Transport Reviews: A Transnational Transdisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ttrv20

Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review

Lei Zhang [a], Frank Southworth [b], Chenfeng Xiong [a] & Anthon Sonnenberg [b]

[a] Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn Martin Hall, College Park, MD, 20742, USA

[b] School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, SEB Building, Atlanta, GA, 30332, USA

To cite this article: Lei Zhang , Frank Southworth , Chenfeng Xiong & Anthon Sonnenberg (2012) Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review, Transport Reviews: A Transnational Transdisciplinary Journal, 32:4, 399-433, DOI: 10.1080/01441647.2012.688174

To link to this article: http://dx.doi.org/10.1080/01441647.2012.688174

Transport Reviews, Vol. 32, No. 4, 399–433, July 2012

Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review

LEI ZHANG*, FRANK SOUTHWORTH**, CHENFENG XIONG* AND ANTHON SONNENBERG**

**Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn Martin Hall, College Park, MD 20742, USA; **School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, SEB Building, Atlanta, GA 30332, USA*

(*Received 19 April 2011; revised 6 April 2012; accepted 21 April 2012*)

ABSTRACT *Since the passage of the Intermodal Surface Transportation Efficiency Act in 1991, a significant number of state highway agencies have started to develop and implement statewide travel demand models to meet policy and legislative development needs. Currently, however, a lack of up-to-date multimodal and inter-regional passenger travel data hampers analysts' ability to conduct quantitative assessments of long-distance travel infrastructure investment needs, at both the national and statewide levels. Despite these data limitations, but also largely shaped by them, long-distance travel modelling has become an increasingly popular topic in recent years. This paper reviews several methodologies for multimodal inter-regional travel demand estimation, drawing examples from both state-specific modelling within the USA and from fully national models being developed and applied in other parts of the world, notably in Europe.*

1. Introduction

This paper provides an assessment of the current state-of-the-art in long-distance passenger travel modelling, including travel demand forecasting, in the USA. Americans travel a great deal. This travel includes many inter-city and other long-distance trips that cross one or more state lines and also international boundaries. In 1995, the date of the last long-distance passenger travel survey (Bureau of Transportation Statistics [BTS], 1997) by the federal government, US households made over one billion personal trips to destinations within the USA and an additional 41 million trips to other countries, logging a total of 827 billion miles of travel, or about 25% of all person miles of travel in the nation in the process. This included a great deal of both business and tourist travel, both major contributors to the national economy, as well as many trips to visit family and friends and to engage in a variety of personal business activities. The financial resources and extra preparation time involved in long-distance travel activities are usually

Correspondence Address: Lei Zhang, Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn Martin Hall, College Park, MD 20742, USA Email: lei@umd.edu

0144-1647 print/1464-5327 Online/12/040399-35

http://dx.doi.org/10.1080/01441647.2012.688174

taken in anticipation of a greater per trip "payoff" in some sense (either monetarily, or socially, or both) than are most of the trips people take. When aggregated over the nation's travelling population these benefits and costs run into many billions of dollars each year. Being able to support continued high levels of personal mobility has therefore become an economic as well as social imperative for the USA, as it has for most countries. Doing so in a cost-effective manner requires that we collect sufficient data to be able to understand current travel patterns as well as project these patterns into the future. Without such information we run the risk of making poor and very expensive investments in our transportation infrastructure and services, as well as failing to anticipate correctly the effects of any federal or state regulation of the transportation system and its use.

With a long history of employing analytical models to guide transportation planning and decision-making, the USA started its travel demand modelling of urban and metropolitan areas as early as the middle of the 20th century (Zhang, Xiong, & Berger, 2010). In contrast, statewide modelling of travel demands received much less attention prior to passage of the Intermodal Surface Transportation Efficiency Act of 1991 (Giaimo & Schiffer, 2005). This situation has started to change. Today more than 35 states have active modelling efforts to meet statewide policy and legislative development needs (Cohen, Horowitz, & Pendyala, 2008; Giaimo & Schiffer, 2005; Horowitz, 2006, 2008; Souleyrette, Hans, & Pathak, 1996). Unlike their urban counterparts, however, these statewide planning efforts have as yet received comparatively limited support in terms of both data collection and modelling advances, where the longer distance forms of trip-making are concerned. By comparison, fully national travel demand modelling has received a good deal more attention in Europe in particular over the past two decades. Recognizing that many of these national modelling efforts equate more closely, at least from the perspective of geography and size of the travelling population, to statewide studies in the USA, this international practice provides some valuable insights into currently available models as well as supporting data collection options. Finally, an additional encouragement for advances in long-distance passenger travel modelling and data collection may also be found in the success of the US Department of Transportation's Freight Analysis Framework, or FAF programme, which in 2010 produced its third generation of inter-regional and multi-modal long-distance freight flow matrices (FHWA, 2011). The resulting national highway traffic flow maps, and notably the projected growth in these flows over time, provide strong visual evidence of the need to consider the implications of such growth for future investment decisions.

2. Policy-Relevant Information Needs

The review of past empirical and policy-directed studies of statewide and national travel activity patterns presented below suggests that the following generic data elements meet the principal data needs of a wide variety of data users in both the public and private sectors:

- Travel purpose
- Mode of transportation and type of vehicle (including access/egress vehicles) used
- Trip origin and destination (OD) locations
- Traveller characteristics

- Travel routes
- Whether domestic or international (foreign) travel
- Travel period duration and type of travel itinerary (days away from home and number of out-of-home stops)
- Travel frequency
- Travel season

Policy sensitive passenger demand studies also require an appropriately quantitative representation of the causes behind personal travel choices, and how these causes result in the trip-making activity patterns we observe in practice. This means that data on the common determinants of personal trip-making need to be collected, and notably data on both the characteristics of travellers and on the types and levels of travel service (LOS) being offered them by the nation's various transportation modes. The empirical literature indicates that the following LOS variables play an important role in the choices made by travellers as to how, how often, and where to travel:

- Monetary travel costs (business and non-business-related expenditures, including transportation, lodging, and other out-of-pocket costs)
- Travel times (including on-time travel reliability and the potential for delays)
- Travel safety
- Travel comfort
- Schedule convenience
- Ease of travel arrangements

For modelling that is nationwide or statewide in scope and whose products are meant to inform planning and policy decisions, this also means being able to represent the transportation LOS variables as well as the OD flows themselves, as costs and movements over specific routes and through specific ports and terminals on the US intermodal transportation network. This includes the derivation of congestion-sensitive travel costs where this traffic uses the same capacity constrained infrastructure as local traffic and/or long-distance freight traffic, as is typically the case on US highways.

The rest of the paper is organized as follows. Section 3 provides a review of current modelling approaches, followed by a summary of some current deficiencies and an example of how we might proceed to address these. Section 4 then provides a brief summary of the major types and major sources of data used to calibrate such models, followed by a summary of current shortcomings and areas for improved data collection.

3. A Synthesis of Multimodal Inter-Regional Demand Analysis Methods

This section provides an overview and assessment of current long-distance travel modelling practices around the world. A search for useful past experience identified three study types:

(1) National models developed in other countries: a number of countries, notably in Europe, have developed and now maintain national travel models. Most of these models (see the reviews and studies reported by Gunn, 2001a; Lundqvist

& Mattsson, 2001; Zhang et al., 2010) include both passenger and freight components, and most combine estimates of short- and long-distance trip-making components.

(2) A number of states in the USA have developed, or are in the process of developing, their own long-distance travel models, seeking to capture travel across their borders as well as between their major metropolitan areas and counties (see FHWA, 1999; Horowitz, 2006).

(3) A third set of studies focus their attention on specific long-distance, high-volume travel corridors, with the most recent corridor studies in the USA and Canada focused on the analysis of high-speed rail feasibility (Bhat, 1995, 1997; Cambridge Systematics, 2006; Volpe Center, 2008), a topic of growing interest worldwide at the present time.

3.1 *Alternative Modelling Approaches*

We categorize multimodal inter-regional travel analysis methods into four groups (see Figure 1). All methods are capable of estimating multimodal OD matrices, and have produced operational models. They differ in whether or not, and in how travel behavioural responses to policy scenarios are considered.

3.1.1 *Direct demand and elasticity analysis*. A number of studies in Australia, Canada, Ireland, Spain, and the UK (Acutt & Dodgson, 1996; Bel, 1997; Domencich & Kraft, 1970; Oum & Gillen, 1983; Wardman, 1997; Worsley & Harris, 2001, among others) have adopted direct demand models. In a typical direct demand model, the aggregate passenger travel demand $\boldsymbol{D}$, between an OD pair by each transportation mode is expressed as a function of economic ($\boldsymbol{E}$), land use ($\boldsymbol{L}$), and socio-demographic characteristics ($\boldsymbol{S}$) of the origin and the destination. Transportation factors influencing the aggregate OD demand by a particular mode include the attributes of that transportation mode ($\boldsymbol{A}$, e.g. travel time, cost, other level of service factors) and its competing modes ($\boldsymbol{B}$) serving the same OD pair. Equation (1) summarizes the general direct demand model structure, where lower-case symbols indicate coefficients.

$$\boldsymbol{D} = f(\boldsymbol{eE}, \boldsymbol{lL}, \boldsymbol{sS}, \boldsymbol{aA}, \boldsymbol{bB}). \tag{1}$$

Typically, a Cobb–Douglas functional form is specified in the direct demand model (Miller, 2004). The coefficient estimates are direct indicators of constant

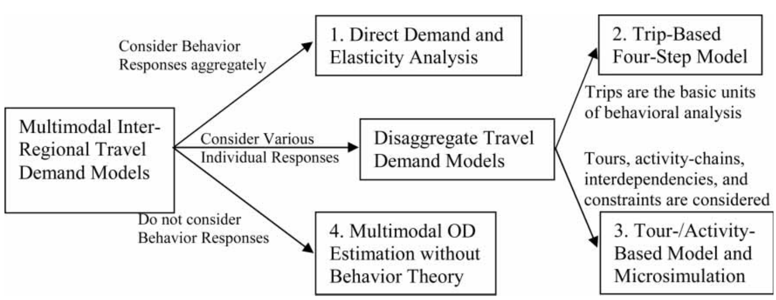

Figure 1. Categorization of multimodal inter-regional travel demand analysis methods.

own- or cross-demand elasticities. Analyses of these demand elasticities can provide direct policy implications. Direct demand models can also produce aggregate forecasts of multimodal travel demand for each OD pair by each mode, given alternative future growth and transportation system scenarios. The aggregate nature of direct demand models with its relatively low cost is appealing for national-level travel analysis. However, they do not take full advantage of the information contained in available travel survey data, and in theory more disaggregate methods based on individual behaviours can produce more accurate forecasts and provide models with improved policy sensitivities.

