Mode Choice Modeling for Long-Distance Travel

Paper submitted to the Transportation Research Board for the Annual Meeting 2013

Rolf Moeckel, Dr.-Ing. (Corresponding author)
Parsons Brinckerhoff
6100 NE Uptown Blvd, Suite 700
Albuquerque, NM 87501
Phone (505) 878-6553
moeckel@pbworld.com

Rhett Fussell, PE
Parsons Brinckerhoff
434 Fayetteville St, Suite 1500
Raleigh, NC 27601
Phone (919) 836-4075
fussell@pbworld.com

Rick Donnelly, PhD
Parsons Brinckerhoff
6100 NE Uptown Blvd, Suite 700
Albuquerque, NM 87501
Phone (505) 878-6524
donnellyR@pbworld.com

4,948 words and 8 figures/tables (equivalent of 250 words) = 6,948 words

Summary

With the ongoing debates from Florida to California and throughout the country concerning the benefits of high-speed rail, there is a renewed interest in intercity mode choice modeling. The investments for improving long-distance travel are substantial and may have serious impacts on travel demand, the environment and the economy. As such, alternatives for improving long-distance travel require careful evaluation before decisions are made on the form and design of long-distance travel infrastructure. A new nested multinomial logit mode-choice model has been developed that is sensitive to travel costs, distance, transit station accessibility, service frequency, number of transfers and parking costs. On the auto side the model considers the modes drive-alone and shared-ride with 2, 3 and 4 or more passengers. The transit side models regional bus, rail and air as modal options. To explore the model sensitivities, scenarios on increased gasoline prices and improved bus service are described.

After a short introduction, the state-of-the-art of mode choice modeling is reviewed. Section 3 explains how total travel demand is generated, and section 4 describes the mode choice model developed in this paper. Section 5 describes the application to the North Carolina Statewide Transportation Model (NCSTM) and section 6 shows the scenario application. The paper ends with conclusions and future research.

Acknowledgement
The development of the long-distance mode choice model has been funded in part by the North Carolina Department of Transportation. Special thanks go to Joe Kolousek and Carlee Clymer for collecting and processing travel data on rail, bus and air long-distance travel.

1. INTRODUCTION

In the last few years, a new interest in mode choice analysis has risen due to the controversy regarding the implementation of high-speed rail in different parts of the U.S. Analysis tools, however, have not caught up with this new demand in transportation modeling. The vast majority of mode choice models developed over the last few decades have been implemented for urban models with a focus on short-distance travel, where modal availability is different from long-distance travel. The travel behavior in long-distance travel is quite different, too, as people tend to be more familiar with modal options for short-distance travel than for long-distance travel. In addition, the composition of travelers differs. While short-distance use of transit is dominated by commuters, long-distance transit modes (particularly rail and air) are heavily used for pleasure trips as well as by business travelers. Given the fact that long-distance travelers tend to stay longer at their destination, travel time tends to be a less dominant factor in mode choice than in short-distance travel.

Investments for improving long-distance travel often are tremendous. Adding a lane to an existing highway or even building a new highway may cost millions of dollars, just as adding a new rail line or improving the speed on an existing rail line may be cost-intensive. Environmental impacts may be serious, as increased auto traffic or air travel may increase gaseous emissions and noise levels substantially. Finally, the economic impact may be significant as well. According to Krugman [1], more accessible regions are ceteris paribus economically more successful. Thus, long-distance infrastructure may affect economic growth of different regions notable. Given the impacts on travel demand, congestion, the environment and the economy, changes to long-distance infrastructure ought to be analyzed carefully before investments are made.

Understanding mode choice is an integral part of transportation analysis. Foremost, mode choice is relevant if any kind of modal availability is analyzed, such as the implementation of a new rail service, extension of existing bus service or making auto travel faster, slower, more expensive or cheaper. Even if only the auto side is under investigation, modal analysis is relevant. Removing the right number of travelers from the highway system that are using transit is important just to get the number of auto travelers correct. Finally, mode choice analysis may be relevant even where no transit is available in the area studied. Often, high-occupancy vehicles and the willingness to pay a toll is treated as a mode choice and can be dealt with more rigidly in mode choice than in trip assignment.

2. STATE-OF-THE-ART

Grayson [2] developed one of the earliest logit-based mode choice models for long-distance travel. His model covers long-distance travel of 100 miles [1 mile = 1.61 km] or more and the estimation is based on the National Travel Survey from 1977. Coefficients were estimated for the purposes business, social and entertainment. The incremental logit model developed by Koppelman [3] is an interesting alternative for mode choice modeling, as it only requires ridership on the existing modes and the properties of the new transit service to be tested. This model allows studying the effects of adding or changing a single mode of transport. The model was developed for the urban context, and to our knowledge it has not been tested for long-distance mode choice modeling.

