Periodica Polytechnica
Transportation Engineering

42(2), pp. 111-117, 2014
DOI:10.3311/PPtr.7465

Reconstructing Surveyed Itineraries and Choices between Inter-City and Regional Train Services

Florian Heinitz / *Erik* Fritzlar

RESEARCH ARTICLE

RECEIVED 3 April 2014

Abstract

This contribution presents findings from a network-based examination of observed route choice behaviour of railway travellers in the German Federal State of Thuringia. The purpose is twofold: On the one hand, the idea was to assure that the VISUM 12 schedule-based network models were capable of ex-post reconstructing the exact itineraries from passenger surveys. When this could be achieved, the second step was to use these RP data to explain the binary choice between the combined usage of the Inter-City and regional network versus the regional network only through econometric modelling. The trade-off between different service level characteristics was of special interest. The O-D specific demand split was incorporated in the State Transport Model to replace a formerly fixed segmentation of regional network captives from choice riders on the entire network.

Keywords

transport demand modelling · route choice · discrete choice · railway mode

Florian Heinitz
Institut Verkehr und Raum, FH Erfurt University of Applied Sciences, Altonaer Strasse 25, D-99085 Erfurt, Germany
e-mail: heinitz@fh-erfurt.de

Erik Fritzlar
Institut Verkehr und Raum, FH Erfurt University of Applied Sciences, Altonaer Strasse 25, D-99085 Erfurt, Germany
e-mail: erik.fritzlar@fh-erfurt.de

1 Introduction

Since the German railway reform of 1994, there has been a clear distinction between the long-distance passenger division offering Inter-City services and the publicly co-funded regional train services. This "dichotomy" became even more pronounced with the differentiation of levels of service followed by the inauguration of further high-speed lines (Garmendia et al., 2011). Moreover, in the course of the Europe-wide railway liberalization process, more and more private (or non-federal) train-operating companies have gained market access to Germany's regional lines. It is therefore of much interest to further develop existing modelling tools to better understand the intra-modal split and match the observed vehicle occupancies with the algorithmiccally assigned paths of passenger flows to the timetable network.

The Integrated Transport Model of Thuringia is a central instrument to provide multi-modal analyses and forecasts for devising state-wide transport policies. It currently comprises about 1,200 model zones within the investigation area, another 300 cordon cells at the Federal State borders and up to about 700 zones across Europe, primarily used for inter-regional transport studies. The heterogeneity of the residents is covered by a segmentation into nine household types, three region types (two within Thuringia), and eight trip purposes. The modal choice set consists of car (driver / passenger), public transport, and non-motorized modes. The network models are very detailed and regularly updated, using a dynamic road network configurator to choose from up to 420 rebuilding/traffic reorganization measures as well as weekday-specific electronic timetable information for most of the railway, bus and tram lines, and their underlying fare systems. The supply side model is implemented in *VISUM 12* software, using its branch and bound techniques for public transit assignment (Friedrich et al., 2007).

The four-step demand computation flow is of the trip-end type (see Ortuzar and Willumsen, 2011), depicted in Fig. 1.

The research rested on the initial assumption that there was a considerable bias in the choice between Inter-City (IC) and regional trains, as suggested by former load pattern analysed

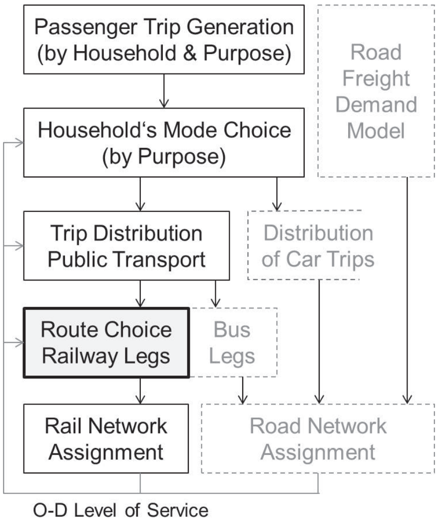

Fig. 1. Railway Branch within the Thuringia State Model (Source: Authors' Representation)

in conjunction with the modal trip matrices. The goal of closer examination was to verify the route search in the schedule-based networks and to understand what factors contributed and to what extent to the travellers' choices between the rail network tiers.