The concept of direct demand analysis can also be applied to the household and even personal levels. For instance, the dependent variable, $\boldsymbol{D}$, can be total household-level travel demand indicators. Different from disaggregate models, direct demand models, when applied at the household or individual level, do not consider different types of behavioural responses. Instead, they rely on statistical/econometric models to derive the relationship between aggregate travel demand indicators and other observed variables from existing data sources.

A good example of direct demand and elasticity analysis applied to national-level forecasting is the 1997 national road traffic forecast (NRTF) model in the UK (Worsley & Harris, 2001). Several groups of direct demand models are empirically estimated for NRTF, which use observed data to estimate vehicle ownership, vehicle use, truck traffic, transportation facility level-of-service, and traffic flows on different types of facilities. Based on the demand elasticity from these models, a hierarchical set of demand switching rules are then defined to analyse the full impact of specific policy scenarios on the road and transit networks. It should be noted that more recent national models in the UK have incorporated disaggregate modelling elements (Rohr et al., 2010).

The advantage of this approach lies in its relatively low development costs, reliance on available data, and provision of base-year multimodal OD matrices. OD matrices developed from these methods have also been routinely used to calibrate/validate national travel demand models developed with behavioural approaches. In terms of disadvantages, a direct OD matrix estimation model in itself is not sensitive to long-range policy alternatives due to its lack of behavioural sensitivities (Miller, 2004). For national studies and states, especially those with the need for a base-year OD matrix and with more short-term policy and planning goals (e.g. evaluating operational improvements), this method is a valuable analysis tool.

3.1.2 *Trip-based four-step method*. Of all the national travel demand models reviewed, the trip-based four-step approach is the most dominant methodology. Variants on the four-step approach have been employed in a number of national models in Europe, including those in Denmark (Fosgerau, 1998), Germany, Hungary, Italy (Lundqvist & Mattsson, 2001), Japan (Yao & Morikawa, 2003), the Netherlands (Daly, Fox, & Tuinenga, 2005; Gunn, Mijjer, Lindveldt, & Hofman, 1997), Norway, Sweden, Switzerland, the UK, and the USA (Ashiabor, Baik, & Trani, 2007; Baik et al., 2008; Cambridge Systematics, 2008) and in pan-European models (Davidson & Clarke, 2004; Gaudry, 2001; Leitham, Downing, Martino, & Fiorello, 1999; Nielsen, 2007). Quite a few states adopt similar methods in their statewide models as well, such as Indiana, Maryland, Massachusetts, and Virginia, among others. As the needs for analysing international travel have become increasingly important, several pan-European models have been developed, including the MYSTIC project (Peter Davidson Consultancy, 2000), the STREAM model (Williams, 2001), and the most recent TRANS-TOOLS project (Burgess et al., 2008).

Since the limited success of long-distance travel modelling in the USA in the 1970s (Weiner, 1976), few efforts were made prior to the late 1990s, except for a steady stream of academic studies on multimodal long-distance travel focusing on mode choice and vehicle ownership (Bhat, 1995; Koppelman & Sethi, 2005; Mannering, 1983; Winston, 1985, among others). However, the past decade has seen a revival of interest in not only statewide but also fully national-level transportation planning and policy analysis. In this last category, researchers at Virginia Tech (Ashiabor et al., 2007; Baik et al., 2008) have developed a four-step trip-based transportation systems analysis model (TSAM) for the entire country, using hybrid aggregate/disaggregate demand modelling framework (see Section 3.1.5). Cambridge Systematics (2008) has also conducted a trip-based framework study with staged process of a national travel demand model (this is not yet an operational model, however). And Epstein, Parker, Cummings, and Hammond (2008) have developed an agent-based microsimulation model of intercity travel for a study of pandemic diseases. The model employs a micro-level implementation of the gravity model to simulate individual travel decisions based on a zip-code-level OD system. While the model is capable of simulating trip-frequency and destination choices, there are no mode-choice or assignment steps. Nevertheless, this study exemplifies the benefit of an inter-regional model beyond transportation systems applications.

Table 1 provides an overview of selected models based on a trip-based approach. Section 4 of this paper provides a listing and discussion of the principal data sources used by individual models. The most advanced trip-based models for national travel analysis usually consist of the following modules, executed in a sequential manner with feedbacks between individual modules:

- Pre-processing (e.g. socio-economic, demographic, and vehicle ownership forecasts)
- Trip generation
- Trip distribution
- Model choice
- Time-of-day switching
- Traffic (route) assignment
- Post-processing (e.g. policy impact analysis, and emission estimation modules).

The traffic analysis zone systems in these models contain several hundred to almost 10 000 zones. Trip purposes typically are divided either into business, personal, and vacation travels or into categories based on trip ends and purposes (the latter is more common for countries with smaller geographic coverage and thus have relatively richer behavioural data for intercity travel in their national surveys). Transportation modes considered include car, bus, regular rail, high-speed rail, air, water, bike, and walk. A few models also include a time-of-day switching module developed from dedicated survey datasets. Traffic assignment methods range from static whole-day algorithms to multi-class multi-period stochastic equilibrium assignment. Feedback between the individual steps ranges from being nonexistent to fully integrated systems.

Table 1. An overview of selected long-distance travel demand studies

Model/study	Geographic detail	Modes	Trip purposes	Demand components and methodology	Explanatory variables	Data sources
USA						
TSAM (Ashiabor et al., 2007)	County level	Car, air, SATS (bus, rail)	Business/non-business	TG, TD, MC, TA. NL and mixed logit.	Travel time and travel cost (LOS), household (HH) income, region type	ATS 1995 for model calibration. In addition Virginia Tech conducted travel surveys
Koppelman (1990)	City/metro pairs (using data from NTS 1977)	Car, air, bus, rail	Business/non-business	TG, TD, MC, service class choice. Disaggregate NL model	LOS, departure frequency, distance between city pairs, HH income, structure, and size, employment	1977 NTS, supplemented with data on intercity level of service
Koppelman and Sethi (2005)	Only mode choice/service class choice from surveys	Car, air, rail sleeper, rail premium coach, rail economy sleeper	N/A	MC, service class choice. Multinomial logit (MNL) model, NL, and generalized NL	Cost, schedule convenience, overnight dummy, quality of service, group size, income, distance	SP surveys of 1000 rail users on-board making trips longer than 250 miles and 400 using other modes
Coldren, Koppelman, Kasturirangan, and Mukherjee (2003)	City pairs in the USA	Air	N/A	Itinerary share models. Aggregate MNL	LOS, connection quality, carrier, carrier market presence, fares, aircraft size and type, and time of day	Passenger booking data from computer reservation systems (CRS). Carrier schedule info from Official Airline Guide (OAG). Market size/fare data from the "Superset" data source (Data Base Products, Inc.).

(*Continued*)

Table 1. Continued

Model/study	Geographic detail	Modes	Trip purposes	Demand components and methodology	Explanatory variables	Data sources
Jin and Horowitz (2008)	Time-of-day choice modelling based on NHTS	Car, air, other	Work, return home, personal business, recreation	Time-of-day choice. MNL	LOS, travel companions, duration, age, gender, education, HH income, HH size, car ownership, presence of a child	2001 NHTS enhanced with the use of a preference survey that was conducted by email or by face-to-face interviews
Epstein et al. (2008)	Zip code pairs	N/A	N/A	TG, TD. Agent-based microsimulation (ABM)		
Individual state studies						
Oregon	2950 zones (instate and within a 50 mile radius). Each zone fits within about 14.5 million grid cells	Drive, carpool, urban transit, air, Amtrak, intercity bus, walk, bicycle	Home-based, work-based	TG, TD, MC, TA. Microsimulation (Monte Carlo) and logit models	Regional economics and demographics, production allocations and interactions, HH allocations, land development, commercial movements, HH travel, and transport supply	Household surveys, OD surveys, on-board surveys, specialized surveys
Michigan	2307 instate TAZs, 85 outstate TAZs	Car	HB work/biz, HB soc/rec/vac, HBO, NHB work/biz, NHB	TG, TD, TA. Used TransPlan and TransCAD	HH size, income, travel cost, area type	NPTS data used for calibration, CTPP for validation
Maryland	1607 zones (MD/DE/DC, and parts of NJ, PA, VA, and WV). 189 external zones	Car, air, rail, bus	HB work, journey to work, journey at work, school, HB shop, HB other	TG, TD, MC, TA. Gravity model and NL. A microsimulation technique is introduced for long-distance travel using the NHTS	LOS, demographics (population, income, occupation status, HH size, number of workers), and socioeconomics	Travel surveys, NHTS, CTPP, census

Corridor studies						
Cambridge Systematics (2006)	TAZs	Car, air, rail, HSR. Access egress (drive, drop-off, rental, taxi, transit, walk/bike)	Business, commute, recreation, other	TG, TD, MC, access/egress MC. Trip frequency MNL models	Employment and HH attributes, trip purpose/distance class, LOS, accessibility, region, travelling party size	A combination of new surveys and existing surveys from regional and state agencies
Volpe Center (2008)	County and MSA level	Car, air, existing rail and HSR, bus	Business/non-business	Direct demand modelling. Logit model	LOS, frequency, income	Base-year demand statistics from Amtrak, DOT Office of Aviation Analysis, 1995 OD data
Bhat (1995)	Corridor: Toronto–Montreal	Car, air, rail	Paid business	Mode choice. Heteroskedastic extreme value model using max likelihood technique	LOS, income, frequency, city type	Travel survey conducted in the corridor covering 2769 travellers
Bhat (1997)	Corridor: Toronto–Montreal	Car, air, rail	Paid business	Mode choice. Endogenous segmentation mode choice model	Income, sex, travel group size, day of travel, distance, service frequency, LOS, large city indicator	1989 Canadian intercity dataset from VIA Rail (3593 travellers)
European studies						
Dutch National Model System (LMS)	National. 1308 zones plus 55 external zones	Driver, carpool, train, bus/tram/metro, slow modes	Work, HB business, NHB business, shopping, education, other	TG, TD, MC, TA. Disaggregate tour frequency model	HH attributes (structure, license holding, car availability, income), sex, age, education, activity. Zonal socioeconomics and demographics, LOS	National Travel Survey (from 1995 with 68 000 HHs). Panel data. SP survey data

(*Continued*)