Forinash and Koppelman [4] compared the traditional multinomial mode choice model with a nested mode choice model. They conclude that a major disadvantage of the multinomial model is that if one mode of travel changes, the relative probability of choosing an unchanged mode is fixed, which is also called the independence of irrelevant alternatives problem. The nested mode choice model, in contrast, affects modes that are grouped in the same nest as the changing mode to a larger extend than modes in a different nest. They tested six different nesting structures with the four modes auto, bus, train and air and found that a reasonable nesting of modes results in more plausible sensitivities of a mode choice model. Specifically to analyze high-speed rail ridership, Wen and Koppelman [5] developed a Nested Logit model with Paired Combinational Logit, Cross-Nested Logit and Product Differentiation Model, which they applied to the planned high-speed rail corridor from Montreal to Toronto. While the multinomial model led to mode-specific constant of up to 8.2, the generalized nested logit model reduced the highest constant to 5.3. Bhat [6] developed a heteroscedastic extreme value model of intercity mode choice that also overcomes the independence of irrelevant alternatives property of the more commonly used multinomial logit model, and found that it worked better than logit or probit models.

Baik et al. [7] developed a complete travel demand model for long-distance travel by Auto, Air and "Air Taxi", including generation, distribution, mode choice and assignment. The estimation is based on the somewhat dated American Travel Survey (ATS) of the Bureau of Transportation Statistics (BTS) from 1995. The mode train was left out due to limitations in the survey data. De Lapparent et al. [8] analyzed stated preference data on mode choice decisions in the Czech Republic, Switzerland and Portugal. They find quite different travel behavior, which they explain with different travelers' preferences in these three counties. Partly, however, this may also be due to comparatively small data samples in this study. Grimal [9] points out that the habit of short-distance mode choice affects long-distance mode choice.

Abdelwahab [10] implemented and calibrated two long-distance mode choice models in Canada, one in the East and one in the West. He tested the transferability of the two models and found a transferred model to be 18-23% less accurate than a locally estimated and calibrated model. After adjusting the mode-specific constants to reflect observed modal shares in the application context, the predictive ability of the models improved by about 10%. Unfortunately, the paper does not document the mode-specific constants that were calibrated. It may well be that smaller mode-specific constants make a model more transferable, as more power is given to the explanatory variables, rather than correcting aspects not captured by the model with constants. Wilson et al. [11] implemented two long-distance mode choice models for the eastern and western regions of Canada. The model distinguishes auto, bus, rail and air. For reasons not further discussed, the eastern model is using unusually high constants (up to 18.016 for rail), while the western model has much more reasonable constants (up to 1.362 for bus).

Blackstone et al. [12] analyzed a survey for airport choice in the corridor from Baltimore to New York (including the airports BWI, EWR, JFK and PHL), which by definition is part of the decision process for long-distance travel. They find that the location of the workplace plays a crucial role when selecting an airport, as many trips by air are business trips that start or end at the workplace. Başar and Bhat [13] developed a probabilistic choice set multinomial logit model for the airport choice in the San Francisco Bay Area (including the airports SFO, SJC and OAK).

They find that access time to the airport is the most important explanatory variable, but also frequency of service has influence on the airport choice.

The review of the literature revealed important benefits of a nested logit model compared to a traditional multinomial logit model. It also showed the limited transferability of existing models to other study areas without recalibrating the model. The relevance of selecting the most likely transit station, which may not be the closest station, became apparent. Most existing models, however, do not take into account special attributes of long-distance travel, such as specific long-distance travel purposes, severe travel time extensions by check-in or security procedures for some transit modes, or the fact that travelers do not necessarily live right next to a transit station, nor do they necessarily travel to destinations right next to a transit station. The model developed in this paper aims at surpassing these limitations.

3. LONG-DISTANCE TRAVEL DEMAND

To generate the travel demand, a Nationwide Estimate of Long-Distance Travel (NELDT) has been developed and implemented to simulate person long-distance travel across the U.S. [14]. Figure 1 shows the workflow of NELDT. The key input is the long-distance element of the National Household Travel Survey (NHTS) 2002[1], which contains 45,165 trip records of long-distance trips of 50 miles or more. Unfortunately, the NHTS 2009[1] did not contain a long-distance element, which made it necessary to use the previous NHTS dataset.