2 Methodology

The investigation started with a literature review on route search on public transportation networks as well as related studies that specifically tackle the discrete choice problem between Inter-City travel and regional trains. A recent overview is provided by Liu et al. (2010). Notable results were achieved e.g. by Vrtic and Axhausen (2003), Xu et al., 2009 or Gao and Wu, 2011. Scherer (2010) examined the preferences of transit riders.

Passenger survey data from two regional railway censuses conducted in the central and eastern (ISUP, 2009) as well as the northern (GVS, 2010) part of Thuringia were used. The coverage is sufficiently good, with 60.2 % of the railway patronage interviewed at the core network, and 48.0% at the supplementary network. The merged disaggregated database comprises 20,858 datasets with personal and trip characteristics, departure and arrival station as well as an itemization of waypoints and train lines used. Up to three transfer stations before and after the observed leg could be recorded with the help of the questionnaires. A simplified formal description of each data set is:

$$\left(i,j,z,\pi,\phi,d,\tau,m_A,m_E,\{r_k,\lambda_k\},\{s_{k'},\lambda_{k'}\}\right) \quad \text{with} \quad -3 \le L^- \le k \le 0 \le k' \le L^+ \le 3 \ . \tag{1}$$

The key to the first set of symbols is given in Table 1.

The first stage of the investigation was the descriptive analysis of the itineraries, followed by a software-assisted matching of itinerary information with the path search results obtained from the timetable network model. This had to be carried out for each data set separately, taking about 56 hours of computation time. The goodness-of-fit will be analysed in Section 3.

Tab. 1. Key to Symbols Used – Part I

Symbol	Variable Name
i	Origin Zone Index
j	Destination Zone Index
z	Trip Purpose
π	Party size
φ	Fare class
d	Day of the week
τ	Time of observation
m_A, m_E	Means of transport for access, egress
r_k, $\overline{r}_k$	Observed, modelled waypoints (i.e. transfer stations) already visited
s_k, $\overline{s}_k$	Observed, modelled waypoints (i.e. transfer stations) still to be visited
λ_k, $\overline{\lambda}_k$	Observed, modelled leg's line code
L^-, L^+	Limits of recorded transfers before/after
k	Leg's index
TT_F, TT_N	Travel time [*min*] for Inter-City (*F*) and Regional train connection (*N*)
UF_F, UF_N	Actual unit fare [*€-cent/Pkm*] for Inter-City (*F*) and Regional train connection (*N*)
NT_F, NT_N	Number of transfers [#] for Inter-City (*F*) and Regional train connection (*N*)
TD_F, TD_N	Travel distance [*km*] for Inter-City (*F*) and Regional train connection (*N*)

Comparisons between the IC and regional routes have revealed that 5,157 or 24.7 per cent of all interviewed passengers had an actual choice between non-identical train routes with or without the inclusion of Inter-City trains, although only a proportion travelled on the IC trains - irrespective of their awareness of the alternative offer. The sub-set of choice riders is used to derive a choice model between routes including intercity trains and routes on the regional network only, presented in Section 4. Since the chosen route ("Inter-City/Regional" or "Regional only") has already been reconstructed in the first step, the service levels of the optimal alternative route ("Regional only" or combined "Inter-City / Regional") were added for each observation. Regional network connections were obtained by disabling Inter-City network links before the route search.

The choice situation to be modelled is depicted in Fig. 2.

Prior to the discrete choice modelling, a descriptive analysis was deployed to detect potentially useful dependencies of the inclination to choose Inter-City trains from personal characteristics, trip characteristics, and differences in service level parameters between the two sub-networks.

The railway itinerary is framed by access and egress links from/to the zone centroid nodes to/from the station used for. The addresses of true origin and destinations remain unknown as the survey only asked for departure/arrival stations. Modal split during access and egress, however, exhibits a less expected frequency distribution. (Fig. 3)

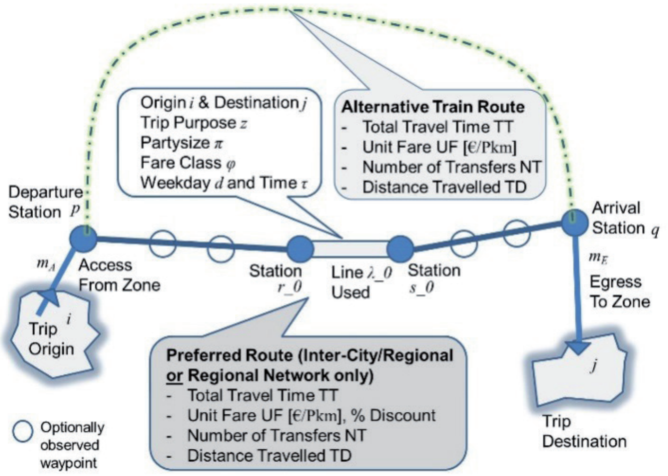

Fig. 2. Choice Situation (Source: Own Representation)