Table 1. Continued

Model/study	Geographic detail	Modes	Trip purposes	Demand components and methodology	Explanatory variables	Data sources
Great Britain long-distance transport model (Rohr et al., 2010)	146 county zones and 406 district zones	Driver, passenger, bus/coach, rail, air	Commute/education, business, visiting friends, leisure, and other	Tour frequency, mode, and destination choices. Nested choice structure	Person type, HH income, HH type, gender, LOS information for each mode	Long-distance component of National Travel Survey (NTS) in 2002–06 National Rail Travel Survey (NRTS) 2004–06
Italian Decision Support System (SISD)	267 national zones, 63 external zones	Car, bus, air, interregional rail, intercity rail, and sleeper rail	Commute, business, education, leisure and tourism, and other	TG, TD, MC, TA. Disaggregate tour frequency model	Sex, age, education, income, employment, license holding dummy, car ownership availability). Zonal socioeconomics and demographics, LOS, frequency	Interviews with 16 000 families, border-crossing interviews, traffic counts
Spain (Bel, 1997)	Spainsh rail network by province	Train, car	N/A	Double logarithmic form	Travel time, dummy variable for "increase in air service frequency"	1987 and 1991 operating data from train operator
Norway National Transport Model 4 (NTM 4)	454 domestic zones	Driver, passenger, transit, slow modes, air, sea	SD: HB Commute, HB business, education, WB business, shopping, social, recreation. LD: work, business, social, recreation, services and other	TG, TD, MC, TA. Disaggregate tour frequency model	Comparable and based on LMS	National Travel Survey (5800 households)

Swedish National Model System (SAMPERS)	700 zones, divided into 9000 subzones. 180 external zones	Car, bus, normal intercity train, X2000 high-speed train, and air	SD: work, business, school, social, recreation, other. LD: private, business	TG, TD, MC, TA. Disaggregate tour frequency model	Comparable and based on LMS	National Travel Survey (RiksRVU) 1994–98, and interviews from fixed link projects
Danish National Transport Model (PETRA)	3010 zones	Walk, cycle, car, car passenger, bus, and train	Home, work, errand, and leisure	TG, TD, MC, TA. Disaggregate tour-/activity-based model	Comparable and based on LMS	National Travel Survey (TU) 1995
BVWP (Austria)	676 domestic zones, 205 foreign zones	Car, train, coach/regional bus	Work, business, school, shopping, leisure, other	TG, TD, MC, TA. Aggregate trip frequency model		Household travel survey 1995–96, OD survey, traffic counts
VALIDATE (Germany), 2005	7000 zones	Motorized, public transit, and combined walking and cycling modes	Home, work, business, shopping, and other	TG, TD, MC, TA. Comparable and based on EVA algorithm		Mobility in Germany (MiG), 2002, Mobility in Cities (SrV), 2003
Switzerland National Travel Demand Model	2949 internal zones, 165 external zones	Motorized travel, public transit, and slow modes	Home, work, education, business, shopping, and leisure	TG, TD, MC, TA. Using EVA algorithm, which adopted growth-factor method for TG, and growth-factor method with multiple balancing factors for the joint MC/DC model		Swiss national travel survey (Mikrozensus)
MATISSE (France)	Links with OD distances varying from 50 to 2500 km	Car, air, rail	Business, private	TG, TD, MC, TA. Disaggregate trip frequency model	Travel time, cost, group size, time of day, car availability, fare reduction, quality of service	French household travel survey "Transports 1981–82"

(*Continued*)

Table 1. Continued

Model/study	Geographic detail	Modes	Trip purposes	Demand components and methodology	Explanatory variables	Data sources
STREAMS (EU)	201 EU zones, 27 zones for the other European countries, 4 for the rest of the world	Car, coach, slow train, high-speed rail, air, and other	SD: commuting-business, personal business-education, visiting. LD: commuting-business, personal business-education, visiting, domestic holiday, and international holiday	TG, TD, MC, TA. Aggregate trip frequency model	Age, employment, car availability, HH structure (aggregate average per distinguished population group), zonal employment, total population, tourism arrivals (bed spaces), gross value added	National travel surveys from seven EU countries, Eurobarometer survey (1998), tourism survey data from the World Tourism Organization (WTO)
DATELINE (EU)	NUTS 1-based zonal system of 90 internal zones and 30 external zones	Car, rail, air, and other		Direct OD estimation from multiple data sources, using synthesizing and matrix merging techniques		
STEMM (EU)	NUTS 3-based zonal system of 1269 zones	Air, rail, and car	Business, private, and vacation	TG, TD, MC quasi-direct demand method		National passenger travel surveys, border crossing survey data
TRANS-TOOLS (EU)	1269 zones	Air, rail, and car	Business/home-work, holiday, and other	TG, TD, MC, TA Four-step method using non-linear logit function	Travel cost, travel time, frequency, number of transfers, population, GDP, employment, car ownership	Pan-European household long-distance trip survey and ETIS-BASE database

Other non-US studies						
Yao and Morikawa (2003), Japan	6 zones from questionnaires, 147 zones from the NTS	Air, rail (conventional rail, Shinkansen), sea, bus, and car	Business and non-business	TG, TD, MC, route choice. Regression model and NL models with route choice	Accessibility, population, working population and its share, zonal GDP, business and non-business attractiveness. LOS, frequency, value of travel time savings	The model utilizes combined estimation across multiple data sources such as SP/RP surveys at six major rail stations, and aggregate data from the 2000 NTS
Aldian and Taylor (2003), Indonesia	Intercity Central Java. Number of zones unknown	Car only	N/A	TG, TD, TA. Fuzzy multi-criteria analysis, adopting disaggregate models with deterministic part of utility function. MNL model for TD	Population density, gross domestic regional product, road user cost (distance, road geometry, ride quality), number of hotel rooms	1996 national origin destination survey

Notes: Purpose: HB, home-based; SD, short distance; LD, long distance. Demand: TG, trip generation; TD, trip distribution; MC, modal choice; TA, traffic assignment; HH, household.

Certain models in Table 1 have departed from the traditional trip-based method, and incorporated elements from other approaches. Two prototypes states who have advanced their statewide models from a conventional four-step method are New Hampshire (employing tour-based components) and California (employing nested logit (NL) models based on disaggregate data sources). Some of these, arguably more advanced, methods, notably tour-/activity-based and microsimulation-based modelling approaches are discussed in the next section.

3.1.3 *Tour-/activity-based and microsimulation approaches*. More recent versions of several national models in the Europe recognize tours, trip chaining, and time-of-day dynamics on the demand side, and/or time-dependent congestion evolution on the supply side, signaling a trend of moving to tour-/activity-based microsimulation approaches. For instance, the most recent Dutch model (Fox, Daly, & Gunn, 2003) has replaced cross-classification and regression-based trip generation modules with tour-based procedures. In addition, it incorporates time-of-day switching propensities on the demand side. The Danish model (Fosgerau, 1998) considers three nested levels of travel representation: trips, tours, and chains (defined as a sequence of tours). The Italian model distinguishes three alternatives in the trip generation step for each trip purpose: not to travel, to make one tour, and to make two or more tours. Agent-based mobility simulation has been successfully conducted on the Switzerland national networks for congestion analysis. Within USA, Ohio and Oregon statewide models are the pioneers using tour-/activity-based microsimulation to account travel demand at the behavioural level (Giaimo & Schiffer, 2005).

A commonality among these advanced national demand models is their consideration of both short- and long-distance trips. While the analysis of short-distance daily travel can certainly benefit from these considerations of behavioural dynamics and interdependencies, the value of the activity-based approach for long-distance travel analysis in large geographies (e.g. the USA, the European Union) is not apparent and no such models have been developed. To be of value to long-distance travel analysis, this approach needs to be extended to multi-day out-of-home travel tours, linked to a traveller's household/family structure and business/employment practices.

We choose to present the Dutch Landelijk Model System (LMS; "Landelijk" is the Dutch for "national") in this section. First, its development started in 1983, and it is one of the earliest national models. Second, it is representative of the disaggregate modelling approach, and has actually been the prototype of several other national models. Third, the Dutch model has been updated several times with the development of both advanced methods and data. Finally, LMS in its present form represents a transition from trip-based to activity-based approaches and therefore provides a good example for the states that aim at improving their statewide models gradually.

The LMS adopts a disaggregate tour-based system built upon its original four-step framework, with stages of license holding, mode choice, and time-of-day decisions all linked with models of car-ownership, trip generation and distribution, and all based on analyses of individual choices (Gunn et al., 1997, 2001a; Hofman, 2001). The linkages among these choice dimensions are considered with a NL specification. From 1997 to 1999, the model was updated as well as the basic data sources. Additional improvements are currently underway

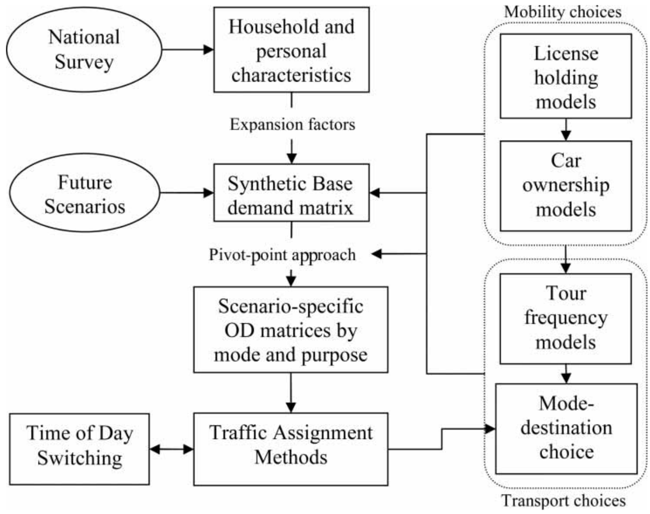

Figure 2. LMS model structure.

with longitudinal survey data. The zoning system has also been upgraded (Daly, 1998), which was shown to have greatly enhanced model accuracy.

Figure 2 illustrates the overall structure of the LMS. In the pre-processing mobility choice step, LMS employs a combination of license holding and car ownership models in the base year and under various policy scenarios. The tour-frequency models employ two interconnected modules (i.e. the 0/1+ module and the stop–go module) to estimate the total number of tours made by each individual. Tours are also segmented by travel purposes. The tour-frequency estimation utilizes information on the household structure, license holding, car ownership, occupation, gender, age, and education. The destination and mode choice modules are NL models. This step depends on the accessibility by each mode and on the level of attractiveness of each zone by travel purpose. A time-of-day switching module has also been developed based on stated-preference data. While a static capacity-constrained algorithm is used for traffic assignment in the LMS, the possibility of combining LMS demand modules with dynamic assignment has been considered in previous research (Ben-Akiva, Gunn, van Vuren, & Hofman, 1998; Gunn & Hofman, 1998). Congestion estimates are then fed back to time-of-day, and mode-destination choice models. Both national economy and land use are considered exogenous, which is also the case for almost all national and European models (the only exception is the recent TRANS-TOOLS European model wherein spatial computable general equilibrium economic models are integrated, Burgess et al., 2008).