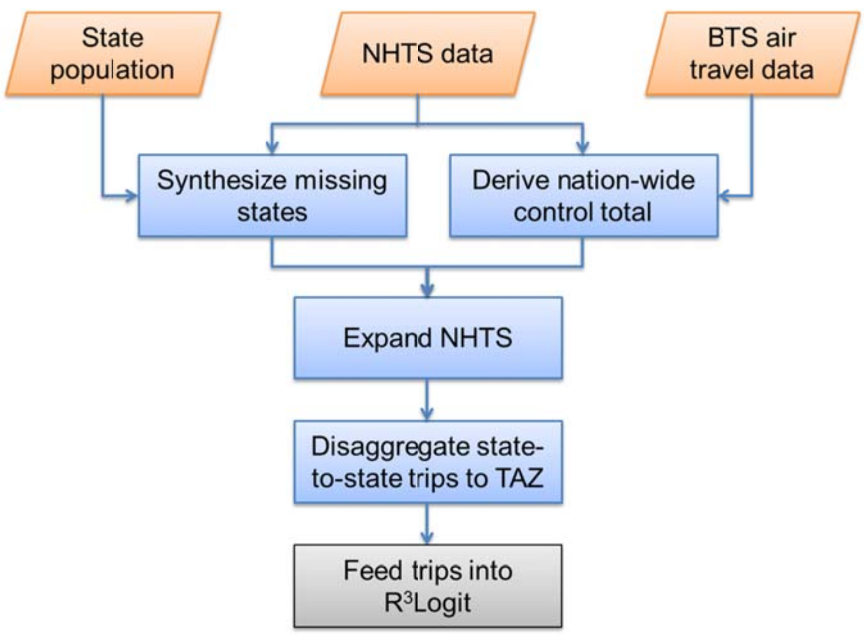

FIGURE 1: NELDT design

NHTS provides long-distance travel records by home state of the traveler. For privacy reasons, the NHTS dataset only reports the origin state for trips from states with a population of 2 million or more. For smaller states, synthetic data records were generated based on using travel data records of surrounding states for which records were available.

[1] Available for download at http://nhts.ornl.gov/download.shtml

The NHTS data records in combination with the synthesized records are considered to be a representative sample of long-distance travel in the U.S. The Bureau of Transportation Statistics (BTS) provides a ten percent sample of all ticketed air passengers[2]. This value is used as a control total to expand the NHTS data to a nationwide long-distance travel data set.

Three trip purposes are distinguished, namely business, personal and commute trips. These purposes intentionally differ from trip purposes commonly used in urban travel demand models (such as home-based work, home-based shop, home-based other, non-home based, etc.). According to NHTS 2002, personal (often also called pleasure) is the most important trip purpose with 59% of all long-distance trips. Business is the second most common purpose with 28% of all long distance trips. As expenses for a business trip commonly are reimbursed by the employer, trips with this purpose tend to be less price-sensitive and more travel-time sensitive. Commute trips (13% of all long-distance trips) are distinguished as a separate purpose, because in contrast to the other two purposes commute trips are frequently repeated trips, and therefore, tend to be more optimized than infrequent personal or business trips. Furthermore, origins and destinations of commute trips are more constrained by home and work locations than long-distance trips of other purposes.

The NHTS provides long-distance trips by home state and destination state. To increase the resolution, state-to-state trips are disaggregated to zone-to-zone trips using population density and employment as trip generators and attractors. For every trip purpose, employment and population are weighted differently when disaggregating trip origins and destinations. A gravity model ensures that the trip length frequency distribution reported in the NHTS is replicated after trip ends were disaggregated from states to zones.

4. DESIGN OF R^{3}LOGIT

R^{3}Logit has been built as a discrete choice mode selection model. It is based on a mode choice model originally developed as GLogit by Gordon Schultz, an early pioneer in travel demand modeling. Schultz initially developed this model for New Orleans in the mid-1980s, and it was later applied to about half a dozen projects across the U.S. For this application, the model has been revised to suit long-distance travel and to reflect most recent findings in mode choice modeling. R^{3}Logit is designed as a nested multinomial logit model.

FIGURE 2 shows the nesting structure. The definition of the nesting structure has an important impact on the mode split modeled, and the grouping of modal alternatives shall reflect the degree sensitivities across alternatives [4: 99]. Accordingly, R^{3}Logit handles auto and transit at the same level as they are similar high-level choices. For auto, the model further distinguishes drive-alone, shared-ride 2, shared-ride 3 and shared-ride of 4 or more people. This level is relevant for two reasons. First of all, the model shall be able to analyze the impact of alternatives that include High-Occupancy Vehicle (HOV) lanes. Depending on whether the HOV lanes allow 2+ person carpools or 3+ person carpools, the assignment would block these HOV lanes to either drive-alone or to drive-alone and shared-ride 2. Furthermore, the explicit distinction of vehicle occupancy improves modeling the number of vehicles on the highway system. Trip generation and trip distribution handle person trips, and several person trips may generate only one high-occupancy vehicle trip. By dividing person trips by the occupancy, person trips can be converted into vehicle trips. On the transit side, the model distinguishes three modes that are most relevant for long-distance transit trips, namely bus, rail and air.