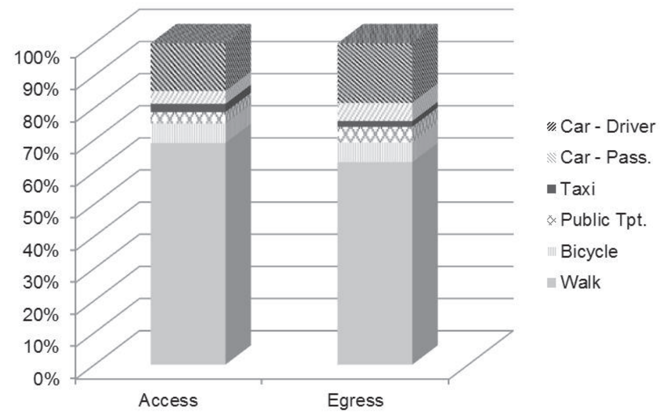

Fig. 3. Mode Choice for Station Access and Egress (Source: Authors' Representation)

Railway stations are mostly accessed and left by pedestrian mode, suggesting that the typical catchment of a train station is small for two thirds of the patronage. The identical modal share of bicycles during access and egress can be explained if the interviewed carried the bicycle with them during the train ride, using it for both access and egress. Although the in-vehicle bike transport was not asked for in the passenger survey, the possibility that bikes are parked at either end of the railway trip has been ruled out. Besides a few IC with designated walk-on bicycle service, the offer to carry bicycles on board is a distinguishing service feature of regional trains in Germany and thus a potential choice criterion.

3 Itinerary Reconstruction Results

For each of the network variants shortest-path searches for even- and odd-hour departures in the morning and afternoon were tested, including the presumptive departure time obtained by a back calculation of the elapsed hours from the time stamp of the interview.

The formal matching criteria to be checked for the line indices and for the agreeing waypoints are:

$$\lambda_k = \bar{\lambda}_k, \forall k, L^- \le k \le L^+ \tag{2}$$

$$\left(r_k = \bar{r}_k, \forall k, L^- \le k \le 0\right) \wedge \left(s_{k'} = \bar{s}_{k'}, \forall k', 0 \le k' \le L^+\right) \tag{3}$$

Based on data compiled by Dömming & Fritzlar, 2013, the accuracies in reconstructing both the used line / mean of transport as well as the waypoints for the railway connections which may involve other public transport services are given in Figure 4 and Figure 5 respectively.

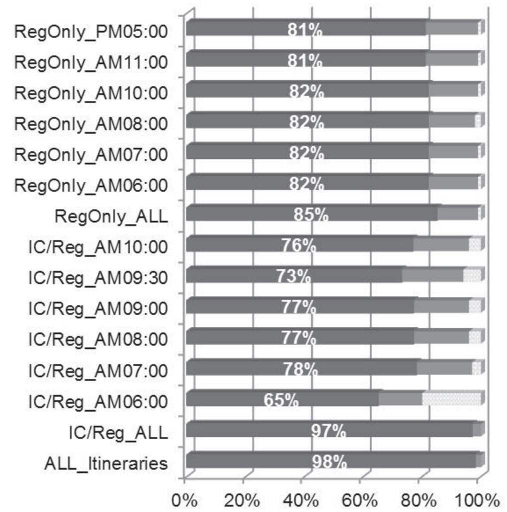

Fig. 4. Itinerary Matching Results – Means of Transport (Source: Authors' Representation)

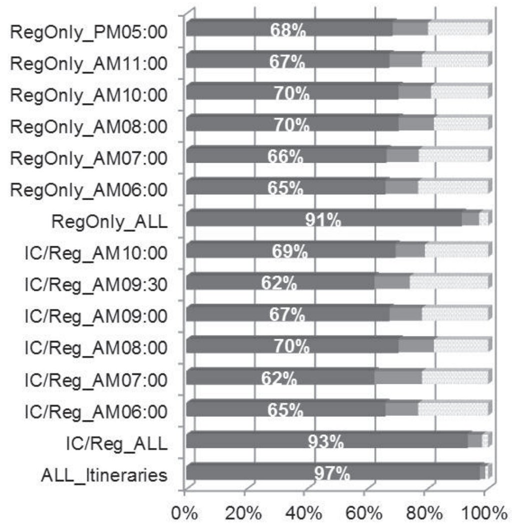

Fig. 5. Itinerary Matching Results – Correct Waypoints (Source: Authors' Representation)

