips recorded in the ESGVS originating in country C and ending in the Austrian NUTS-3-region L ($F_{CL}$), must be equal to the sum of calculated trips originating in country C and ending in the district j of the NUTS-3-region considered. P represents all countries considered and L all NUTS-3-regions in Austria.

$$\mathrm{F_{CL}} = \sum_{j=1}^{N_L} f_{Cj} \quad \mathrm{L} = 1, \ldots, \mathrm{M} \quad \mathrm{C} = 1, \ldots, \mathrm{P} \tag{4}$$

III. The sum of all trips recorded in the ESGVS ($F_{KC}$) originating in NUTS-3-region K in Austria and ending in foreign country C must be equal to the sum of all calculated trips from the district of the Austrian NUTS-3-region considered to country C.

$$\mathrm{F_{KC}} = \sum_{i=1}^{N_K} f_{iC} \quad \mathrm{K} = 1, \ldots, \mathrm{M} \quad \mathrm{C} = 1, \ldots, \mathrm{P} \tag{5}$$

IV. The number of domestic trips of an OD-relation calculated on the basis of entropy maximisation must be at least equal to the number of domestic trips of HGVs registered in Austria and reported for this OD-relation (SGVS), with d representing an index for trips of HGVs registered in Austria: $f_{ij} \geq f_{ij}^{d}$

V. The number of trips originating in Austria with a foreign destination which belong to an OD-relation calculated on the basis of entropy maximisation must be at least equal to the number of the same kind of trips of HGVs registered in Austria and reported for this OD-relation in the SGVS: $f_{iC} \geq f_{iC}^{d}$

VI. The number of trips ending in Austria and starting abroad which belong to an OD-relation calculated on the basis of entropy maximisation must be at least equal to the number of the same kind of trips of HGVs registered in Austria and reported for this OD-relation in the SGVS: $f_{Cj} \geq f_{Cj}^{d}$

VII. The sum of all trips ($Z_L$) ending in a NUTS-3-region in Austria and recorded in the ESGVS must be equal to the number of all trips calculated for Austria plus the sum of all trips originating in all countries considered:

$$Z_L = \sum_{i=1}^{N} \sum_{j=1}^{N_L} f_{ij} + \sum_{C=2}^{P} \sum_{j=1}^{N_L} f_{Cj} = \sum_{C=1}^{P} F_{CL} \quad L = 1, \dots, M \tag{6}$$

In analogy, the same applies for destination traffic as well

$$Q_K = \sum_{i=1}^{N_K} \sum_{j=1}^{N} f_{ij} + \sum_{i=1}^{N_K} \sum_{C=2}^{P} f_{iC} = \sum_{C=1}^{P} F_{KC} \quad K = 1, \dots, M \tag{7}$$

As political districts in Austria are significantly smaller in size compared to NUTS-3-regions, each district can be allocated to a particular NUTS-3-region. By using the iteration procedure to maximise the entropy the number of trips of all unknown (empty) OD-relations on the level of political districts in Austria are calculated first. Thus, the target function

$$\sum_{ij} f_{ij} \ln f_{ij} \tag{8}$$

comprises trips within Austria only. Thus it is assumed, that the initial values for all HGVs are the same as for the HGVs registered in Austria (as reported in the special analysis of the national statistics).

$$f_{i_K j_L}^{(0)} = F_{KL} \frac{f_{ij}^{d}}{\sum_{i=1}^{N_K} \sum_{j=1}^{N_L} f_{ij}^{d}} \tag{9}$$

| Notation | Description |
|---|---|
| $f_{i_K j_L}^{(0)}$ | Initial value of trips of HGVs registered in Austria from political district i (belonging to NUTS-3-region K) and j (belonging to NUTS-3-region L) |
| $f_{ij}^{d}$ | All trips of HGVs registered in Austria from political district i to j in Austria [number of trips / year] |
| $F_{KL}$ | All HGV trips between NUTS-3-regions K and L in Austria [number of trips / year] |

The target function for the calculation of traffic from abroad (in analogy, the same applies for destination traffic as well) is defined as:

$$\sum_{Cj} f_{Cj} \ln f_{Cj} \tag{10}$$

The unknown OD-relation $f_{Cj}$ defines trips from the country C to the political district j in Austria under the condition

$$F_{CL} = \sum_{j=1}^{N_L} f_{Cj} \quad L = 1, \dots, M \quad C = 2, \dots, P \tag{11}$$

$$f_{Cj} \geq f_{Cj}^{d}$$

This procedure results in a matrix of domestic-, origin-, destination-, and transit-HGV traffic of Austria for the year 2009 and all recorded HGVs. Spatial references relate to the political districts in Austria – which are significantly smaller than NUTS-3-regions – and NUTS-3-levels for regions abroad.