Despite not being a completely activity-based model, the LMS is sensitive to many socio-economic, land use, transportation systems, and policy factors. Applications of the LMS include the forecast of rail demand for railway investment analysis, impact assessment of increased fuel prices, performance impact of improved roadway signalization, evaluation of roadway investment packages, analysis of high-speed trains, and estimation of the transportation of socio-economic, demographic, technological, and international driving forces (Hofman, 2001).

3.1.4 *Multimodal OD estimation without behaviour theory*. The aforementioned methods can be referred to as top-down approaches, because they all start with zonal socio-economic, demographic, and land use information, all require comprehensive demand models estimating behavioural responses, and all utilize data for calibration and validation of model parameters. However, OD matrices can also be estimated with a bottom-up approach directly from available data sources, including household surveys, user surveys on facilities, and link-level traffic counts. It is seldom seen in current US practices except some early statewide modelling attempts (e.g. Kentucky estimated the OD demand from traffic count data, Bostrom, 1998).

The European Commission has funded a stream of studies that aim to estimate multimodal OD matrices from available data sources, including the OD-ESTIM project (Hilferink, 1997; Cost-efficient *O*rigin-*D*estination *ESTIM*ator), the MYSTIC project (Peter Davidson Consultancy, 2000; *M*ethodolog**y** and evaluation framework for modelling pa*S*sengers and freigh*T* on transport *I*nfrastructure *S*cenarios), and the DATELINE project (Brog, Erl, & Schulze, 2004; Davidson & Clarke, 2004; *D*esign and *A*pplication of a *T*ransport survey for *L*ong-distance trips based on an *I*nternational *N*etwork of *E*xpertise). The MYSTIC team has developed a heuristic harmonization procedure to directly merge various data sources from seven EU countries into consistent pan-European OD matrices. The more recent DATELINE project has refined the MYSTIC methodology, and also incorporated data from a new pan-European long-distance travel survey involving 16 countries.

The DATELINE used *matrix builder* to count the numbers of journey records between zone pairs (Davidson & Clarke, 2004), which are then multiplied by expansion factors to get the estimated total journeys between OD pairs. Variance estimates are also established for these initial OD demand estimates based on expansion factors. In this *observed* OD matrix there are empty cells many of which are not true-zero cells. The *matrix synthesis* step first develops a gravity model based on non-empty cells in the observed matrix, and then applies this model to estimate the demand in all cells, which produces a *synthetic* matrix. Finally, the *matrix merge* step computes the final estimates of mode-specific OD demand as the variance-weighted sum of the observed and the synthetic matrices, ensuring that more reliable OD demand estimates from the observed or the synthetic matrices play a greater role in producing the final estimates.

Another interesting methodology for directly estimating OD matrices from data sources has been developed for the Dutch LMS model (Gunn et al., 1997). As described, the Dutch model analysed the policy scenarios based on the demand changes from a base-year OD matrix, which requires a procedure for updating this base matrix. The procedure takes the national household survey, OD data from intercepted travellers, rail ticket sales, and an *a-priori* matrix (a previous, but now outdated OD matrix) as inputs. The matrix estimation can be described as a statistical method that treats each observed data item as a piece of statistical evidence to be weighed against others based on their relative accuracy. This method is most useful when data for OD estimation originate from different sources, when the data items in these sources have different levels of accuracy or reliability, and when the data items follow different statistical distributions.

There is also a large-body of literature on estimating/updating OD demand from link-level traffic counts. This method can be operationalized with proportional assignment (Bell, 1991), linear programming (Nie, Zhang, & Recker, 2005; Sherali, Sivanandan, & Hobeika, 1994), or bi-level mathematical programming (i.e. Upper level: minimizing the discrepancies between observed link counts and the link counts implied by the estimated OD matrix; Lower level: traffic equilibrium conditions; Yang, Sasaki, Iida, & Asakura, 1992). Input data include an *a-priori* matrix and at least partial traffic counts on a significant number of links. This method should be of value to states and countries that have annual or daily traffic counts for a large portion of their transportation system (e.g. the Highway Performance Monitoring System in the USA). While this method could be computationally intensive on large-scale time-dependent networks, its applications on national networks for long-distance OD demand estimation may not require considerations for congestion evolution or time dependencies. There have been successful demonstrations of this method on large-scale networks such as the pan-European transportation network (Hansen, Nielsen, & Frederiksen, 2008; Nielsen, 1998).

3.1.5 *Mixed method and multi-step models*. Over the past three decades it has become an increasingly common practice to use hybrid aggregate/disaggregate demand modelling frameworks in transportation planning studies. These frameworks try to offer the best of both worlds: a behavioural basis for determining representative traveller utility functions and their associated travel cost elasticities, tied to a mechanism for expanding the resulting disaggregate model's results to match regional travel activity totals.

To date, for example, the most ambitious effort to construct a complete four-step long-distance transportation planning model for USA is attributable to researchers at the Virginia Polytechnic Institute and State University, whose TSAM uses the hybrid modelling method and produces estimates of annual long-distance trips by air and auto on a county-level basis (Baik et al., 2008).

With an initial focus on air travel, the TSAM starts with aggregate trip generation and distribution steps. The model uses purpose-specific trip rates multiplied by a set of exogenously supplied and household income stratified county population estimates (and forecasts) as the trip generation step. These Os and Ds are then distributed between county pairs using an aggregate spatial interaction model and an iterative proportional fitting routine to match county-level ODs to survey expanded, State-specific trip-making totals. The modal choice is solved as a two-step NL model, which assigns each OD flow to the air taxi, commercial airline, or automobile after first determining the average or "composite" cost of commercial air travel options by solving a logit model for travel between each OD pair's most common embarkation-debarkation airport pairings. Ashiabor et al. (2007) describe this modelling as well as the application of a mixed logit model to the same data. They also illustrate the use of door-to-door travel cost functions that incorporate airport access/egress as well as airport waiting time and flight delay costs, and the difficulties of getting accurate data on trip OD locations for this purpose from past surveys. Finally, traffic assignments for the commercial air travel are estimated using travel time and fare-based disutility functions to calibrate a multinomial logit model of alternative airport-to-airport route selections.

The TSAM framework exemplifies the effort required at the present time in combining a broad range of data sources and modelling techniques to obtain a set of spatially disaggregated long-distance Origin-Destination-Mode-Purpose travel matrices for the entire USA. Similar multi-stage and multi-sourced travel modelling frameworks are being used in the EU and elsewhere. Of these, the MYSTIC, STEMM, STREAMS, and TRANS-TOOL modelling systems listed in Table 1 have been applied on a continental scale in Europe. Also of note, these and a number of the more elaborate national travel models are also moving towards a merger of passenger and freight forecasting methods in order to capture a complete set of transportation sector activities as well as to assign mixed passenger-freight volumes to regional and national networks. Finally, some of the non-US modelling systems listed in Table 1 are beginning to explore feedback loops between the assignment stage and the effects of any congestion costs captured in this step on the generation, distribution, and mode choice.

Finally, a category of long-distance travellers not represented in US household surveys is foreign visitors and tourists. Limited analysis of the within-US travel activity patterns of these visitors appears to have been carried out. The TSAM discussed above does offer one beginning in this area, modelling international passenger enplanements at the nation's 66 international airports, using regression based on gross domestic product and historical enplanement data for nine world regions (Baik et al., 2008). While detailed air travel data on these travellers is collected from all of the commercial airlines making stops at US airports (refer to Table 4), data on how these travellers move around the country once they leave the air travel system is not collected, although data on the traveller's principal destination should be reported on their landing declaration.

Table 2. Qualitative comparison of multimodal inter-regional of analysis methods

Model properties	Direct demand	Trip-based four-step	Tour-/activity-based and microsimulation	Direct OD estimation
Behavioural foundation	Good	Good	Great	None
Transparency of the Model	Good	Average	Low	Good
Data requirement	Average	Average	High	Low-average
Development cost	Low-average	Average	High	Low
Development risk	Low	Low-average	High	Low
Maintenance/application cost	Low	Average	High	Low
Policy sensitivity	Average	Good	Great	None
Model properties	Direct demand	Trip-based four-step	Tour-/activity-based and microsimulation	Direct OD estimation
Short-term operational analysis?	Yes	Yes	Yes	Yes
Long-range planning analysis?	Yes	Yes	Yes	No
Produce base-year matrix	No	No	No	Yes

Note: Shaded cells represent properties that are desirable.

3.2 *Comparison of Alternative Analytical Methods*

Table 2 offers a qualitative comparison of the four methods discussed above. The four-step and the tour-/activity-based models enjoy sound behavioural foundations, and have the greatest capabilities for analysing operational and planning policy scenarios. However, these two methods are also the most costly options, involve relatively higher development risks, and have less model transparency (important for statewide models when the method needs to be explained to non-technical audiences). The direct OD estimation method is probably the cheapest and can produce base-year matrices for other methods, but has limitations for long-range policy analysis. The direction demand model is the average-performing alternative in most categories.

The choice of methodology is therefore a task that requires careful considerations based on funding resources, policy analysis needs, data availability and future data collection plans, and the value of increased forecast accuracy. We have observed staged development of travel demand models in current practices. For instance, the UK national modelling system has evolved from a single disaggregate auto-ownership module in the 1970s, to direct demand models in the 1980s, to hybrid (direct demand and discrete choice) NRTF system in 1997, and now towards more advanced behavioural models with even greater policy sensitivity. The pan-European travel demand model started with direct OD matrix estimation without behavioural models in the MYSTIC (Peter Davidson Consultancy, 2000) in the early 1990s, and gradually advanced to the 201-zone NUTS2 (Nomenclature of Territorial Units for Statistics version 2) STREAM (Williams, 2001) model with aggregate methods, then to the 1275-zone NUTS3 STEMM model (Gaudry, 2001) with joint aggregate–disaggregate models, and finally to the most recent disaggregate TRANS-TOOLS model that integrates European transportation and economic models (Burgess et al., 2008). Within USA, the Oregon statewide model represents a gradual transition from traditional four-step approach to the more advanced activity-based microsimulation structure (MW Consulting and PB Consult, Inc., 2002). Similar cases also include Michigan (Faussett, 2005) and California (Caltrans, 2009). These experiences may soundly inspire the other statewide models at their phased improvement programme.