The results show that both network types and a combination of departure times were needed to ultimately reconstruct the observed itineraries. More than nine out of ten connections of either regional or IC connections could be reproduced entirely. Although some problems with the reconstruction of the travellers' departure times - especially in the early morning - remain, the achieved total maximum of >97 per cent of more than 20,000 data sets assures a very good overall conformance to observations, thus a mostly valid basis for the next step of analysis. A residue of about one per cent of observations remains. These exhibit atypical routings which are non-explicable by optimality goals underlying a shortest-path search algorithm. On a cautionary note one must acknowledge that the network model's complexity was just medium-scale, but fully in agreement with the real-world timetable.

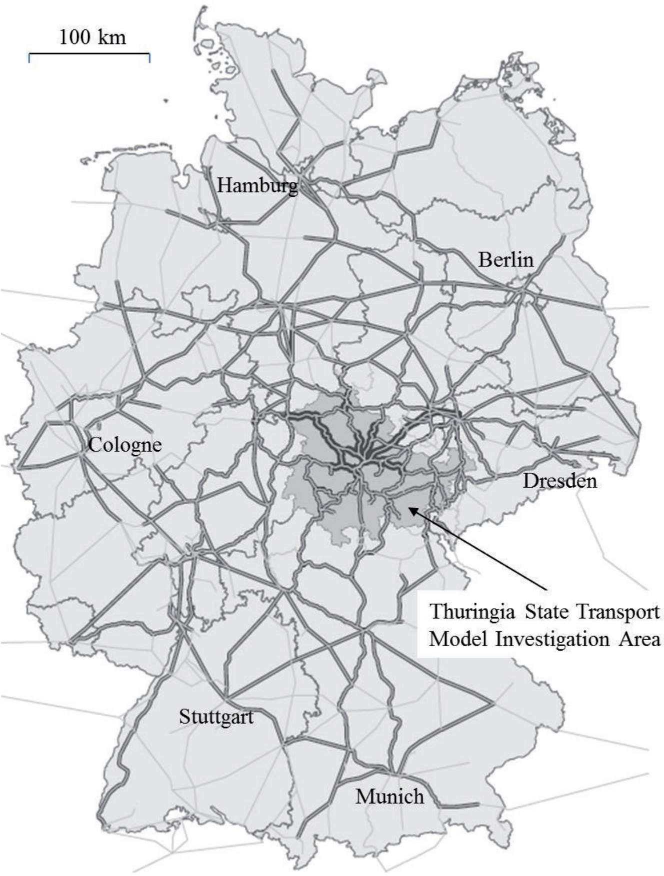

Fig. 6. Nationwide Network Assignment Graph of the Trip Data Set Used (Source: Authors' Representation)

The finally obtained load pattern is depicted in Fig. 6.

4 Choice Modelling Results

First exploratory results are given in Fig. 7, 8, 9 and 10, illustrating distinct dependencies of the preferences of choice riders in favour of an Inter-City train connection from the trip distance, trip purpose, and trip party sizes in terms of average values in the non-weighted subsamples.

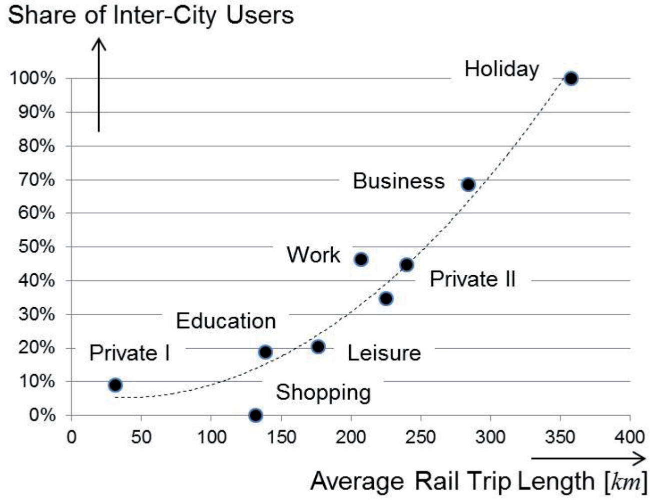

Fig. 7. Usage of the Inter-City Network by Trip Purposes (Source: Own Representation)