Subsequently, the data are further disaggregated to the corresponding traffic zones of the Austrian transport demand model. These traffic zones are Austrian municipalities and even smaller units in densely populated areas, whereas in other European countries the size of traffic zones increases with the distance from Austria (from NUTS-3-level to NUTS-0). The matrix of the Austrian transport model for the year 2005 calculated on the basis of an economic model provides the basis for this disaggregation (TRAFICO et al. 2009). The share of transport flows between municipalities in the total flow within one cell was identified for each pair of political districts in this 2005-matrix. This relation was transferred to the preliminary step-1-matrix in order to break down the traffic flows from political districts to municipalities.

### *4.2. Step-2: Transformation of the annual HGV OD-matrix into a matrix of the average daily traffic (workday)*

The matrix resulting from step-1 represents the total road freight transport volume of 2009. However, annual matrices cannot be processed since the Austrian transport models' constraint function and capacities of its road network are based on hourly and daily traffic. Thus, the matrix has to be transformed into a matrix of the average daily HGV traffic of a workday. The transformation factor needed is derived from data of toll gantries on motorways and express roads. These road toll data provide the percentage of HGV traffic of the total transport volume on workdays. Since the data are restricted to motorways and express roads, they are finally crosschecked by using data of the permanently counting stations.

The following equation provides the factor described above:

$$\boldsymbol{T} = \frac{\sum_{\boldsymbol{m=1}}^{\boldsymbol{M}}(\boldsymbol{V}_{m,a} - \boldsymbol{V}_{m,f})}{\sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} \boldsymbol{V}_{m,a}} \tag{12}$$

| Notation | Description |
|---|---|
| $T$ | Transformation factor used to transform the HGV matrix of a year into a matrix of the average daily workday |
| $V_{m,a}$ | Annual HGV traffic at toll bridge m, all days of the year 2009 [number of vehicles / year] |
| $V_{m,f}$ | Annual HGV traffic at toll bridge m, all Saturdays, Sundays, holidays of the year 2009 [number of vehicles / year] |
| $M$ | Number of toll gantries |

In 2009, the sum of workday HGV transits through toll gantries was 513.436.453. These data are based on the 886 toll gantries. For the respective year, the total number of HGV traffic is 556.058.379. The first figure corresponds to a proportion of 0.9233 of the second figure. This factor differs, depending on the registration country of the vehicle. However, since this information is not considered in the transport model, this average factor is used to convert annual values into daily HGV traffic for average workdays. Accordingly, the 2009 matrix of HGV traffic of step-1 is multiplied by the respective factor and divided by 251 which was the number of workdays in 2009:

$$F_{ij,w(2009)} = \frac{F_{ij,a} \times T}{W_{2009}} \tag{13}$$

| | |
|---|---|
| $F_{ij,w(2009)}$ | Number of all trips of the annual average daily HGV traffic from traffic zone i to zone j in the reference year 2009 [number of trips / workday] |
| $F_{ij,a}$ | Sum of all trips of the annual HGV traffic from traffic zone i to zone j [number of trips / year] |
| $T$ | Transformation factor used to transform the annual HGV matrix into a matrix of the average daily workday |
| $W_{2009}$ | Number of workdays in the year 2009 [workdays] |

### 4.3. Step-3: Consideration of cross Alpine freight data (CAFT)

On-street data collection for the CAFT-survey on trans-alpine and border-crossing freight transport is conducted at 18 survey points on different reference days. The objective of this survey is to get accurate data of important OD-relations across the Alps and national borders. The number of HGV trips from origin i to destination j is the sum of all trips between traffic zones i and j of all possible routes passing the CAFT survey point z:

$$q_{ijz} = \sum_r v_{ijrz} \tag{14}$$

| Notation | Description |
|---|---|
| $v_{ijrz}$ | Number of trips from origin i to destination j passing CAFT survey point z choosing route r; index r represents routes with at least one CAFT survey point calculated by the Austrian Transport Model [HGVs / year] |
| $q_{ijz}$ | Total number of trips from origin i to destination j passing CAFT survey point z independently from the chosen route r [HGVs / year] |

#### 4.3.1. Step-3a: Taking multiple counting into account

Trips passing along more than one survey point have a higher chance to get captured in the CAFT than others. In the calculation of the OD-matrix, this has to be considered as multiple surveyed trips. An example is illustrated in figure 1. Route 2 and 3 (R2, R3) are passing one CAFT survey point, whereas route 1 (R1) is passing two CAFT survey points (A and B). A HGV using route 1 can be surveyed at both survey points. That is why the number of trips between traffic zones i and j is overrepresented.