3.3 *Other Methodological Issues*

3.3.1 *Identifying and estimating travel costs*. As both a significant determinant and important policy lever, accurate estimates of travel costs are an essential component of any travel demand model. To date there has been comparatively limited treatment of how these costs differ between long-distance travel and the rest of a household's trip making activity. A look at the "Explanatory Variables" column of Table 1 shows that monetary costs, travel times, and frequency of service (as a surrogate for inconvenience) are the most common LOS variables. For example, multi-day trips incur lodging costs as well as actual travel costs. Time constraints, and the time spent getting from one place to another, appear to be valued differently than they are for most daily trip activities (with some en-route activities, such as sight-seeing, having positive utility). Business travellers have different (generally much higher) time valuations (Gunn, 2001b; Wardman, 2001) than leisure travellers—although this assumption is often based on the commonly used notion that travel time can be evaluated in relation to pre-defined notions of the value of the specific activity being delayed or foregone. In reality, the reliability of a given model service or the immediacy of a good intermodal connection, can be very situation specific. Limited theoretical research and even less quantitative modelling have been done to establish how such time should be valued. New travel "cost" terms may need to be devised or cost functions allowed to vary according to the type of trip being made: its purpose, the number of days away from home, the number of travellers in the group, and other factors.

3.3.2 *Constructing "synthetic" OD matrices*. The more elaborate national and statewide modelling frameworks listed in Table 1 create a base-year set of OD passenger travel matrices by merging data from a number of different sources, using spatial interaction ("gravity") models or similar iterative proportional fitting routines to reconcile these data to known marginal activity totals, and in the process estimate the many missing OD data cells created by applying limited size data samples to spatially detailed traffic analysis zoning systems. The resulting matrices are a combination of observed and model generated flow estimates. Two lines of statistical modelling have been particularly useful:

- log-linear modelling of missing-value cells in large, multi-dimensional (travel demand) matrices (Agresti, 1990; Deming & Stephan, 1940; Goodman, 1971; Southworth & Peterson, 2005; Willekens, 1983; Wrigley, 1985), and
- the merging of observed OD data with observed traffic counts, to arrive at a set of flow estimates that come closest to satisfying both sets of input data, and suitably weighted by the confidence the analyst has in each data set (see, for example, Nielsen, 1998; Rios, Nozick, & Turnquist, 2003; Yang et al., 1992).

Both approaches offer some very appealing benefits to data programmes that are consistently and often severely limited in their data collection budgets. Their appeal lies not in any behavioural or explanatory content: but in their ability to generate detailed OD flow estimates that are reconcilable, to a greater or lesser degree, with observed planning level (i.e. aggregate) data totals from otherwise disparate data sources. For example, the current US multimodal national freight flows matrix is based on a log-linear modelling approach that uses maximum likelihood estimation coupled with iterative proportional fitting and other data fusion methods to estimate missing values in a 138 × 138 × 43 × 7 (over 5.7 million cell) origin-destination-commodity-mode matrix (Southworth & Peterson, 2005).

3.3.3 *Intermodal travel: access and egress options*. Many long-distance trips involve over one mode. This includes nearly all trips to and from the airport, and many also to and from rail and bus stations. The modes include most common private automobile, taxi, local rail or bus transit line, and regional shuttle bus service. In order to be able to compare travel costs by different line-haul modes these access, egress, and inter-modal transfer times and costs need to be accounted for. That is, a proper model comparison involves door-to-door travel analysis. Gosling (2008) provides a synthesis of airport access modelling, noting a trend away from aggregate and multinomial logit models towards the use of disaggregate, NL models based on airport passenger surveys. This includes NL models involving several levels of nesting and four or more market segments: with practice tending towards separate treatments, and different explanatory variables, associated with resident business trips, resident non-business trips, non-resident business trips, and non-resident non-business trips.

With a number of metropolitan areas, such as London and New York, operating multiple major airports, we also need to better understand the tradeoffs people make between airport access times and airport specific departure frequencies, alternative fares, and alternative route choices (i.e. direct versus two or more leg flights) in their selection of airports: and whether driving to an airport in another metropolitan area to get the desired ticket price or LOS is also part of their perceived choice set. This too has become a topic for a number of recent research efforts (see Hess, 2004, for example, who uses a mixed multinomial logit model, and Blackstone, Buck, & Hakim, 2006, who calibrate a probit model based on telephone survey data; and Gelhausen & Wilken, 2006, who calibrate a generalized NL model of combined airport and access mode choice). In contrast, the modelling of rail or bus station access modes and distances have been less popular subjects, though Nuzzolo, Crisalli, and Gangemi (2000) describe the calibration of a "nested-logit service/run/class/access–egress mode choice model" of medium–long-distance railway service. It is noticeable that in each case, a feature of these efforts has been a movement away from the comparatively simple formulation of the multinomial logit towards more involved discrete choice forms.

3.3.4 *Future year OD forecasts and mode shifts*. Various statistical methods have been used to forecast future travel volumes. These include a variety of time series-based methods appropriate to corridor planning, especially where historic data on mode-specific ridership is available from the carriers involved (FHWA, 1999). For longer-term forecasting of multiple OD flows within a region of state, it is usual to apply the same cross-sectional four-step, or direct demand model using base-year model parameter values applied to forecast year travel originations and attractions (destination totals). These Os and Ds are estimated using the trip-making rates derived from the base-year trip generation and attraction models but applied to forecast year economic activity totals which are themselves estimated by combining time series data with population cohort–survival and inter-industry input-output models (see Section 4.4 below for US sources of these forecasts). The most common method for updating demand forecasts using logit or similar discrete choice models is the "pivot point" or incremental logit approach (see FHWA, 1999 for a description of the method, and Hague Consulting Group, 2000 for its use in a number of European models).

3.4 *Assessment of Current Modelling Practice*

Miller (2004) provides an excellent summary of, and suggestions for, needed improvements in long-distance travel demand modelling that is still relevant today. The following list of modelling needs draws directly from his list, while adding to and commenting further on it:

- *Limitations on OD and trip purpose details*: This is the greatest weakness of all efforts to model long-distance travel, with the limited sample size of passenger and household surveys preventing expansion of estimates on a sound statistical basis to rather broad regional OD matrices, and in many cases also to rather broad trip purpose categories.

- *Treatment of access and egress modes*: The true effects on access/egress mode availability and user-perceived costs need to be better captured in both our datasets and demand models. Miller (2003) points to the use of NL modelling as one means being used to capture such costs in a theoretically consistent manner. All OD travel costs should be "door-to-door" costs. If multi-destination tours are modelled these costs should be put on home-back-to-home tour cost basis.
- *Treatment of travel costs and LOS attributes*: Extensions of traveller disutility or generalized cost functions are needed that go beyond the "fare, time, and service frequency" approach. These cost functions should be allowed to vary according to the types of trips or tours being made: by trip purpose, by number of days away from home, and by number of travellers in the group, etc.
- *Alternatives to discrete choice modelling*: To date, many of the choices simulated by microsimulation/ABM methods still rely heavily on the partial travel choice probabilities generated by logit or similar discrete choice, disaggregate demand models. In the future, alternative rule-based choice systems might also be explored, taking advantage of the less restrictive functional forms these methods make possible. Support for such methods will, however, require supporting data collection efforts, including more in-depth study of how travellers make long-distance travel decisions.
- *Making traffic congestion endogenous to the modelling process*: The effects of increased traffic volumes on congestion-induced delays need to be modelled explicitly if policy analysis is to place reliance on a national or regional model's ability to evaluate the effects on traveller benefits of adding or removing significant modal capacity. Feedback from the assignment to the other steps in the four-step modelling process is one way to do this. Other less computationally intensive ways also need to be explored.
- *Alternatives to a trip-based approach to behavioural response*: While the number of trips between places is an important planning input, the behavioural basis for generating these volumes needs to be tied closer to the daily and seasonal activity patterns of travellers who often organize their long-distance travel activities in the form of multi-destination out-of-home trip tours. Household characteristics need more attention here, notably where leisure trips are concerned.
- *Foreign Visitor Trips:* More attention needs to be given to modelling the travel activity schedules and destinations of foreign visitors, principally those of foreign tourists.

4. Data on Long-Distance Travel Activity

4.1 *Current Data Sources Supporting Long-Distance Passenger Travel Analysis*

Many of the shortcomings of current models, however, are closely tied to the limitations of existing datasets. Much past "travel modelling" has in fact been focused on filling gaps in current data sources or on finding ways to cope with limitations on the travel as well as traveller details provided by past household surveys. A variety of data collection methods are being experimented with currently (see Southworth & Hu, 2010), including:

- Household and personal travel surveys;
- Comprehensive, multimodal surveys of all travel taken by persons or households;

- Passenger carrier customer surveys (including data collection tied to mandatory government reporting requirements);
- User interception surveys on roadsides, transit terminals, airports, and borders;
- Special-purpose stated preference (SP) and SP-RP (reveal preference) surveys;
- Tourism data;
- Dedicated long-distance travel surveys;
- Transportation network link-based vehicle or vessel traffic counts, and
- Vehicle or vessel en-route tracking data.

Table 3 summarizes the type, coverage, frequency, collection method, and quality of various datasets used in selected national and European travel demand models. The primary demand-side data sources are cross-sectional household and personal travel surveys, conducted with various methods including mail surveys, telephone interviews, and in-person interviews. Most of the national surveys are not designed specifically for long-distance travel analysis because of a country's size. They involve 15 000–70 000 households. The recent pan-European survey, which covers a geographic region similar to the USA, includes about 870 000 person samples and only considers medium- and long-distance travel. Almost all survey data are cross-sectional or repeated cross-sectional, though panel data exist in the Netherlands for selected years. Typically, data are collected annually (e.g. Swedish SAMPERS used RiksRVU national travel survey collected annually from 1994 to 1998, see Beser & Algers, 2001; Sveder, 2001), once every 5 years (Switzerland and Japan conducted national surveys every 5 years, see Swiss Federal Statistical Office, 2006; Yao & Morikawa, 2003), or just one time (most pan-European studies and some national studies such as Danish PETRA model were based on one-time cross-sectional surveys). There are also cases (e.g. Italy) where data are collected twice in a year to account for seasonal demand variation (probably due to interests in tourism trends). One-day diary is the most popular method, with a 7-day diary adopted in the recent UK survey. In countries where repeated cross-sectional surveys have been conducted, the quality of data in different years is often inconsistent due to changes in the sampling framework, recruitment methods, non-response handling, and questionnaire design. Oversampling, special SP surveys, and modal-specific data collection are often employed to supplement and improve the quality of national survey datasets. Section 3.4 describes common limitations of the various data sources, and presents tested strategies that can address data quality and completeness issues.