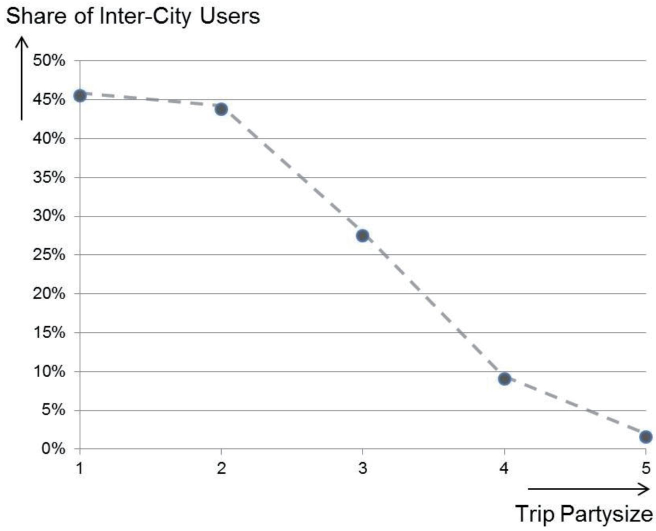

Fig. 8. Usage of the Inter-City Network by Trip Party Size (Source: Authors' Representation)

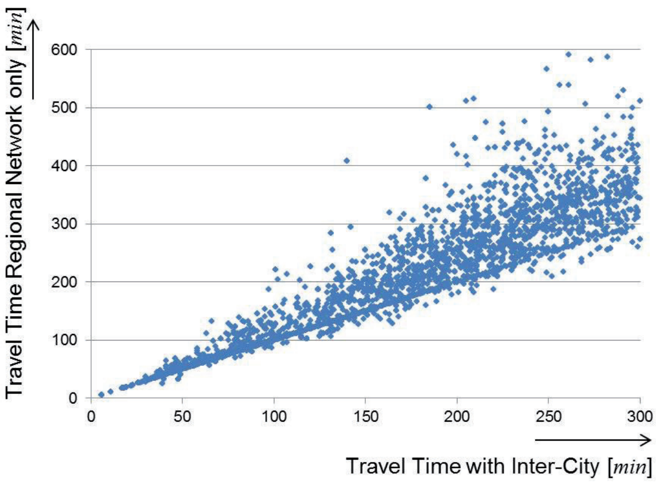

Fig. 9. Entire Network vs. Regional Network Travel Times (Source: Authors' Representation)

For groups of more than two persons except for holiday trips, the share of IC users drops considerably to reach low single-digit values for party sizes of five and bigger. Regional trains cost about 30 per cent less, with 11.92 €-cent per person-kilometre on average rather than 15.48 €-cent for the long-distance trains. This defines the trade-off for the backlog in terms of speed and the necessity of transfers, especially for distant O-D pairs. According to analyses of the sample, regional train users have to accept about 25 per cent more transfers on average.

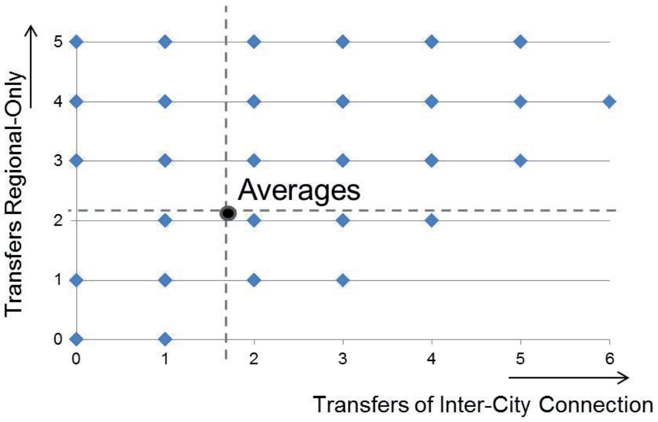

Fig. 10. Entire Network vs. Regional Network Transfers (Source: Authors' Representation)

With the holiday trip purpose left aside, since travellers in this segment have used Inter-City trains without any exception, the remaining set for the model estimation comprises revealed preferences for 3,365 individual trips, amended with alternative-specific service levels from the workflow so far.