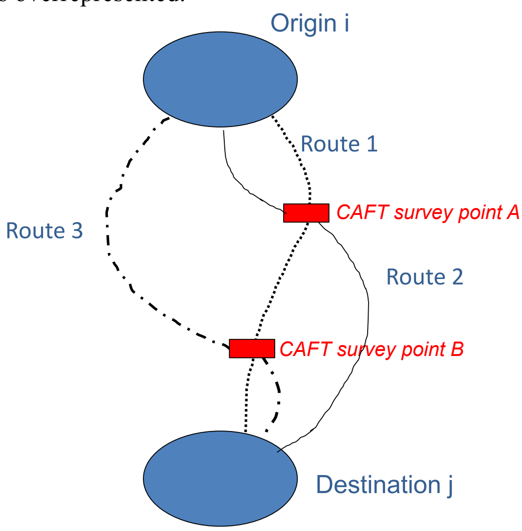


Fig. 1. Route analysis to take multiple surveying at CAFT survey points into account

The term $q_{ijz}^{C}$ represents the number of trips from origin i to destination j counted at CAFT survey point z (including multiple counting due to other survey points passed). Multiple counting is identified by means of a network assignment of the matrix of step-2 using an equilibrium approach resulting in the adjusted number of trips $\hat{q}_{ijz}^{C}$. This results in a factor $\beta_{ijz}$ allowing for the adaptation of HGV traffic volumes of each affected OD-relation.

$$\beta_{ijz} = \frac{\hat{q}_{ijz}^{C}}{q_{ijz}^{C}} \quad (15)$$

| Notation | Description |
|---|---|
| $\beta_{ijz}$ | Factor for the correction of the HGV traffic volumes of each affected OD-relation dependent on the number of passed survey points |
| $q_{ijz}^{C}$ | Number of trips counted at CAFT survey point z including multiple counting due to more than one CAFT survey point passed [HGVs / year] |
| $\hat{q}_{ijz}^{C}$ | Adjusted number of trips at CAFT survey point z (excluding multiple counting due to more than one CAFT survey point passed) [HGVs / year] |

*4.3.2. Step-3b: Consideration of alternative routes*

HGVs can choose more than one route to reach destination j when starting in i. In most cases, none of them will pass a CAFT survey point, while in few cases of trans-alpine and cross-border relations all routes will pass one survey point. In other cases, some, but not all routes pass one of these locations. This has to be taken into account when developing the matrix of step-3. An example is given in figure 2. The route search algorithm of the Austrian Transport Model was applied to calculate all possible routes of all OD-relations passing the Brenner Alpine pass. Fig. 2 shows all routes avoiding the Brenner of those OD-relations of which at least one route passes the Brenner. Due to the structure of the road network in the Alpine region, these are mainly long distance detours through Switzerland (neighbouring country in the west of Austria). If one considers CAFT counting values only, the total number of trips between origins and destinations passing the Brenner is underestimated.

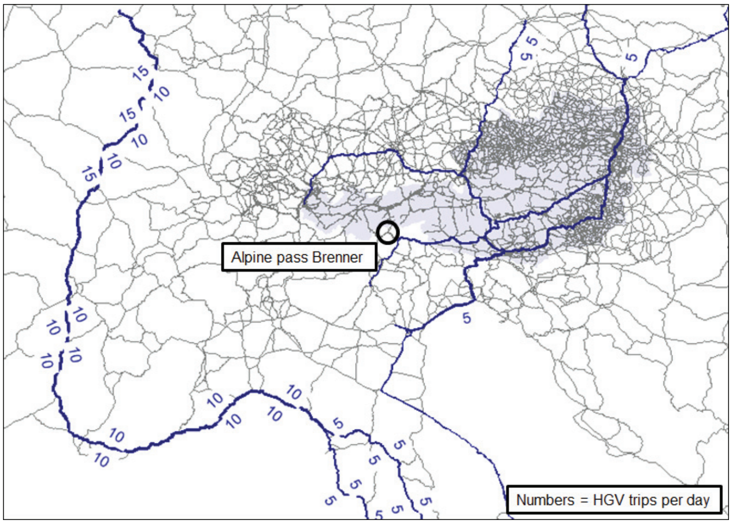


Fig. 2. Alternative routes of HGV traffic to the use of the Brenner Alpine pass (based on an assignment of the matrix of step-2)

Thus, all OD-relations of which some, but not all, possible routes pass CAFT survey points have to be identified. This is again done by an equilibrium assignment of the matrix of step-2 using the Austrian Transport Model. The proportion of all trips between traffic zones i and j passing a CAFT survey point, of all trips between i and j, is expressed in the following equation:

$$\alpha_{ij} = \frac{\sum_z \hat{q}_{ijz}^C}{f_{ij}} \quad (16)$$

The recorded CAFT result of an OD- relation needs to be divided by the resulting factor:

$$\hat{Q}_{ij}^C = \frac{\hat{q}_{ij}^C}{\alpha_{ij}} \quad (17)$$

| Notation | Description |
|---|---|
| $\alpha_{ij}$ | Factor for considering alternative routes |
| $f_{ij}$ | Number of all trips between origin i and destination j according to the European freight transport statistics |
| $\hat{Q}_{ij}^C$ | Number of trips between origin i and destination j including alternative routes [HGVs / year] |

*4.3.3. Step-3c: Merging of the step-2-matrix and the CAFT-data*

The following approach was chosen to merge the step-2-matrix with the CAFT survey data: For all OD-relations, which pass at least one CAFT-survey point, the values resulting from step-3 are used to substitute the corresponding values of the matrix of step-2 (even if the value of step-3 is zero). This procedure is based on the fact that the CAFT result is significantly more accurate than the database of step 1, because the CAFT survey has no non-respond-problems („item-non-response“) as does the ESGVS data set. For the remaining OD relations the value of the step-2-matrix is used. There are no detailed research results to determine the veracity of this hypothesis, but some evidence shows its plausibility. In particular, the modelled traffic volume as a result of the assignment of the matrix of step-2 in comparison to data from counting stations or road toll gantries shows an underestimated value of the official statistics. In order to quantify this effect, the quotient between the results of step-2 and step-3 is calculated, considering OD relations included in both data sets only. Results show that in most cases this quotient is less than 15 %, except for transit traffic (50 %), i.e. an underestimation of HGV trips in the data of the national and European statistics is quantified (Table 1).