---

[2] http://www.transtats.bts.gov/databases.asp?Subject_ID=3&Subject_Desc=Passenger%20Travel&Mode_ID2=0

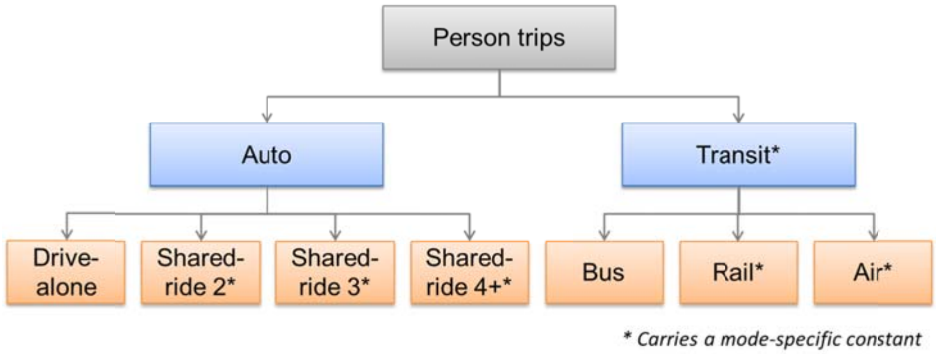

FIGURE 2: Nesting structure of R³Logit

Mode-specific constants are used to account for unincluded attributes, such as the comfort of seats in a certain mode, having a radio in your car, or the perception that a certain mode was more convenient than another mode. At every choice set, there is one option without a constant (i.e. auto, drive-alone and bus), and the constants for the other choices describe the unincluded relative benefit of that choice in comparison to the constant-less choice. The nesting coefficient for the upper level (choice between auto and transit) was set to 0.3, and utilities at the lower level are calculated using the following equations. The utility for drive-alone and shared-ride trips takes into account travel time and distance, auto-operating costs and parking cost and is calculated by equation 1:

$$u_{i,j,m,p} = ivtc \cdot tt_{i,j} + ovtc \cdot autoEgr + prkc_p \cdot \frac{0.5 \cdot p_j}{occ_m} + aocc_p \cdot \frac{dist_{i,j} \cdot aoc}{occ_m} \qquad \text{Equation 1}$$

TABLE 1 lists the coefficients and variables used in this calculation. The out-of-vehicle time coefficient is twice as negative as the in-vehicle time coefficient, as time walking to and from the vehicle tends to be perceived as more burdensome than the travel time in the vehicle.

The auto-egress time (i.e. the time to walk from the parking location to the final destination) is set to be 5 minutes. In future version, this value might change depending on the destination. While five minutes are very reasonable in urban areas, they are somewhat high for suburban and rural areas where travelers usually can park right in front of their destination. However, as most long-distance trips are destined to urban areas, a fairly high value was chosen for this variable.

The number of persons traveling in a single vehicle for the occupancy of 4 or more was based on the assumption that most vehicles do not carry more than 5 passengers, and therefore, the average number of passengers in the category 4+ is likely to be close to 4.

The coefficient on auto-operating costs has less impact on mode choice than parking costs. This is based on the assumption that out-of-pocket costs for parking are perceived as more onerous than auto-operating costs that include hidden expenses for purchase, maintenance and insurance of a vehicle.

In Equation 1, parking costs are multiplied by 0.5 as they equally apply to the outbound and the return trip. As travel time and costs are accounted for one way only, parking costs are divided in half to be consistent. Monetary costs are divided by the number of passengers to account for savings of car-pooling.

TABLE 1: Coefficients and variables for utilities of drive-alone and share-ride modes

Coefficient	Description	Value
$u_{i,j,m,p}$	Utility of mode *m* (drive-alone, shared-ride 2, shared-ride 3 or shared-ride 4+) from zone *i* to zone *j* for purpose *p*	
ivtc	In-vehicle time coefficient: Parameter evaluating the travel time spend in the vehicle	-0.025[*]
$tt_{i,j}$	Travel time from zone *i* to zone *j* by auto	
ovtc	Out-of-vehicle time coefficient: Parameter evaluating the travel time spend out of the vehicle	-0.05[*]
autoEgr	Auto-egress time: Time spent to walk from the parking location to the final destination	5 min
$prkc_p$	Parking costs coefficient by purpose *p*	(b) -0.006, (p) -0.012, (c) -0.010[*]
p_j	Parking costs in zone *j*	
occ_m	Number of persons traveling by drive-alone, shared-ride 2, shared-ride 3 and shared-ride 4+	1, 2, 3, 4.1 persons
$aocc_p$	Auto-operating costs coefficient by purpose	(b) -0.0009, (p) -0.0039, (c) -0.0029[*]
$dist_{i,j}$	Distance from zone *i* to zone *j* in miles	
Aoc	Auto-operating costs	0.0874 cents per mile