With a binary logit model as a special case of a MNL (Ben-Akiva and Lerman, 1985; Hensher and Green, 2005, and Train, 2009) a commonly used model form is deployed. Under the assumptions of independently, identically varying random error terms, the probabilities of individual (or group) n choosing "Inter-City / Regional" (F) versus "Regional trains only" (N) are

$$P_n(F) = \frac{\exp(V_{F,n})}{\exp(V_{F,n}) + \exp(V_{N,n})} = \left(1 + \exp(V_{N,n} - V_{F,n})\right)^{-1} \quad (4)$$

$$P_n(N) = 1 - P_n(F) \quad (5)$$

The finally selected linear specification of the systematic utility functions V of the alternatives is as follows:

$$V_{N,n} = \beta_{NT} \cdot NT_{N,n} + \beta_{PS} \cdot \pi_n + \beta_{WOR} \cdot \delta(z,1)_n + \beta_{BUS} \cdot \delta(z,3)_n + \beta_{P\_N} \cdot UF_{N,n} + \beta_{TA3} \cdot \delta(d,3)_n \quad (6)$$

$$V_{F,n} = \beta_{0\_F} + \beta_{NT} \cdot NT_{F,n} + \beta_{TT} \cdot TT_{F,n} + \beta_{D\_F} \cdot TD_{F,n} \quad (7)$$

All symbols used are given in Tab. 2.

Tab. 2. Key to Symbols Used – Part II

Symbol	Name and Description
$P_n(F)$, $P_n(N)$	Choice probability of alternatives F, N - n-th observation
$V_{N,n}$, $V_{F,n}$	Alternatives' systematic utilities - n-th observation
β_0	Alternative-specific constant (ASC_F)
β_{NT}	Parameter to attribute "Number of Transfers" (BETA_NT)
β_{PS}	Parameter to attribute "Trip Party Size" (BETA_PSIZE)
β_{WOR}	Parameter to dummy variable "Work Trip" (BETA_WOR)
β_{BUS}	Parameter to dummy variable "Business Trip" (BETA_BUS)
β_{P_N}	Parameter to attribute "Unit Cost" (BETA_BUS)
β_{TA3}	Parameter to dummy variable "Sunday" (BETA_TA3)
β_{D_F}	Parameter to attribute "Travel distance" (BETA_D_F)
δ ( )	Dummy variables for selected trip purpose z or weekday d

The nine functional parameters were estimated by the standard software *BIOGEME 2.0* (Bierlaire, 2009), obtaining an adjusted rho-square of 0.444, whereas all parameters' p-values < 0.001.

The estimated results of the final version 4.2.4 are listed in Tab. 3.

Tab. 3. Estimation Results (Source: Own Representation)

Parameter	Value	Std.error	t-test
ASC_F	-1.50	.225	-6.68
BETA_BUS	-1.76	.135	-13.05
BETA_D_F	.0104	.000502	20.79
BETA_NT	-.635	.0598	-10.61
BETA_PSIZE	1.44	.190	7.58
BETA_P_N	-.0231	.00521	-4.43
BETA_T	-.0121	1.01 e-009	-11995
BETA_TA3	-.461	.107	-4.30
BETA_WOR	-.817	.121	-6.76

All coefficients have both correct signs and significance. Dummy variables were introduced for influential trip characteristics, that is, party size, certain trip purposes and weekdays. Travel time and transfers contribute most of systematic utility.

With an unexpectedly low beta for the unit fare (€ per person-kilometer), the resulting valuations for an hour of extra travel time and an additional required transfer are considerable, yet fast declining with the distance (Tab. 4).

Tab. 4. Valuations of Time, Transfers (Own Representation)

Trip distance [km]	Equivalent [€/Pkm] for additional transfer	Equivalent [€/Pkm] per hour travel time
100	.275	.314
200	.137	.157
500	.055	.063

A sensitivity analysis was conducted in order to overview and interpret the model's practical implications. In Fig. 11 the choice probabilities for two trip purposes and varying average speeds were depicted along the distance axis.