Table 1. Comparison of the number of HGV trips between step-2 and 3 based on the OD relations which can be found in both data sets; domestic, destination, origin and transit HGV trips are related to the area of Austria

| | Number of OD relations | Step-2: HGV trips per year (2009) | Step-3: HGV trips per year (2009) | Proportion: Step-2 / Step-3 |
|---|---|---|---|---|
| Domestic traffic | 1,907 | 71,955 | 979,663 | 7.3 % |
| Destination traffic | 2,503 | 182,510 | 1,760,444 | 10.4 % |
| Origin traffic | 2,604 | 214,104 | 1.707.386 | 12.5 % |
| Transit traffic | 1,000 | 495,568 | 988,809 | 50.1 % |

A comparison between the values of the matrix resulting from step-2 and the matrix calculated in step-3 shows that almost one million HGV trips per year were added (+4%). Transit traffic through Austria has to be adjusted the most (+37%). Almost 800,000 HGV trips had to be discarded. In this step it comes to only marginal changes of Austrian domestic traffic. This is plausible due to the location of the CAFT survey points at Alpine passes and cross border stations of Austria covering mainly transit traffic.

*4.4. Step-4: Matrix correction with the help of data of toll gantries and traffic counting stations*

Step-4 is a matrix correction using data from traffic counting stations and toll gantries. The matrix correction method follows the principle of “fuzzy logic”. The matrix correction iteratively minimizes the difference between the modelled traffic volume and the traffic counts by assigning a given matrix, comparing modelled and counted values, adapting the matrix and repeating this until a final condition is reached (Fig. 3).

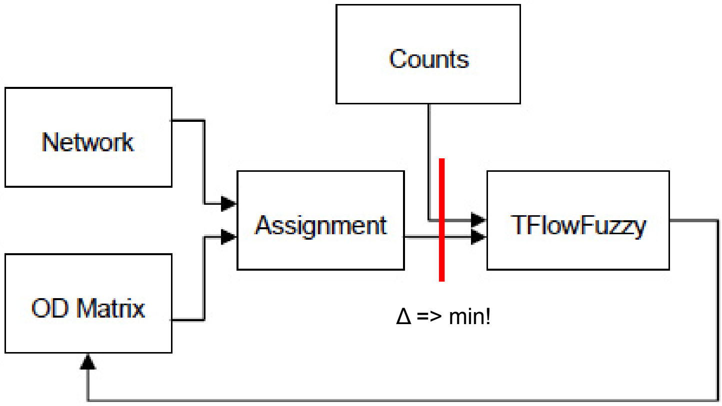


Fig. 3. Principle of the VStromFuzzy software tool (source: PTV AG 2012)

The step-3-matrix was used as starting matrix, the software program "VStromFuzzy" from the transport model software VISUM of the German PTV AG (PTV AG 2012) was applied for the procedure. 1,967 vehicle counting stations were used for the procedure. A maximum deviation of 10% of counted from modelled values was permitted.

Admittedly, the PTV software is not specified to address this kind of problem. However, the client requested the use of software tools available on the market. This fact led to problems which had to be overcome while the project was running. A large number of counting points is available, and they are concentrated along the higher ranked road network. This can lead to mathematical redundancy, so that the algorithm might not fit in a satisfying way. Therefore the number of counting points had to be adapted to achieve an acceptable result. An adjustment of the software tool according to the needs of this methodology could improve the result.

### *4.5. Plausibility checks of the results of the individual steps*

Matrices resulting from steps 2, 3, and 4 were assigned using the Austrian Transport Model. The accuracy of the matrices can be assessed by a comparison of the modelled values with counted values from a representative selection of vehicle counting stations. However, this quality control includes inaccuracies resulting from the assignment algorithm of the transport model used. For this particular comparison 1,201 counting stations were used, of which 864 were toll gantries of motorways and express roads, and 337 automated counting stations at minor roads. The result of step-4 depicts an average absolute deviation of the median modelled traffic volume from the counted number of vehicles of merely 17 HGVs per 24h. The median of the deviation drops from -30% in step-2 to -20% (step-3). In step-4, the median deviation modelled is less than one percent (-0.6%). 94% of all values are within a maximum deviation of 10%. With the exception of singular outliers, the modelled values correspond very well to the independent figures collected at the vehicle counting stations (Fig. 4).