[*]In line with nested mode-choice theory, parameters are scaled by the nesting coefficient

The utility for transit trips takes into account time and cost of the transit trip, transit access and egress as well as frequency of service:

$$u_{i,j,m,p} = ivtc \cdot tt_{i,iStat,auto} + ovtc \cdot trnAcc_m + ivtc \cdot tt_{iStat,jStat,m} + ntrc \cdot trnsf_{iStat,jStat,m} + trfc_p \cdot fare_{iStat,jStat,m} + tfqc_p \cdot \frac{frqu_{iStat,jStat,m}}{dist_{i,j}} + ovtc \cdot trnEgr_m + ivtc \cdot tt_{jStat,j,auto}$$

Equation 2

The utility for a transit trip includes

(a) the trip from the origin to a transit station,
(b) access time to transit,
(c) transit travel time, number of transfers, transit fare and frequency of service,
(d) egress time and
(e) the trip from a transit station to the final destination.

TABLE 2 lists the coefficients and variables used to calculate the utilities of transit modes. The transit fare coefficients are set equal to the parking cost coefficients to ensure consistency across modes. Under the assumption that travelers of trips with a very long distance are less concerned about infrequent travel connections, the frequency of service is divided by the distance traveled. For example, if someone is traveling 2,000 miles by air, this person is likely to stay for a longer period of time, and therefore has some flexibility regarding departure and arrival time. On the other hand, a traveler making a trip of 75 miles is likely to stay for a shorter time, and therefore, expects a more frequent service to suit her or his travel plans.

TABLE 2: Coefficients and variables for utilities of transit modes bus, rail and air

Coefficient	Description	Value
$u_{i,j,m,p}$	Utility of mode *m* (bus, rail or air) from zone *i* to zone *j* for purpose *p*	
Ivtc	In-vehicle time coefficient: Parameter evaluating the travel time spend in a vehicle	-0.025*
$tt_{i,j,m}$	Travel time from zone *i* to zone *j* on mode *m*	
Ovtc	Out-of-vehicle time coefficient: Parameter evaluating the travel time spend out of the vehicle	-0.05*
$trnAcc_m$	Time to access the transit, which includes the walk from the vehicle to the transit platform or gate, time for check-in and time for security checks	Bus: 15, Rail: 30, Air: 60
Ntrc	Coefficient on number of transit transfers	-0.01*
$trnsf_{iStat,jStat,m}$	Number of transfers to travel from *iStat* to *jStat* on mode *m*	
$trfc_p$	Transit fare coefficient by purpose	(b) -0.006, (p) -0.012, (c) -0.010*
$fare_{iStat,jStat,m}$	Transit fare from *iStat* to *jStat* on mode *m*	
$tfqc_p$	Coefficient on frequency of service per day by mode	(b) 0.2, (p) 0.1, (c) 0.1*
$Frqu_{iStat,jStat,m}$	Frequency of service from *iStat* to *jStat* on mode *m* per day	
$dist_{i,j}$	Distance from zone *i* to zone *j*	
$trnEgr_m$	Time to egress the transit, which includes the walk from the transit platform or gate to a vehicle and time for collecting baggage	Bus: 10, Rail: 15, Air: 20

*In line with nested mode-choice theory, parameters are scaled by the nesting coefficient

Special attention is given to finding the best transit stations for a given origin-destination pair. In some cases, it might be beneficial to choose a transit station that is further away from the trip origin in order to catch a non-stop transit connection or to save on fare. Therefore, the three closest transit stations to both the origin and the destination are evaluated for a trip (FIGURE 3). Using Equation 2, the utilities are calculated for every *iStat*/*jStat* combination. Out of these nine alternatives, the *iStat*/*jStat* combination with the highest utility is chosen to be evaluated in the mode choice model.