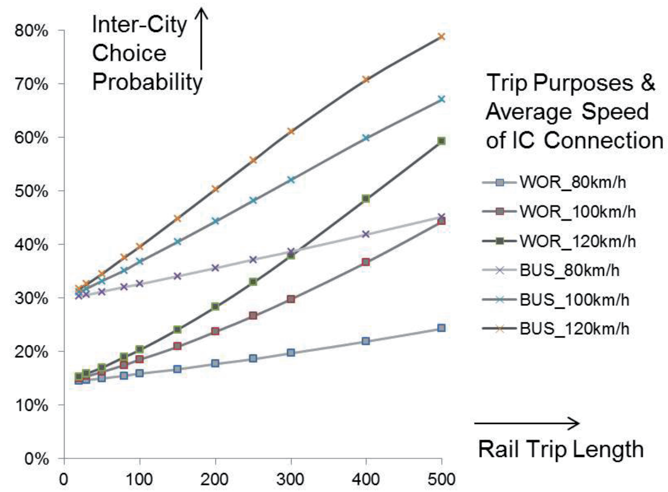

Fig. 11. Analysis of IC Choice Probabilities (Source: Authors' Representation)

For the same setting, the time elasticities with respect to variations in Inter-City travel speeds and distances were computed for business and work trips (Fig. 12). Note that this is not the reaction to a global demand level for railways. Only the intra-modal split is affected by this.

The results shown for considerable speed-up scenarios need to be interpreted with some caution in the case of minor itinerary adjustments. Hjorth and Fosgeraus (2012) and others sought to explain the phenomenon of low marginal value of travel time for just small improvements.

Finally, Fig. 13 depicts the sensitivity of the decision makers to the number of necessary train changes, exemplary for a small to medium-range O-D pair, an average speed of 90 km/h and a unit fare of .14 €/Pkm while using regional trains.

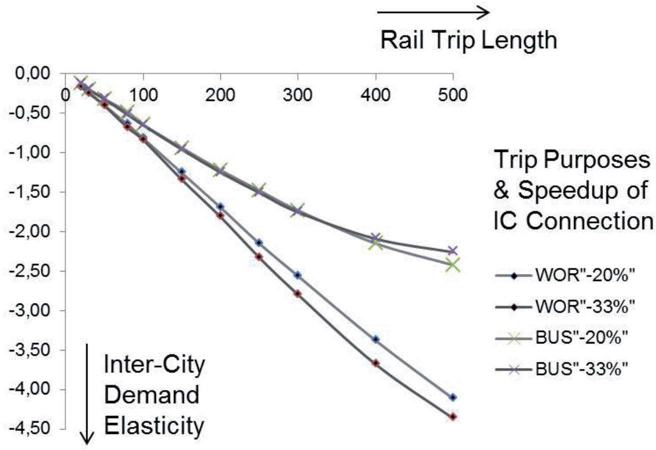

Fig. 12. Time Elasticities (Source: Own Representation)

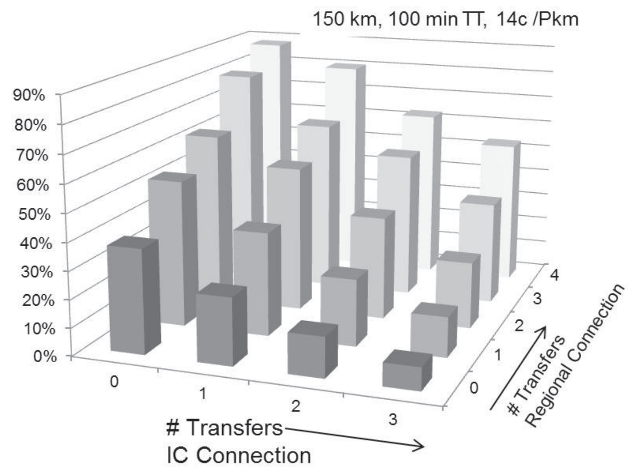

Fig. 13. Sensitivity to the Numbers of Transfers (Source: Authors' Representation)

5 Conclusions

The first task was to assure that the timetable network models of the Integrated State Transport Model of Thuringia were capable of ex-post reconstructing the exact itineraries from two passenger surveys successfully, providing the basis for applications such as ticket fare clearing (Kusakabe et al., 2010).

Further to this step's completion, the railway route choice model was created to refine the assignment of passenger flows in the context of the state-wide model. The choice probability could be described as a function of alternative-specific service characteristics and trip-/trip-maker-specific constants. As a result, the intra-modal proportion of Inter-City passengers versus regional network captives can be obtained for different settings, highly improving the accuracy level of load patterns.

Moreover, it can be stated that there is no general inclination to use regional train network to a higher extent at longer distances than one would expect. The different intercepts and valuations of time, costs and transfers of demand segments were realized by introducing binary variables.