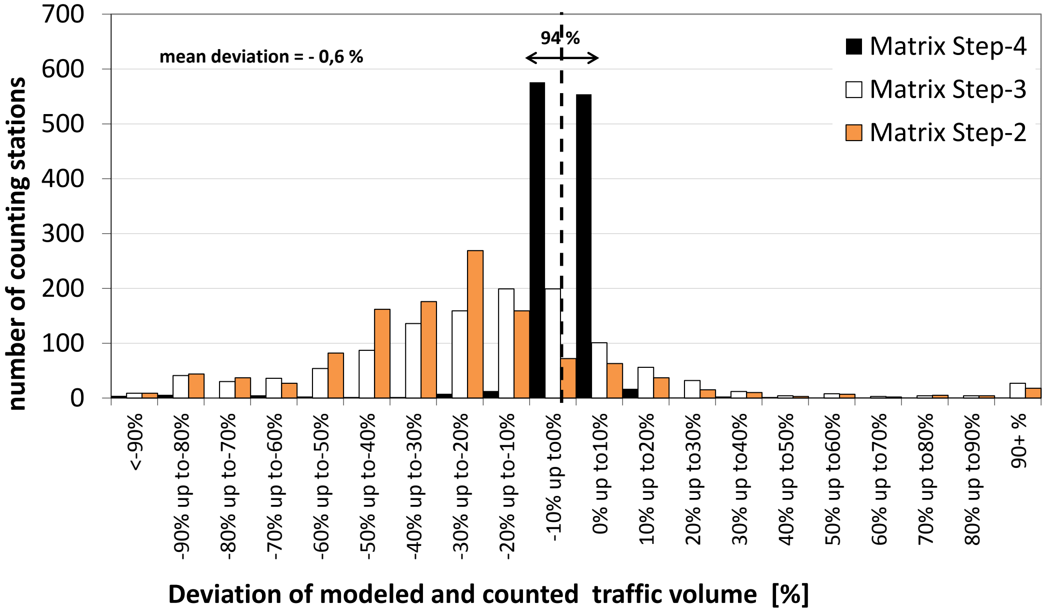


Fig. 4. Distribution of the relative difference of modelled and counted values at counting points of the road network for the matrices 2 to 4 for freight traffic

The percentage root mean square error (PRMSE) compares the fit of two data series and can be used as additional plausibility check. According to Sammer (2006), the PRMSE is calculated as

$$PRMSE\ [\%] = 100 \times \left[\frac{1}{n} \times \sum \left(\frac{V_{1i} - V_{2i}}{V_{2i}}\right)^2\right]^{0,5} \tag{18}$$

| Notation | Description |
|---|---|
| $n$ | Number of counting stations |
| $V_{1i}$ | Modelled transport volume at counting station i |
| $V_{2i}$ | Counted transport volume at counting station i |

If 1,201 counting stations on minor roads as well as toll gantries on motorways are taken into account, the PRMSE decreases from 42% for the step-2-matrix to 12% of the final matrix. The same tendency can be shown for a separated analysis of the 864 toll gantries on motorways with a decrease from 35% to 6% and of 559 counting stations in the entire road network (including those 222 on motorways) from 47% to 16%. An increase of the PRMSE between the second and third step can only be observed for the 559 counting stations, for which the PRMSE-value worsened by one percentage point which is not that surprising since the CAFT-integration aims at an improvement of the mapping of international transports. These results show that the quality of the results depends on the number and quality of counting stations which is significantly better in the motorway network.

## 5. Application of the method

The resulting matrices were integrated into the Austrian Transport Model and assigned with the help of the transport model software VISUM of PTV AG using an all-or-nothing choice algorithm. This means that all drivers chose the route with the lowest impedance value, congestions or capacity restraints are not reconsidered; it is assumed that all trips between i and j are made using the same route. This algorithm was chosen since it reflects the situation of HGV traffic quite well and is used in many HGV transport models. Underlying data of motorized passenger road traffic are derived from the Austrian Transport Model.

### *5.1. HGV traffic volume 2009*

According to the matrix resulting from step-4, the average workday traffic in Austria amounts to 260,630 vehicle trips per day, which is the equivalent of 70.85 million HGV trips per year. Compared to the matrix of step-2, this shows an increase of 44% which corresponds to an underestimation of -31% by the official statistic (-31% = [100/144-1.0] x 100). Differences between the second and third as well as the second and forth step vary, depending on the origins and destinations of trips. The major increase of the origin, destination, and domestic traffic related to Austria results from step-4, whereas there is only little change in transit traffic in step-4. On the other hand, the transit traffic volume increases especially in step-3 because CAFT-survey stations are taken into account; they offer a valid and reliable reflection of international HGV trips since due to their location, mainly transit but also origin and destination traffic is covered (Table 2).