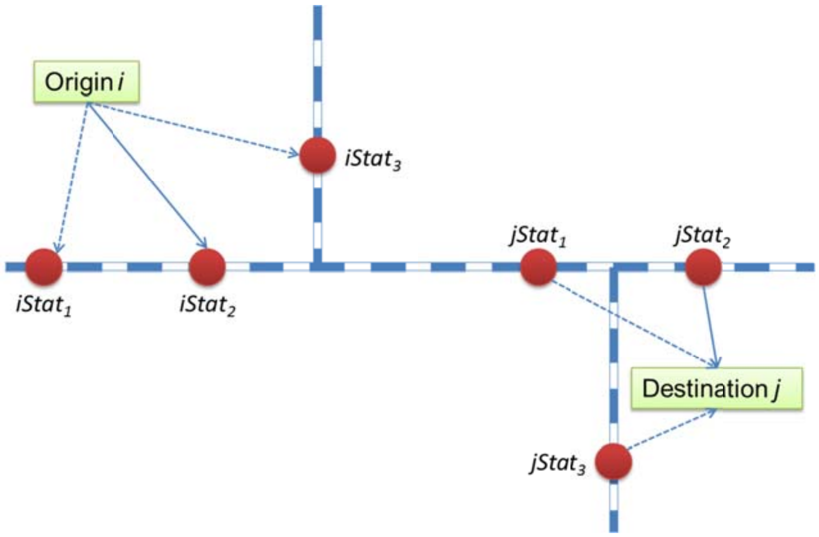

FIGURE 3: Evaluation of three transit stations near origin and destination

Equations 1 and 2 are used to calculate the utilities of all seven modes at the lowest level. The utilities at the higher level (to choose between auto and transit) are calculated using mode choice logsums, including all modes that belong into this nest:

$$u_{i,j,\dot{m},p} = nc \cdot \ln\left(\sum_{m \in \dot{m}} \exp\left(u_{i,j,m,p}\right) \cdot \exp\left(c_m\right) \right) \qquad \textbf{Equation 3}$$

Equation 3 sums up the utilities calculated either in equation 1 (for auto) or in equation 2 (for transit) and adjusts them by their respective mode-specific constant. Using the respective share of utilities, the mode split is calculated.

TABLE 3: Coefficients and variables for utilities of the auto and transit choice

Coefficient	Description	Value
$u_{i,j,\dot{m},p}$	Utility of mode $\dot{m}$ (auto or transit) from zone i to zone j for purpose p	
nc	Nesting coefficient	0.3
$u_{i,j,m,p}$	Utility of mode m (which is part of the nest of $\dot{m}$) from zone i to zone j for purpose p	
c_m	Mode-specific constant to account for unincluded attributes of mode m	

The coefficients used in R³Logit are not estimated but rather heuristically derived. No data were available to estimate coefficients. The NHTS data, which was used to generate the total travel demand, has a rather low share of train and bus travel records, making it questionable to estimate coefficients using these data. A comparison of estimated coefficients found in the literature reveals that many estimations reveal very similar coefficients. Therefore, the coefficients used in this model are considered to be generic enough to be applied in different study areas. To confirm that the setting of a given parameter does not disturb model sensitivities, a series of sensitivity test were run to ensure that no single parameter drives the model results. Adjustments were made to the parameters $trfc_p$ (Coefficient on transit fare), $ntrc$ (Coefficient on number of transit transfers), *autoEgr* (Auto egress time) and $Frqu_{iStat,jStat,m}$ (Frequency of transit service for selected origin/destination pairs). The model results changed in the expected direction, though the impact of changing a single parameter on the overall mode split was very small.

5. APPLICATION IN NCSTM

The North Carolina Department of Transportation (NCDOT) did not have the ability to get a relatively quick and approximate but consistent and defensible, estimate of how different patterns of future development change key measures of transportation performance, and a tool that can contribute to discussion and other evaluation tools that address how future transportation investments may affect future development patterns. Specifically, NCDOT wanted to test high level decisions regarding non-highway modes. The State is looking towards high-speed rail options, improvements to airports statewide (both access and service frequency) as well as upgrades to Amtrak rail service and intermodal connections. The R^3Logit model allows for this high-level analysis of long-distance travel choices.

A multi-state region was chosen as the boundary for development of the long-distance transit system. Included are all states east of the Mississippi River. This allowed capturing almost all long-distance trips that have an actual choice between modes, with longer trips being captured by air almost exclusively. Key cities in these states were included in the transit network, which were chosen by having both a connection with Amtrak and a larger airport. The transit network is limited to trips with at least one trip end within North Carolina. The mode share of through trips (External-to-External) is given by the NHTS data and kept static.

Transit data including fare, frequency, number of transfers and travel duration was manually collected for Amtrak, Greyhound, GotoBus and Coach America NC using their passenger routing websites. For air travel, a meta search engine[3] was used to gather similar travel data.

Any mode choice model can only account for a limited number of travel parameters. To correct for unincluded attributes, mode-specific constants were calibrated to match observed travel behavior. As no comprehensive mode share data were available for North Carolina, the NHTS 2002 was used to estimate the mode split. Still, with only 1,822 records of long-distance trips in North Carolina, the number of records was deemed as being too small to calculate the observed mode split. Therefore, the target mode share was calculated using records from the Southeast of the U.S. (including FL, GA, KY, NC, SC, TN, VA and WV), which provided a total of 10,022 records. After calibrating the mode-specific constants, the mode split shown in TABLE 4 is matched precisely by R^3Logit.