4.2 *US Data Sources Supporting Long-Distance Passenger Travel Analysis*

Table 4 lists the principal sources of data on long-distance travel activity within, into, and out of the USA, organized by type of data collection and primary mode of transportation (air, auto, bus, and rail). These are for the most part federally supported data sources, with coverage of nationwide travel activity, albeit at different levels of spatial (i.e. OD) resolution. Of the modal (carrier)-specific data sources, only the data on air travel provide OD flow totals, of which the 10% sample of air passengers also allows for statistically robust disaggregation of passenger traits by OD corridor. However, as shown in Table 4, no single database pieces together all of these datasets to obtain a single passenger trip volumes matrix by airport OD pair and trip purpose. It is also worth noting that none of

Table 3. Data sources used in current national and pan-European travel demand models

The models	Primary data sources	Survey period	Data type	Data coverage range	Collecting method	Comments
Dutch National Model System (LMS)	Netherlands National Travel Survey (OVG), Special SP surveys	1985-present	Repeated cross-sectional used. Panel available	10 000–68 000 households Supplemented by several hundred SP surveys	Computer Assisted Telephone Interview (CATI), and a 1-day travel diary	
Great Britain (NTM)	(1) National Travel Survey (NTS)	(1) 1988-present	Repeated cross-sectional	5000–15 000 households	Home interview, and a 7-day travel diary	
	(2) National Rail Travel Survey (NRTS)	(2) 2004–06				
	(3) Household Interview	(3) 2009				
Italian Decision Support System (SISD)	(1) Household-based survey	(1) 7/94 and 4/95	(1) Twice in 1 year	(1) 8500 households in summer, 10 000 in winter	(1) Household telephone interview,	Surveys were held in summer and winter separately to capture seasonal variation
	(2) Border-crossing interviews	(2) 7/94 and 3/95	(2) Twice in 1 year	(2) 16 000 interviews in summer, 12 000 in winter	(2) Border face-to-face interview,	
	(3) Manual traffic counts	(3) 7/94 and 3/95	(3) Twice in 1 year	(3) 138 traffic counts	(3) Bidirectional traffic counts	
Swedish National Model System (SAMPERS)	National Swedish Travel Survey (RiksRVU)	1994–98	Repeated cross-sectional	30 000 personal interviews	CATI, and a 1-day travel diary	
Danish National Transport Model (PETRA)	National Travel Survey (TU)	1995	One year cross-sectional	13 793 personal interviews	CATI, and a 1-day travel diary	One year cross-sectional data was insufficient to produce variation over time

German National Travel Demand Model (validate)	(1) Mobility in Germany (MiG) (2) Mobility in Cities (SrV)	(1) 2002 (2) 2003	(1) One-year cross-sectional (2) One-year cross-sectional	(1) 49 000 households (2) 34 000 persons	(1) CATI, and a 1-day travel diary (2) N/A	
Swiss National Travel Demand Model	Swiss National Travel Survey (Mikrozensus)	2000	Cross-sectional, collected every 5 years	27 918 households	CATI	
STREAMS	National travel surveys, border crossing, roadside, tourism	Mostly 1994	Cross-sectional	7 EU countries: Denmark, Finland, France, Germany, the Netherlands, Sweden, and the UK		
Japanese integrated intercity travel demand model	(1) The Inter-regional Travel Survey	(1) 2000	(1) cross-sectional, collected every 5 years		(1) Approximately 500 000 passengers	(1) Separate one weekday sample interview taken for five inter-regional mode systems
(2) SP data	The latest survey in 2005 includes weekend days	(2) 2000	(2) 1 year	(2) 2000 interviews in total	(2) interview at six main rail stations on the HSR line	
DATELINE	Pan-European Household Long-Distance Trip Survey	2000–02	Cross-sectional	86 969 persons in 16 European counties; over-sampling on very long-distance trips	Combined postal, telephone, and in-person interviews with a two-phase design	Special journey-tour-trip design for long-distance survey

Table 4. US national and international person travel activity data sources

Database name and agency source	Travel modes covered	Type of data collection	Currency, frequency, and size of data collection effort	Known limitations for flows modelling
Department of Transportation (DOT)/ BTS ATS	All modes, including intermodal	P	National sample of long-distance travel (less than 100 miles one way) by US households, including foreign trips originating in the USA, door-to-door travel, and all trip purposes	Sample size only large enough to support a set of state-to-state and 55 largest MSA-to-MSA OD flows. Not repeated since 1995
DOT/FHWA NHTS and NPTS	All modes	P	National sample of daily trip-making by US households. All trip purposes. Rich in data on traveller characteristics. Trip length data and data by metro/suburban/non-metro area classes. Data collected in 2001–02 and 2008–09. The 2008–09 survey sampled some 155 000 households. Data collected by predecessor NPTS is available for 1969, 1977, 1983, 1990, and 1995	Data on long-distance trips (over 50 miles) is less reliable than short distance trip-making data. There are not enough data to construct OD tables
DOC/Tourism Administration Canadian Travel Report (CTR)	All modes	P	Data on number of Canadian tourists to USA, by US State visited. Data on trip purposes and modes, by travel season, State visited, stay duration, lodging type and activities, spending, and by residence of traveller. US Customs data from reports by individual States. Data has been collected since 1979	OD details only by State visited
DOT/BTS Office of Airline Statistics Form 41 T-100 Non-Stop Segment & On-Flight Market Data for Commercial Carriers	Air	C, V	National sample of all domestic and foreign flights with a stop in the USA. Passengers transported, scheduled departures and OD enplanements and debarkations recorded. Traffic volumes on each airport-to-airport flight leg. Seats available, aircraft type and service class data is recorded, with a complete sample of airport to airport traffic volumes on a monthly basis	Air only, no data on ground access or egress travel (i.e. ODs are airport-to-airport only). The database does not include foreign flights with less than 60 seats

DOT/BTS Office of Airline Statistics Passenger Origin-Destination (OD) data	Air	C, V	Full trip itinerary data collected via a 10% continuous ticket sample. Domestic and international inbound and outbound flights are included. Ticket origin, destination and all en-route airport stops are identified, ticket fares are reported, as is mileage for each trip segment.	Routings are classified as unidirectional by applying arbitrary trip-breaking rules to itinerary data.
DOT/BTS Office of Airline Statistics Commuter Air Carrier Statistics	Air	C, V	National sample of lights and passengers transported between city pairs on scheduled commuter services. Quarterly reporting of aircraft miles, hours, departures, and revenue-passenger miles and available seat-miles	Point-to-point airport to airport data only, no data on true ODs
Department of Commerce (DOC)/ Office of Trade and Tourism Industries (OTTI) International Air Travel Statistics (INS I-92 and I-94 data)	Air	C, P	Data on all international flights entering or departing the USA. The data are reported as monthly, quarterly, and annual statistics	Does not contain data on flights to/from Canada. Point-to-point airport-to-airport data only, no data on true ODs
DOC/OTTI In-Flight Survey of International Air Travellers	Air	P	Annual statistics on travellers to and from the USA. Data on traveller characteristics and spending patterns, including data on traveller's residence, destinations visited, domestic transport used, demographics and length of stay	
DOT/Federal Aviation Administration (FAA) General Aviation/Air Taxi Activity and Avionics Survey	Air	C	Sample survey of flight hours, airframe hours and, number of landings, and the state where the aircraft is based.	No OD data
DOT/FAA Airport Carrier Activity Information System (ACAIS)	Air	C	Data on annual number of enplanements by scheduled and non-scheduled flights out of US airports, by airport Hub size	

(*Continued*)

Table 4. Continued

Database name and agency source	Travel modes covered	Type of data collection	Currency, frequency, and size of data collection effort	Known limitations for flows modelling
DOT/FHWA Highway Performance Monitoring System (HPMS)	Highway	V	Data on average annual daily traffic counts submitted annually by States. A time series of annual data from 1978 exists, with reporting through FHWA's *Highway Statistics* publication.	Auto, bus, motorcycle, and single unit 4-tire truck count data, no OD data. Sample coverage of non-Interstates is variable and difficult to expand to national totals.
DOT US-Canada and US-Mexico Border-Crossings Data	Highway, Rail, and Pedestrian	V	Data are for people and vehicles entering the USA from Canada or Mexico. Data are collected monthly by the US Customs Service. Annual data are summarized at port of entry level, starting in 1997	No data on traffic exiting USA into Canada or Mexico
Amtrak Intercity Rail Data	Rail	C	Nationwide intercity and long-distance annual ridership by advertized Amtrak routes. Data availability at discretion of Amtrak	Not OD data. Strict limits on data made available to the public

Notes: C, carrier survey based; P, household or passenger survey based; V, vehicle/vessel based.

these datasets address the issue of airport access and egress connections. For modal cost/modal choice comparisons, as well as LOS studies, there is therefore a need to develop a set of true origin-to-true destination "air-inclusive" passenger flow matrices and an associated matrix of the multimodal journey costs. The publicly available data on long-distance rail and bus (motorcoach) travel, as compiled by Amtrak and American Bus Association, respectively, contain limited OD details at the present time.

The largest data gap, however, is in automobile travel. Since the American travel survey (ATS) of 1995, very little data on the details of long-distance trip-making has been collected, other than travel by air, making it difficult to establish just how much this travel has grown over the past 15 years or how the geographic pattern of such movements, notably travel by automobile, has changed. As a result, there is no single database or a data analysis programme that can provide robust statistical information on where, by which modes, and how often people are moving around the nation. Nor can it effectively assess the response of travellers in the aggregate to changes in travel costs and how these costs affect trip frequencies, trip distances (and hence destination selections), and choice of modes. By maintaining mode specific datasets that are rarely combined, it is also a difficult and resource-intensive activity to compare modal travel options with existing data sources. Rather, this information must be pieced together from a variety of sources with little or no consistency or coordination in the data collection methods being used. A well constructed, statistically sound and multimodal long-distance passenger travel database is needed prior to the development of a nationwide multimodal long-distance travel demand model of the USA.

4.3 *Some Areas for Data Improvement*

It is clear that significant gaps exist in our knowledge about long-distance travel in America, with data on air travel proving the most comprehensive, and data on private automobile travel the least satisfactory in terms of either its national activity totals or its ability to generate detailed origin-to-destination flow estimates. Areas where significant improvement would be beneficial include the following:

Household survey data needs: Lack of up-to-date and statistically robust data on both how and why US travellers choose different long-distance modes, destinations, and routes is the single biggest problem currently facing policy analysis. Many of the current model shortcomings listed at the end of Section 3 can be linked directly or indirectly to poor or insufficient data coverage. Limited sample size means a limited ability to break trip-making up into more behaviourally sensible trip purpose classes, while spatial, notably OD, detail is either constrained to broad regional and annual volumes, or must be created using synthetic methods involving a mix of different data types as well as sources.

Data on access and egress mode choice: Data also needs to be collected on how. as well as how often, people access regional air, rail and bus travel options. Methods are needed for combining this data with information gathered from other data sources, such as trip purpose data collected by public transit agencies (e.g. "travel to airport", or "travel to rail station" responses).