Table 2. Average number of HGV trips per workday of the matrices 2 to 4 for the reference year 2009 and the relative changes; the domestic, destination, origin, and transit HGV trips are related to the area of Austria

| | Average number of HGV trips per workday [HGV / 24h (workday] | | | Relative changes [%] between different steps | |
|---|---|---|---|---|---|
| | Step-2 | Step-3 | Step-4 | Step-2-3 | Step-2-4 |
| Transit traffic | 9,230 | 12,660 | 12,900 | 37% | 40% |
| Domestic traffic | 150,360 | 150,360 | 214,610 | 0% | 43% |
| Origin and destination traffic | 21,560 | 25,600 | 33,120 | 19% | 54% |
| Total traffic volume | 181,150 | 188,610 | 260,630 | 4% | 44% |

### *5.2. HGV kilometrage 2009*

The average daily driving kilometrage per workday of HGVs on the Austria road network rises between step-2 and 4 from 11.64 to 15.69 m. kilometres per workday. This is an increase by 35% and indicates an underestimation of -26% in the original survey (-26% = [100/135-1.0] x 100); the biggest change in the process happens in step-4. This change is slightly smaller than that of the HGV traffic volume (carried tons of goods within 24 h by HGV trips/ workday). This is mainly due to a decrease in the average distance travelled per HGV trip, at every step of the methodology. In this context, one can see a discrepancy between steps 3 and 4. Due to the integration of CAFT survey data at step-3, the HGV traffic performance rises more steeply than the actual number of trips. This is plausible, since mainly trips of origin-, destination- and transit traffic related to Austria are concerned. On the other hand, in step-4, mainly domestic traffic is concerned as these trips are on average shorter. This also indicates a significant underestimation of domestic traffic within the ESGVS data (Table 3).

Table 3. Average daily HGV kilometrage of the matrices 2 to 4 on the Austrian road network (absolute figures and relative changes); the domestic, destination, origin, and transit HGV trips are related to the road network in Austria

| | Average daily HGV kilometrage per workday [m. HGV-km/24 h (workday)] | | | Relative changes [%] | |
|---|---|---|---|---|---|
| | Step-2 | Step-3 | Step-4 | Step-2-3 | Step-2-4 |
| Transit traffic | 2.09 | 2.78 | 2.88 | 33% | 38 % |
| Domestic traffic | 7.10 | 7.05 | 9.41 | -1% | 32 % |
| Origin and destination traffic | 1.22 | 1.55 | 1.78 | 29% | 45 % |
| Total traffic volume | 11.64 | 13.00 | 15.69 | 12% | 35 % |

### *5.3. HGV transport volume 2009*

The transport volume is the entire amount of goods transported. It is measured in carried tons of goods per 24 h. This transport volume depends on the number of HGV trips and their loading. The average daily HGV transport volume without transit traffic on the Austria road network rises between step-2 to 4 from 1.39 to 1.96 m. t/24 h per workday. This amounts to an increase of 41% and an underestimation of -29% by the original matrix (-29% = [100/141-1.0] x 100).

It has to be noted that during the first two steps there is no differentiation between “external” traffic and transit traffic, hence these data are not depicted in the table. “External” traffic is defined as traffic which is considered as transit according to European statistics, but is routed on roads circumventing Austria by the Austrian Transport Model due to the route search algorithm. The reason for this issue could be either incorrect official mapping by the national statistical bureaus or it could be an inappropriate resolution used by the route search algorithm. Between steps 2 and 3 there is a slightly smaller increase in transport volume neither originating in, nor destined to Austria. The same applies to combined transit- and foreign traffic. This seems obvious, since at least at Alpine passes and border counting stations, origin-, destination- and transit-traffic can be recorded with high probability. Marked changes can be found especially during steps 3 and 4 (Table 4).

Table 4: Average daily HGV transport volume of matrices 2 to 4 per workday in the Austria road network (absolute figures and relative changes); the domestic, destination, origin, and transit HGV trips are related to the area of Austria

| | Average daily HGV transport volume per workday [m. t/24 h (workday)] | | | Relative changes [%] | |
|---|---|---|---|---|---|
| | Step-2 | Step-3 | Step-4 | Step-2-3 | Step-2-4 |
| Domestic traffic | 1.13 | 1.13 | 1.59 | 0% | 41% |
| Origin and destination traffic | 0.26 | 0.29 | 0.37 | 12% | 42% |
| Total transport volume (without foreign traffic) | 1.39 | 1.42 | 1.96 | 2% | 41% |
| Transit- and foreign traffic | 0.16 | 0.21 | | | |
| Transit traffic | | | 0.18 | | |
| Total transport volume | 1.56 | 1.63 | 2.14 | 4% | 37% |

## 6. Conclusions

The integration of independent data sources significantly improves the quality of the official national freight transport statistics which are based on a community flow survey. This refers to both the integration of data of roadside OD-interviews at Alpine passes and border crossing stations and data of the electronic toll services for

HGV on motorways and of the automatic counting stations. Both lead to a significant improvement of data quality, whereas the CAFT is particularly important for correcting international traffic flows starting in, ending in, or passing through Austria. Data of counting stations or toll gantries are needed to adapt domestic traffic.