[3] Available at http://matrix.itasoftware.com

TABLE 4: Observed mode split and calibrated mode-specific constants

Mode	Observed share			Mode-specific constants		
	Business	Personal	Commute	Business	Personal	Commute
Auto	47.5%	66.3%	91.8%	0	0	0
Drive-alone	18.5%	6.5%	62.3%	0	0	0
Shared-ride 2	14.1%	20.9%	18.7%	0.1227	0.5414	-0.1659
Shared-ride 3	7.5%	15.4%	7.5%	0.0536	0.5657	-0.3241
Shared-ride 4	7.4%	23.5%	3.3%	0.1443	0.7833	-0.4808
Transit	52.5%	33.7%	8.2%	4.1564	2.3716	0.1655
Bus	31.9%	10.2%	2.3%	0	0	0
Rail	0.3%	0.5%	1.7%	-4.9869	-3.0159	-0.8836
Air	20.3%	23.0%	4.2%	0.4784	2.2847	2.4338

The calibrated mode-specific constants are small on the auto side, which is desirable. On the transit side, constants are somewhat higher than desired, even though these constants are lower than many constants published by previous papers. Higher constants reduce the model sensitivities, as a larger share of the model result is explained by constants. The commute purpose constants are comparatively small, and thus provide the most reliable model. The high negative constant on rail travelers for the business and personal purposes is probably caused by limitations of the observed data. While BTS air travel statistics in comparison with AMTRAK ridership suggest that there should be 16-times as many air passengers as train passengers, the NHTS data for the southeastern states suggest that there were 41-times as many air passengers as train passengers. Rail passengers may be largely underrepresented in this dataset, which is likely to be the cause for the relatively high constants on rail.

6. SCENARIO ANALYSIS

To analyze the model sensitivities, two sample scenarios where modeled using R^3Logit in NCSTM. One scenario implements an express bus service between Raleigh and Charlotte in North Carolina. It is assumed that this bus receives a reserved lane, allowing the bus to travel at free-flow speed. Travel time is 2.5 hours, costs are assumed to be \$10, and a frequency of 10 busses per day is assumed. This is an improvement over current bus service of 2 hours and 50 minutes for \$13, 7-times per day. Another scenario analyzes the impact of increased auto-operating costs. In this scenario, the price for gasoline increases, resulting into tripled auto-operating. As the first scenario only affects the Raleigh and Charlotte areas, only trips that have their origin in Wake County (Raleigh) and their destination in Mecklenburg County (Charlotte), or vice versa, are included in the summaries of TABLE 5.

TABLE 5: Modal share for trips between Wake and Mecklenburg Counties by scenario

Mode	Base	Express Bus		Tripled AOC	
	Mode Share	Mode Share	Difference	Mode Share	Difference
Drive-alone	9.9%	8.4%	-1.4%	8.9%	-1.0%
Shared-ride	5.7%	5.3%	-0.5%	5.7%	0.0%
Bus	44.8%	56.7%	11.9%	45.1%	0.3%
Rail	0.0%	0.0%	0.0%	0.0%	0.0%
Air	39.6%	29.7%	-10.0%	40.4%	0.7%

Improving the bus service between Charlotte and Raleigh attracts additional bus riders, drawing passengers in particular from air travel. The scenario with tripled auto-operating costs, in contrast, has little impact on mode shares. This is in line with the observed changes in mode split during the recent gas price peak in 2008, where oil prices rose quickly but mode split was not affected by a large margin.

7. CONCLUSIONS

Research in the U.K. suggests that while trips with a distance of more than 50 miles make up only 2 percent of all journeys, these trips account for 31 percent of all vehicle miles traveled [15: 8]. A similar relationship was found for other European countries with 0.5 percent of all trips being more than 100 km long and contributing roughly 20 percent of total kilometers traveled [16]. Thus, long-distance travel contributes significantly to vehicle-miles traveled, congestion and emissions. Most research in mode-choice modeling, however, focuses on urban, short-distance travel. While it might be easier to implement transit options at the urban scale, statewide and mega-regional planning agencies are required to understand modal options for long-distance travel. R^3Logit is meant to contribute to analyzing scenarios that affect the mode share for long-distance travel.

Though the model is able to analyze a variety of scenarios in its current state, improvements are envisioned to enhance the applicability of the model. At this point, most concerning are the relatively high constants needed to match observed mode shares. Even though these constants are lower than those used in most other long-distance mode split models found in the literature, the size of the constants may still limit the policy sensitivity of the model. It appears that a bias in the observed mode split data might be a major cause for these relatively high constants. As it is very expensive to obtain better surveys for long-distance (which need to cover auto, bus, rail and air travelers), alternative forms of data collection, such as mobile phone data, appears to be promising to overcome this shortcoming. Aguilar et al. [17] found a reasonable accuracy of mobile phone GPS data on different modes, which may be used to determine the mode of transport of long-distance trips. Given the large quantity of data records that possibly could be retrieved from mobile phone data, the modal share derived this way is expected to be much more representative than the limited number of data records from NHTS 2002 that where available for this study.