Several hypotheses on service variables of future interest such as

- Capacity constraints (Nuzzolo and Crisalli, 2012)
- Reaction thresholds with regard to travel time improvements (Obermeyer et al., 2014)
- amenities of different long-distance products (e.g. Wi-Fi, restaurants offered on board most of the ICE fleet),
- on-time performance or related reliability measures (see e.g. Rochau et al., 2011),
- headways of (return) connections,
- different sensitivity by frequent users, and
- seasonality factors

should also be considered by envisaged model improvements. With the databases currently available, these variables could not be reconstructed for the past and/ or cannot be operationalized for the application yet.

References

Ben-Akiva M. E. and Lerman S. R. (1985) *Discrete Choice Analysis*. Cambridge: MIT Press.

Bierlaire M. (2009) *Estimation of discrete choice models with BIOGEME 1.8*. Tutorial. EPFL Lausanne.

Dömming A. and Fritzlar E. (2013) *Modellgüte des Routingverfahrens für das Verkehrsmodell Thüringen.* Internal Report. Erfurt: Inst. Verkehr und Raum.

Friedrich M., Hofsaess I. and Wekeck S. (2007) *Timetable-Based Transit Assignment Using Branch and Bound Techniques*. Transportation Research Record, 1752, pp. 100-107. DOI: 10.3141/1752-14

Gao S. and Wu Z. (2011) *Modeling Passenger Flow Distribution Based on Travel Time of Urban Rail Transit*. Journal of Transportation Systems Engineering and Information Technology, 11 (6), pp. 124-130. DOI: 10.1016/s1570-6672(10)60156-0

Garmendia M., Urena J. M. and Coronado J. M. (2011) *Long-distance trips in a sparsely populated region: The impact of high-speed infrastructures.* Journal of Transport Geography, 19 (4), pp. 537-551. DOI: 10.1016/j.jtrangeo.2010.06.002

Gesellschaft für Verkehrsberatung und Systemplanung mbH (GVS) (2010) *Verkehrserhebungen in Nordthüringen*. Final Report, Hannover

Hensher D. A., Rose J. M. and Greene W. H. (2005) *Applied Choice Analysis.* United Kingdom: Cambridge University Press.

Hjorth K. and Fosgeraus M. (2012) *Using prospect theory to investigate the low marginal value of travel time for small time changes.* Transportation Research Part B: Methodological, 46 (8), pp. 917-932. DOI: 10.1016/j.trb.2012.04.001

ISUP (2009) *Verkehrserhebung im Dieselnetz Ostthüringen. Untersuchung-für die NVS mbH Erfurt*. Dresden.

Kusakabe T., Iryo T. and Asakura Y. (2010) *Estimation method for railway passengers' train choice behavior with smart card transaction data*. Transportation, 37 (5), pp. 731-749. DOI: 10.1007/s11116-010-9290-0

Liu Y., Bunker J. and Ferreira L., (2010) *Transit Users' Route Choice Modelling in Transit Assignment: A Review*. Transport Reviews, 30 (6), pp. 753-769. DOI: 10.1080/01441641003744261

Nuzzulo A. and Crisalli U. (2012) *A schedule-based assignment model with explicit capacity constraints for congested transit networks*. Transportation Research Part C: Emerging Technologies, 20 (1), pp. 16-33 DOI: 10.1016/j.trc.2011.02.007

Obermeyer A., Treiber M. and Evangelinos C. (2014) *Thresholds in choice behaviour and the size of travel time savings*. ITEA Annual Conference, Toulouse.

Otuzar J., and Willumsen G. (2011) *Modelling Transport*. 4th edition, New York: John Wiley and Sons.

Rochau N., Noekel K., and Bell M. A. (2011) *Strategy-based transit assignment model for unreliable networks*. European Transport Conference, Glasgow.

Scherer M. (2010) *Is Light Rail More Attractive to Users Than Bus Transit?* Transportation Research Record, 2144 (2), pp. 11-19. DOI: 10.3141/2144-02

Train K. E. (2009) *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Vrtic M. and Axhausen K. W. (2002) *The impact of tilting trains in Switzerland: A routechoice model of regional- and long distance public transport trips.* Proc. 82nd Annual Meeting of the TRB. Washington D.C.

Xu R-H., Luo Q. and Gao P. (2009) *Passenger Flow Distribution Model and Algorithm for Urban Rail Transit Network Based on Multi-route Choice*. Journal of the China Railway Society, 31 (2), pp. 110-114.