A plausibility check was made by comparing data from traffic counts with modelled values resulting from an assignment of the final matrix. The results show a considerable match of the counted and modelled values. The fit of modelled and counted data is higher for the motorway network; the average difference there is only eight vehicles (0.3 %). A comparison of the modelled values with data from counting stations in the entire road network shows a difference of 14 trucks on average (1%). Obviously, the quality of results depends on the quality and amount of counting stations which is higher on motorways.

The changes in the matrix sum underpin the thesis that the official national and European road freight transport statistics under-report road transport. This under-reporting accounts for -31% of HGV traffic volume, -26% of HGV kilometrage and -29% of HGV transport volume on the Austrian road network in the year 2009. Domestic traffic accounts for 82% of HGV traffic volume, 60% of HGV kilometrage travelled on the Austrian road network and 81% of HGV transport volume.

The methodology presented offers significant improvements of official statistics based on the commodity flow survey at comparable low costs and efforts. In a further step, it should be analysed if the procedure also produces satisfying results if the integration of the CAFT-data is dropped. If this were possible, the methodology could also be applied in other countries. The results of step-4 and also the plausibility checks depend on the impedance function used for the network assignment. The function used for the Austrian transport model is a fairly simple one; a more advanced impedance function might further improve the OD-matrix.

## 7. Acknowledgements

The authors gratefully acknowledge the support of the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit) which funded the project IMoVe-Gueter within the national funding framework program IV2Splus and the Motorway and Expressway Financing Limited Company of Austria (Asfinag).

## 8. References

Asfinag (Autobahn- und Schnellstrassen-Finanzierungs-Aktiengesellschaft; Motorway and Expressway-Financing Limited Company), 2012. Special toll for vehicles with a maximum permissible gross weight of up to 3.5t. http://www.asfinag.at/maut/go-box-kraftfahrzeuge-ueber-3-5t-hzg. (date of access: 5 July 2012).

Combes F., Leurent F., 2007. Advances in freight transport demand modelling: An assessment with research perspectives. Proceedings of the European Transport Conference 2007. Noordwijkerhout (The Netherlands).

de Jong G., Vierth I., Tavasszy L., Ben-Akiva M., 2012. Recent developments in national and international freight transport models within Europe. In: Transportation. Planning - Policy - Research – Practice. Editor: Springer, ISSN: 0049-4488, Heidelberg (Germany).

DTV-Verkehrsconsult GmbH, 2011. Auswertung und Darstellung der Ergebnisse der automatischen Straßenverkehrszählung 2009 (Analysis and description of the results of the automatic traffic counting stations 2009). Aachen (Germany). In German.

European Commission, 2014. A European classification system of spatial areas of reference. http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Nomenclature_of_territorial_units_for_statistics_%28NUTS%29. (date of access: 30 September 2014).

Eurostat, 2011. Road freight transport methodology. Editor: Publications Office of the European Union, Luxembourg.

Grafl, W., Neumann, A., Sammer, G.; Bittner, L., Fuerst, E., Rudlof, M., 2007. Inter-Dat – Verfügbarkeit von Verkehrsdaten von öffentlichem und gewerblichem Interesse, Schlussbericht (research project Inter-Dat – Availability of transport related data in the public and commercial interest, Final Report). Commissioned by the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit) within the framework I2 of the Austrian Research Promotion Agency (FFG). Vienna (Austria). In German.

Herry Consult, Snizek,& Co, Intra Performance®, 2011. Alpenquerender Güterverkehr - Österreich 2009 (Cross-Alpine freight transport 2009 – Austria). Forschungsarbeiten aus dem Verkehrswesen (Research Works in Transportation). Volume 147. Commissioned by the Austrian Federal Ministry of Transport, Innovation and Technology (bmvit). Vienna (Austria). In German.

Highways Agency, Department for Transport, UK, 1997. Design Manual for Roads and Bridges (DMRB). Volume 12, Section2, London, (UK).

IMoVe-Gueter Consortium, 2014. Innovative Modellierung zur Verbesserung der Grundlage der Güterverkehrsstatistik für Österreich (Innovative model concept to improve the Austrian transport statistics). Research project commissioned by the Austrian Federal Ministry of Transport, Innovation and Technology (bmvit) funded within the framework programme IV2plus of the Austrian Research Promotion Agency (FFG). Vienna (Austria). In German.

Karner, T., Rudolf, M., Scharl, S., 2012. Österreichische Verkehrsstatistik 2010 (Austrian Transport Statistics 2010). Editor: Statistics Austria. Vienna (Austria). In German.

King, G.; Rosen O.; Tanner M. A. (2004): Ecological Inference, Cambridge University Press

Kockelmann, K., Browne, M., Leonardi, J. 2009. The acquisition of long-distance freight data: synthesis of a workshop. In: Bonnel, P., Lee-Gosselin, M., Zmud, J., Madre, J. (Ed.). Transport survey methods: Keeping up with a changing world. Bringley. pp. 311.

Krieger, D., 2011. A Synthesis of Highway Practice. Freight Transportation Surveys. NCHRP. Washington (USA).