The current model does not aim at quantifying induced demand. Induced demand, as defined by Lee et al. [18], describes travel demand that is generated as a consequence of infrastructure improvements. In other words, by making a certain destination more accessible, more people will decide to travel to that destination. The magnitude of induced demand may differ by which mode has been improved. Scherer [19] found that improved light rail tends to generate more induced demand than improved bus service. Weis and Axhausen [20] were able to quantify the induced demand based on historic auto travel demand in Switzerland, though findings are probably not transferable to different modes and different settings of competing modal alternatives. A common approach to estimate induced demand is the use of mode choice logsums, which are an aggregate across different modes that describe the ease of traveling between two locations. However, the coefficient used in such estimation is mostly guesswork, and has to be set individually for every application. With specific projects emerging, induced demand will be estimated based on similar projects elsewhere when applying R^3Logit in production mode.

8. REFERENCES

1. Krugman, P., *Increasing Returns and Economic Geography.* Journal of Political Economy, 1991. **99**(3): p. 483-499.
2. Grayson, A., *Disaggregate Model of Mode Choice in Intercity Travel.* Transportation Research Record: Journal of the Transportation Research Board, 1981. **835**: p. 36-42.
3. Koppelman, F.S., *Predicting Transit Ridership in Response to Transit Service Changes.* Journal of Transportation Engineering, 1983. **109**(4): p. 548-564.
4. Forinash, C.V. and F.S. Koppelman, *Application and Interpretation of Nested Logit Models of Intercity Mode Choice.* Transportation Research Record: Journal of the Transportation Research Board, 1993. **1413**: p. 98-106.
5. Wen, C.-H. and F.S. Koppelman, *The generalized nested logit model.* Transportation Research Part B: Methodological, 2001. **35**: p. 627-641.
6. Bhat, C.R., *A Heteroscedastic Extreme Value Model of Intercity Mode Choice.* Transportation Research Part B: Methodological, 1995. **29**(6): p. 471-483.
7. Baik, H., et al., *Forecasting Model for Air Taxi, Commercial Airline, and Automobile Demand in the United States.* Transportation Research Record: Journal of the Transportation Research Board, 2008. **2052**: p. 9-20.
8. de Lapparent, M., A. Frei, and K.W. Axhausen. *Choice of mode for long distance travel: current SP-based models from three European countries*. in *ETC 2009*. 2009. Leeuwenhorst.
9. Grimal, R., *Spatial analysis of long-distance mobility*, in *World Conference on Transport Research*2010: Lisbon, Portugal. p. 24.
10. Abdelwahab, W.M., *Transferability of intercity disaggregate mode choice models in Canada.* Canadian Journal of Civil Engineering, 1991. **18**: p. 20-26.
11. Wilson, F.R., S. Damodaran, and J.D. Innes, *Disaggregate mode choice models for intercity passenger travel in Canada.* Canadian Journal of Civil Engineering, 1990. **17**: p. 184-191.
12. Blackstone, E.A., A.J. Buck, and S. Hakim, *Determinants of Airport Choice in a Multi-Airport Region.* Atlantic Economic Journal, 2006. **34**: p. 313-326.
13. Başar, G. and C. Bhat, *A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area.* Transportation Research Part B, 2004. **38**: p. 889-904.
14. Moeckel, R. and R. Donnelly. *Nationwide Estimate of Long-Distance Travel (NELDT). Generating External Trips for Local Travel Demand Models*. in *Annual Meeting of the Transportation Research Forum*. 2011. Long Beach, CA.
15. Dargay, J. and S. Clark, *The determinants of long distance travel: An analysis based on the British national travel survey*, in *World Conference on Transport Research*2010: Lisbon, Portugal. p. 21.
16. Axhausen, K.W., *Preparing for the long of it: Methodological research for an European survey of long distance travel*, in *TRB Conference: Personal Travel: The long and short of it*2000: Washington, D.C.
17. Aguilar, D.P., et al., *Quantifying Position Accuracy of Multimodal Data from Global Positioning System–Enabled Cell Phones.* Transportation Research Record: Journal of the Transportation Research Board, 2007. **1992**: p. 54-60.

18. Lee, D.B., L.A. Klein, and G. Camus, *Induced Traffic and Induced Demand.* Transportation Research Record: Journal of the Transportation Research Board, 1999. **1659**: p. 68-75.

19. Scherer, M., *Is Light Rail More Attractive to Users Than Bus Transit? Arguments Based on Cognition and Rational Choice.* Transportation Research Record: Journal of the Transportation Research Board, 2010. **2144**: p. 11-19.

20. Weis, C. and K.W. Axhausen, *Induced travel demand: Evidence from a pseudo panel data based structural equations model.* Research in Transportation Economics, 2009. **25**(1): p. 8-18.