Highway travel control totals: Given that the majority of long-distance trips that households make are by private automobile, and the desire for metro-area-to-area, or better, OD detail, regionally based travel activity control totals are likely to be needed, in addition to those provided by a sample-expanded household travel survey. Innovative ways to use traffic count data need to be explored as a means of at least validating the travel activity totals derived from sample size-constrained household surveys. It is conceivable that eventually, these electronically collected data sources may become a, if not the, principal source of much of our travel data needs.

Bus and rail data: While detailed OD data exist for both bus and rail travel in America, this information is for the most part held as proprietary, with only summary activity totals generally released to the public. Some arrangement to gain access to this sort of data in a more detailed form is worth exploring, recognizing the sensitive nature of such data requests, perhaps offering benefits in terms of data fusion products that might benefit the data supplying agencies involved.

Multimodal data synthesis needs: Methods are needed for combining existing datasets in ways that get the most out of current data sources. For the creation of OD matrices, this means using the latest synthetic flow matrix generation techniques, using such methods as iterative proportional fitting to create OD flow estimates that strike an acceptable balance between carrier-reported, household survey-expanded, and (eventually) traffic monitoring based-projections of travel activity totals.

Arguably, such additions and improvements to our current data sources should more than pay for themselves by providing greater accuracy in present day assessments of long-distance travel activity as well as by providing a more robust basis for understanding, and forecasting, future long-distance travel needs.

5. Conclusions and Recommendations

5.1 *Some Possible Model Development Pathways*

Based on our synthesis of methodological options and data sources for multimodal long-distance passenger travel analysis, a number of different modelling avenues appear worth exploring. If we consider "Available Data Sources" as our origin, and "Long-Distance Passenger Travel Demand Model" as our destination or goal, Figure 3 presents some alternative modelling "routes" available to us. Note that this is not an all-inclusive overview of potential routes. Each route in the figure represents a unique approach towards the development of a statewide or national long-distance passenger travel model. Each node (i.e. box) represents an intermediate product, and node IDs (i.e. A through G) do not necessarily imply sequencing. For instance, a model development path could be ACFDG. In this case, a base matrix is first estimated from the available data sources, which are both used for the developing disaggregate travel models (e.g. trip frequency, destination, model, and route choices). These disaggregate models are then linked together to form a trip-based four-step model.

The long-distance travel modelling method most suitable for a particular country should depend on available data sources, resources for new data collection, resources for model development, policy analysis needs, and possibly other factors. We recommend several guidelines under loosely defined resource availability (for data collection and model development) and policy analysis needs scenarios respectively (see Table 5). The selection of modelling methods

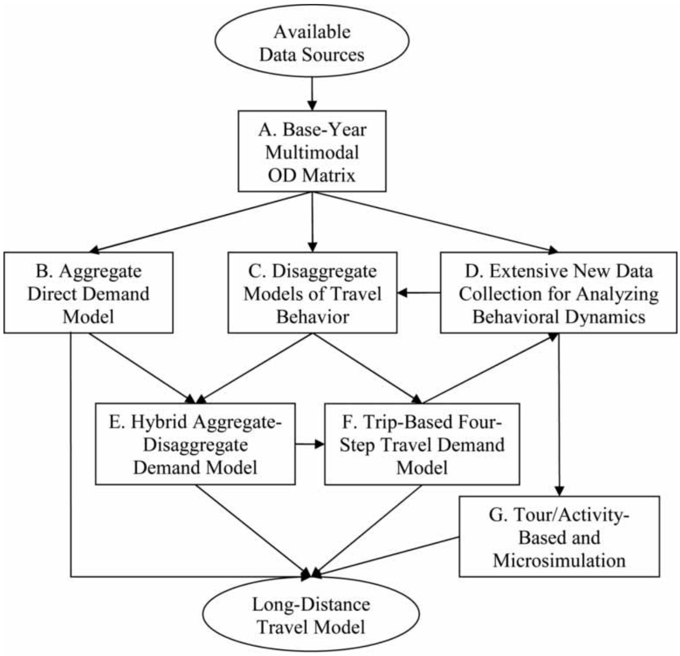

Figure 3. Alternative approaches to long-distance passenger travel demand modelling.

Table 5. Selecting the most suitable modelling approach

	Low needs	Medium needs	High needs
Low resource	AB	ABE	Need more resource
Medium resource	ACE	ACF	ACFDG
High resource	Unlikely	ADCF	ADG

may also be based on acceptable model development risks. For instance, the most aggressive and highest-risk path would be ADG, while the lowest-risk incremental approach would be ABEFDG.

5.2 *Future Research on Long-Distance Passenger Travel Data Collection*

Experiences accumulated from the ATS, nationwide household travel survey (NHTS)/nationwide personal transportation survey (NPTS), and recent long-distance travel surveys in Europe, as well as recent studies on emerging technologies for travel surveys, have revealed a few key technical challenges in the effective design of a long-distance travel survey instrument.

The advantage of employing traditional household survey methods for long-distance trip data collection is that they can gather most information required for travel analysis and modelling. But it also has disadvantages including relatively high cost, respondent burden, and data reporting and measurement errors. Since the frequency of long-distance travel is relatively low for the majority of the households, it is often difficult and costly to obtain a sufficiently large sample of long-distance travel records with traditional random sampling methods.

Survey methods enabled by emerging technology (e.g. web-based, GPS, License Plate, cell/Smartphone, Bluetooth, social media) can overcome several limitations of the traditional household-based survey method, but may not provide all required long-distance travel information. For instance, GPS, Smartphone, Bluetooth, and License Plate technologies can directly provide a wealth of information on long-distance passenger travel, but not all information (e.g. trip purpose and traveller characteristics). Data collection based on social networking websites could encounter sampling bias issues. Our review on long-distance passenger travel data sources suggests several promising future research directions:

- Further testing the feasibility of emerging technological options for long-distance passenger travel data collection;
- Exploring event-prompted recall surveys and advanced data imputation algorithms that can effectively gather/estimate long-distance travel information (e.g. trip purpose, mode) to supplement data directly collected from GPS, Smartphone, License Plate, and/or Bluetooth technologies;
- Developing practical methods and procedures that employ advanced technologies to identify long-distance trips, subsequently re-identify long-distance travellers, and finally conduct follow-up long-distance travel surveys on these travellers, which could provide an effective solution to various sampling problems in long-distance travel surveys; and
- Researching into the transferability of long-distance models developed with various data sources for different countries, including the transferability of model specification, coefficients, mathematical methods, and behaviour postulates.

Acknowledgement

This research was supported in part by the US Department of Transportation (USDOT), Federal Highway Administration (FHWA). The opinions in this document do not necessarily reflect the official views of USDOT and FHWA. USDOT and FHWA assume no liability for the content or use of this document. The authors are solely responsible for all statements in this paper.

References

Acutt, M., & Dodgson, J. (1996). Policy instruments and greenhouse gas emissions from transport in the UK. *Fiscal Studies*, *17*(2), 65–82.

Agresti, A. (1990). *Categorical data analysis*. New York, NY: Wiley.

Aldian, A., & Taylor, M.A.P. (2003). Fuzzy multicriteria analysis for inter-city travel demand modelling. *Journal of the Eastern Asia Society for Transportation Studies*, *5,* 1294–1307.

Ashiabor, S., Baik, H., & Trani, A. (2007). Logit models to forecast nationwide intercity travel demand in the United States. *Journal of the Transportation Research Record*, *2007,* 1–12.

Baik, H., Trani, A., Hinze, N., Swingle, H., Ashiabor, S., & Seshadri, A. (2008). Forecasting model for air taxi, commercial airline, and automobile demand in the United States. *Journal of the Transportation Research Record*, *2052,* 9–20.

Bel, G. (1997). Changes in travel time across modes and its impact on the demand for inter-urban rail travel. *Transportation Research*, *33E,* 43–52.

Bell, M. (1991). The estimation of origin–destination matrices by constrained generalized least squares. *Transportation Research*, *B25,* 13–22.

Ben-Akiva, M., Gunn, H.F., van Vuren, T., & Hofman, F. (1998). *The AIUTO project, implementation of a time dependent assignment at the Randstad test site*. Presented at the 8th World Conference on Transport Research, Antwerpen.

Beser, M., & Algers, S. (2001). SAMPERS—The new Swedish national travel demand forecasting tool. In L. Lundqvist & L.-G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 101–118). Stockholm: Swedish Transport and Communications Research Board.

Bhat, C.R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological*, *29*(6), 471–483.

Bhat, C.R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, *31*(1), 34–48.

Blackstone, E.A., Buck, A.J., & Hakim, S. (2006). Determinants of airport choice in a multi-airport region. *Atlantic Economic Journal*, *34,* 313–326.

Bostrom, R. (1998). *The Kentucky statewide transportation focusing process*. Irvine, CA: Kentucky Transportation Cabinet.

Brog, W., Erl, E., & Schulze, B. (2004). *DATELINE: Concept and methodology*. Paper presented at the 2004 European Transport Conference, Strasbourg, France.

Bureau of Transportation Statistics. (1997). *1995 American travel survey*. Washington, DC: US Department of Transportation. Retrieved February 11, 2009, from http://www.bts.gov/publications/1995_american_travel_survey/

Burgess, A., Chen, T.M., Snelder, M., Schneekloth, N., Korzhenevych, A., Szimba, E., … Christidis, P. (2008). *TRANS-TOOLS (TOOLS for TRansport forecasting ANd Scenario testing)* (Final Report, Deliverable 6, Funded by the 6th Framework RTD Programme). Delft, The Netherlands: TNO Inro.

Caltrans. (2009). *Statewide model framework*. Sacramento, CA: California Department of Transportation.

Cambridge Systematics. (2006). *Bay area/California high-speed rail ridership and revenue forecasting study. Interregional model system development*. Cambridge, MA: Author.

Cambridge Systematics. (2008). *National travel demand forecasting model phase I final scope*. NCHRP Project 08-36, Task 70, National Cooperative Highway Research Program. Washington, DC: Transportation Research Board.

Cohen, H., Horowitz, A., & Pendyala, R. (2008). *Forecasting statewide freight toolkit*. Washington, DC: Transportation Research Board.

Coldren, G.M., Koppelman, F.S., Kasturirangan, K., & Mukherjee, A. (2003). Modeling aggregate air travel itinerary shares: Logit model development at a major US airline. *Journal of Air Transport Management*, *9*(6), 361–369.

Daly, A. (1998). Updating and extending national models. In L. Lundqvist & L.-G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 35–49). Stockholm: Swedish Transport and Communications Research Board.

Daly, A., Fox, J., & Tuinenga, J.G. (2005). *Pivot-point procedures in practical travel demand forecasting*. Paper presented at European Regional Science Association meeting, Amsterdam.

Davidson, P., & Clarke, P. (2004). *Preparation of OD matrices for DATELINE*. Paper presented at the 2004 European Transport Conference, Strasbourg, France.