McKinnon, A. C., Leonardi, J. 2009. The collection of long-distance road freight data in Europe. In: Bonnel, P., Lee-Gosselin, M., Zmud, J., Madre, J. (Ed.). Transport survey methods: Keeping up with a changing world. Bringley. pp. 295.

Park, M., Hahn, J., 2014. Regional Freight Demand Estimation Using A Commodity Flow Survey In Korea. Poster Presentation at the 10th International Conference on Transport Survey Methods. Leura (Australia).

Pischinger, R., Hausberger, S., Sudy, C., Sammer, G., Thaller, O., Schneider, F., Stiglbauer, A., 1997. Volkswirtschaftliche Kosten-Wirksamkeitsanalyse von Maßnahmen zur Reduktion der CO2-Emissionen des Verkehrs in Österreich (Economical cost-effectiveness-analysis of measures appropriate to reduce the CO2 emissions of transport in Austria). Commissioned by the Austrian Federal Ministry of Economy, Family and Youth, Series of the Institute for Internal Combustion Engines and Thermodynamics. Graz University of Technology. Graz (Austria). In German.

PTV AG, 2012. PTV VISUM 12.5. User Manual. Karlsruhe (Germany).

Sammer, G. 2006. Processing, Analysis, and Archiving of Travel Survey Data. In: Stopher, P., Stecher, C. (Ed.). Travel Survey Methods: Quality and Future Directions. Pergamon, Elsevier, Oxford (UK).

Sammer, G., Raser, E., Roider, O., 2008. Aktuelle Fragen zum Fahrverhalten auf dem Hochrangigen Straßennetz (Analysis of the current transport user's behaviour on the high ranked road network). Resarch project commissioned by the Austrian Motorway and Expressway-Financing Limited Company (Asfinag). Forschungsbericht (Research Report) 08/2007 of the Institute for Transport Studies at the University of Natural Resources and Life Sciences Vienna (Austria). In German.

Sammer, G., Klementschitz, R., Steininger, K., Schmidt, Ch., Hausberger, St., Rexeis, M., 2009. Oe-Güter, Problemanalyse und Lösungskonzepte für den Güterverkehr in Österreich aus der Sicht der Bundesländer – Stufe 1 (Problem analysis and solution concepts for freight transport in Austria from the viewpoint of Austrian provinces – step-1). Commissioned by the Province of Upper Austria in co-operation with the governments of the Austrian Provinces. Vienna, Graz (Austria). In German (unpublished).

Sammer, G., Roeschel, G., Gruber, Chr. 2012. Qualitätssicherung für die Anwendung von Verkehrsnachfragemodellen und Verkehrsprognosen (Quality Management and Validation for Application of Transport Demand Modelling and Forecasts). Schriftenreihe Straßenforschung des Bundesministeriums für Verkehr, Innovation und Technologie. Heft 604. Vertrieb: http://www.fsv.at, Vienna (Austria).

Spiegel, T., 2003. The European road freight transport according to the regulation (EU) 1172/98 in comparison to the Cross-Alpine freight transport survey. Document Road/2003/12/DE in Road Transport Statistics. Vienna (Austria), Brussels (Belgium).

Spiegel, T., 2007. Gesamtaufkommen des Straßengüterverkehrs in Österreich (Total volumes of road freight transport in Austria). In: Austrian Transport Statistics 2005. Vienna (Austria). In German.

Southworth, F., Davidson, D., Hwang, H., Peterson, B., Chin, S, 2010. The Freight Analysis Framework. Version 3: Overview of the FAF3 National Freight Flow Tables. Federal Highway Administration Office of Freight Management & Operations. Washington (USA).

Tavasszy, L., Ruijgrok, K. I., 2010. Incorporating logistics in freight transportation models. State of the art and research opportunities. Proceedings of the 12th World Conference on Transport Research. Lisbon (Portugal).

Tavasszy, L., Ruijgrok, K., Davydenko, I., 2012. Incorparating logistics in freight transportation models: state of the art and research opportunities. Transport Reviews: a transnational transdisciplinary journal. Transport Reviews II 2012. Delft, Tilburg (The Netherlands)

Trafico, Ivwl, Eth, Panmobile, Joanneum Research, Wifo, 2009. Verkehrsprognose Österreich 2025+, Endbericht (Transport forecast Austria 2025+, Final Report). Commissioned by the Austrian Federal Ministry of Transport, Innovation and Technology (bmvit). Vienna (Austria). In German.

U.S. Census Bureau, 2011. Commodity Flow Survey. www.census.gov/econ/cfs/how_data_collected.html. (date of access: 27 July 2011).

U.S. Department of Transportation, 2015. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/commodity_flow_survey/index.html. (date of access: 30 July 2015).

Wermuth, M., Neef, C., Steinmeyer, I., 2004. Goods and business traffic in Germany. Resource Paper. In: Stopher, P., Stecher, C. (Ed.). Travel Survey Methods: Quality and Future Directions. Pergamon, Elsevier, Oxford (UK).