Deming, W.E., & Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, *11,* 427–444.

Domencich, T.A., & Kraft, G. (1970). *Free transit*. Lexington, MA: Heath Lexington Books.

Epstein, J.M., Parker, J., Cummings, D., & Hammond, R.A. (2008). Coupled contagion dynamics of fear and disease: Mathematical and computational explorations. *PLoS ONE*, *3*(12), e3955. doi:10.1371/journal.pone.0003955.

Faussett, K.M. (2005). *MDOT's statewide household travel survey—design, implementation, and lessons learned*. Unpublished paper. Retrieved March 29, 2010, from http://www.morpace.com/TRB-Conference-2.PDF

FHWA. (1999). *Guidebook on statewide travel forecasting*. Washington, DC: Federal Highway Administration. Retrieved February 9, 2009, from http://www.fhwa.dot.gov/planning/statewide/swtravel.pdf

FHWA. (2011). *Freight analysis framework, version 3*. Washington, DC: Federal Highway Administration. Retrieved April 3, 2011, from http://www.cta-gis.ornl.gov/faf/

Fosgerau, M. (1998). *PETRA—an activity-based approach to travel demand analysis*. Paper presented at the 8th World Conference on Transport Research, Antwerp, Belgium.

Fox, J., Daly, A., & Gunn, H. (2003). *Review of RAND Europe's transport demand model systems*. Leiden, The Netherlands: RAND Corp.

Gaudry, M. (2001). Test of nonlinearity, modal captivity and spatial competition within the STEMM multicountry applications for passengers. In L. Lundqvist & L.-G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 170–187). Stockholm: Swedish Transport and Communications Research Board.

Gelhausen, M.C., & Wilken, D. (2006). Airport and access mode choice: A generalized nested logit model approach. Munich Personal RePEc Archive. Paper No. 11957. Retrieved March 29, 2010, from http://www.mpra.ub.uni-muenchen.de/11957/2/MPRA_paper_11957.pdf

Giaimo, G.T., & Schiffer, R. (Eds.). (2005, August). Statewide travel demand modeling: A peer exchange. *Transportation Research Circular, #E-C075*. Washington, DC: Transportation Research Board.

Goodman, L.A. (1971). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, *13*, 33–61.

Gosling, G.D. (2008). *Airport ground access mode choice models. A synthesis of airport practice. ACRP synthesis 5. Airport Cooperative Research Program*. Washington, DC: Transportation Research Board.

Gunn, H.F. (2001a). An overview of european national models. In L. Lundqvist & L.G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 19–34). New York, NY: Springer.

Gunn, H.F. (2001b). Spatial and temporal transferability of relationships between travel demand, trip cost and travel time. *Transport Research Part E*, *37*, 163–189.

Gunn, H.F., Mijjer, P.H., Lindveldt, K., & Hofman, F. (1997). Estimating base matrices: The combined calibration method. In *Proceedings of the PTRC Summer Annual Meeting*, London, UK.

Hague Consulting Group (HCG), et al. (2000). EXPEDITE. EXPert-system based PrEdictions of Demand for Internal Transport in Europe. Review of European and national passenger and freight market forecasting systems. Prepared for the European Commission Directorate-General for Energy & Transport, Brussels, Belgium.

Hansen, S., Nielsen, O.A., & Frederiksen, R.D. (2008). *Adjustment of trips-matrices in large-scale networks*. Paper presented at the 2008 European Transport Conference, Leeuwenhorst, The Netherlands.

Hess, S. (2004). *An analysis of airport-choice behaviour using the mixed multinomial logit model*. Retrieved February 9, 2009, from http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa04/PDF/61.pdf

Hilferink, P. (1997). *Cost-efficient origin-destination estimator: OD-ESTIM* (Final Summary Report). Transport Research & Innovation Portal (TRIP). Directorate General for Mobility and Transport, European Commission.

Hofman, F. (2001). Application areas for the Dutch national model. In L. Lundqvist & L.-G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 81–92). Stockholm: Swedish Transport and Communications Research Board.

Horowitz, A.J. (2006). *Statewide travel forecasting models* (NCHRP Synthesis #358). Washington, DC: Transportation Research Board.

Horowitz, A.J. (2008). *White paper: Statewide travel demand forecasting*. NCHRP Project 08-36, Task 76c, National Cooperative Highway Research Program. Washington, DC: Transportation Research Board.

Jin, X., & Horowitz, A.J. (2008). Time-of-day choice modeling for long-distance trips. *Transportation Research Record*, *2076*, 200–208.

Koppelman, F.S. (1990). Multidimensional model system for intercity travel choice behavior. *Transportation Research Record*, *1241*, 1–8.

Koppelman, F.S., & Sethi, V. (2005). Incorporating variance and covariance heterogeneity in the generalized nested logit model: An application to modeling long distance travel choice behavior. *Transportation Research Part B: Methodological*, *39*(9), 825–853.

Leitham, S., Downing, J., Martino, A., & Fiorello, D. (1999). European transport forecasts for 2020: The STREAMS model results. In Proceedings of the ETC '99, Cambridge.

Lundqvist, L., & Mattsson, L.G. (Eds.). (2001). *National transport models: Recent developments and prospects*. New York, NY: Springer.

Mannering, F.L. (1983). An econometric analysis of vihicle use in multivehicle households. *Transportation Research Part A: General*, *17*(3), 183–189.

Miller, E.J. (2003). Microsimulation. In K.G. Goulias (Ed.), *Transportation systems planning: Methods and applications* (Chap. 12, pp. 12-1–12-22). Boca Raton, FL: CRC Press.

Miller, E.J. (2004). The trouble with intercity travel demand models. *Transportation Research Record*, *1895*, 94–101.

MW Consulting and PB Consult, Inc. (2002). *The Oregon Modeling Improvement Program: An Overview.* Salem, OR: Oregon Department of Transportation.

Nie, Y., Zhang, H.M., & Recker, W.W. (2005). Inferring origin-destination trip matrices with a decoupled GLS path flow estimator. *Transportation Research*, *39B*, 497–518.

Nielsen, O.A. (1998). Two new methods for estimating trip matrices from traffic counts. In H.D. Ortúzar, D. Hensher, & D. Jara-Díaz (Eds.), *Chapter in travel behaviour research: Updating the state of play* (pp. 221–250). Oxford: Elsevier Science Ltd.

Nielsen, O.A. (2007). *European TRANS-TOOLS transport model course*. Lyngby, Denmark: Technical University of Denmark. Retrieved November 16, 2009, from ftp://ftp.jrc.es/users/transtools/public/Documentation/

Nuzzolo, A., Crisalli, U., & Gangemi, F. (2000). A behavioural choice model for the evaluation of railway supply and pricing policies. *Transportation Research A*, *34*, 395–404.

Oum, T.H., & Gillen, D.W. (1983). The structure of intercity travel demands in Canada: Theory tests and empirical results. *Transportation Research Part B: Methodological*, *17*(3), 175–191.

Peter Davidson Consultancy. (2000). *MYSTIC toward origin-destination matrices for Europe*. London: Author.

Rios, A., Nozick, L.K., & Turnquist, M.A. (2003). Value of different categories of information in estimating freight origin-destination tables. *Transportation Research Record*, *1783*, 42–48.

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S., & Tsang, F. (2010). *Modelling long-distance travel in the UK*. Paper presented at the 2010 European Transport Conference, Glasgow, Scotland, UK.

Sherali, H.D., Sivanandan, R., & Hobeika, A.G. (1994). A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes. *Transportation Research B*, *28*, 213–233.

Souleyrette, R.R., Hans, Z.N., & Pathak, S. (1996, November). *Statewide transportation planning model and methodology development program*. Ames: Iowa State University.

Southworth, F., & Hu, P.S. (2010). *The American long distance personal travel data and modeling program: A roadmap. Revised draft*. Prepared for the Federal Highway Administration, Washington, DC.

Southworth, F., & Peterson, B.E. (2005). *Generation of a US commodity flows matrix using log-linear modeling and iterative proportional fitting*. Prepared for the Office of Operations, Freight Management and Operations, Federal Highway Administration, Washington, DC. Retrieved February 9, 2009, from http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/faf2_reports/report3/index.htm

Sveder, G.J. (2001). The new Swedish national model—SAMPERS: System and validation. In L. Lundqvist & L.-G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 93–100). Stockholm: Swedish Transport and Communications Research Board.

Swiss Federal Statistical Office. (2006). *Ergebnisse des Mikrozensus 2005 zum Verkehrsverhalten*. Neuchatel: Swiss Federal Statistical Office.

Volpe Center. (2008). Evaluation of high-speed rail options in the Macon-Atlanta-Greenville-Charlotte rail corridor. Volpe National Transportation Systems Center Cambridge, MA. Prepared for the Georgia Department of Transportation, Atlanta, GA.

Wardman, M. (1997). Inter-urban rail demand, elasticities, and competition in Great Britain: Evidence from direct demand models. *Transportation Research*, *33E*, 15–28.

Wardman, M. (2001). A review of British evidence on time and service quality valuation. *Transportation Research Part E*, *37*(18), 107–128.

Weiner, E. (1976). Assessing national urban transportation policy alternatives. *Transportation Research*, *10*, 159–178.

Willekens, F.J. (1983). Log-linear modeling of spatial interaction. *Papers of the Regional Science Association*, *52*, 87–205.

Williams, I. (2001). Designing the STREAMS model of Europe. In L. Lundqvist & L.-G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 159–169). Stockholm: Swedish Transport and Communications Research Board.

Winston, C. (1985). Research on intercity freight and passenger transportation; an economist's perspective. *Transportation Research Part A: General*, *19*(5–6), 491–494.

Worsley, T.E., & Harris, R.C.E. (2001). GB traffic forecasts—status and development. In L. Lundqvist & L.-G. Mattsson (Eds.), *National transport models: Recent developments and prospects* (pp. 146–155). Stockholm: Swedish Transport and Communications Research Board.

Wrigley, N. (1985). *Categorical data analysis for geographers and environmental scientists*. Caldwell, NJ: The Blackburn Press.

Yang, H., Sasaki, T., Iida, Y., & Asakura, Y. (1992). Estimation of origin–destination matrices from link traffic counts on congested networks. *Transportation Research B*, *26*, 417–434.

Yao, E., & Morikawa, T. (2003). *A study on integrated intercity travel demand model*. Paper presented at the 10th International Conference on Travel Behavior Research, Lucerne.

Zhang, L., Xiong, C., & Berger, K. (2010). Multimodal inter-regional origin-destination demand estimation: A review of methodologies and their applicability to national-level travel analysis in the US WCTR. Lisbon, Portugal, July, 2010.