Transportation
DOI 10.1007/s11116-013-9497-y

Hubris or humility? Accuracy issues for the next 50 years of travel demand modeling

David T. Hartgen

Abstract This study reviews the 50-year history of travel demand forecasting models, concentrating on their accuracy and relevance for public decision-making. Only a few studies of model accuracy have been performed, but they find that the likely inaccuracy in the 20-year forecast of major road projects is ±30 % at minimum, with some estimates as high as ±40–50 % over even shorter time horizons. There is a significant tendency to overestimate traffic and underestimate costs, particularly for toll roads. Forecasts of transit costs and ridership are even more uncertain and also significantly optimistic. The greatest knowledge gap in US travel demand modeling is the unknown accuracy of US urban road traffic forecasts. Modeling weaknesses leading to these problems (non-behavioral content, inaccuracy of inputs and key assumptions, policy insensitivity, and excessive complexity) are identified. In addition, the institutional and political environments that encourage optimism bias and low risk assessment in forecasts are also reviewed. Major institutional factors, particularly low local funding matches and competitive grants, confound scenario modeling efforts and dampen the hope that technical modeling improvements alone can improve forecasting accuracy. The fundamental problems are not technical but institutional: high non-local funding shares for large projects warp local perceptions of project benefit versus costs, leading to both input errors and political pressure to fund projects. To deal with these issues, the paper outlines two different approaches. The first, termed 'hubris', proposes a multi-decade effort to substantially improve model forecasting accuracy over time by monitoring performance and improving data, methods and understanding of travel, but also by deliberately modifying the institutional arrangements that lead to optimism bias. The second, termed 'humility', proposes to openly quantify and recognize the inherent uncertainty in travel demand forecasts and deliberately reduce their influence on project decision-making. However to be successful either approach would require monitoring and reporting accuracy, standards for modeling and forecasting, greater

D. T. Hartgen (✉)
University of North Carolina at Charlotte, Charlotte, NC, USA
e-mail: david@hartgengroup.net

model transparency, educational initiatives, coordinated research, strengthened ethics and reduction of non-local funding ratios so that localities have more at stake.

Keywords Accuracy · Travel demand · Forecast · Uncertainty · Optimism bias · Ethics

"All [traffic forecasting] models are wrong; by how much determines their usefulness." George Box.
"The future isn't what it used to be." Yogi Berra.
"Pay no attention to that man behind the curtain!" The Wizard of Oz.

Introduction

Martin Richards, longtime Editor-in-Chief for *Transportation*, has kindly asked me to contribute to this issue and I am most pleased to do so. Under Martin's 41 years of leadership *Transportation* has actively participated in virtually all of the issues and advancements mentioned below, and without his direction it is unlikely that *Transportation* would have achieved its current status. To him we all therefore owe our heartfelt thanks and our applause, and I must say that personally it has been a great honor to serve with him. As *Transportation* moves into its next era, I am confident that those who follow will build on his foundation and that *Transportation* will continue as a central point of focus for thoughtful discussion of the issues that confront us.

The topic of this study is travel demand modeling and forecasting, particularly what needs to be improved regarding modeling accuracy. My thesis advisor Martin Wachs once said, *"Never put a number and a date in the same sentence"*. Of course I proceeded to do exactly that for the next 45 years, carrying both to at least four digits of precision. Sometimes, standard practice encouraged it, sometimes the boss or client, and sometimes hubris. In perhaps a thousand cases varying from site-specific studies and intersection turning movements to broad country-level studies, I have willingly participated in this most basic analysis. In short, I'm a travel demand model-holic.

During the first 25 years of my career there was occasional soul-searching about travel demand forecasting methods and how to improve them, but little discussion about the institutional contexts surrounding them. Of course, there were periodic reviews of model weaknesses (e.g., Transportation Research Board 1973; Stopher and Meyburg 1976), but most in this business know that those are just the surface of much deeper problems that fundamentally challenge our knowledge and our procedures—and which we have only gingerly discussed. I am speaking here of the ethics of forecasting, including the potential for biased assessment, misrepresentation, advocacy, collusion and possibly even fraud.

These are strong words, not to be used lightly in any profession. So this will not be the usual *Transportation* paper with hypothesis, data, model coefficients and interpretation. Instead, it will be a 'frank and honest' (to use the language of diplomacy) assessment of various issues. Many, if not most, of my colleagues will disagree with my views; I ask only for thoughtful and constructive response.

Most of my discussion applies to passenger travel demand modeling, particularly estimating road traffic volumes and transit ridership, and mostly in urbanized areas. While one could make similar observations regarding freight modeling, the additional complexity of that topic and its even greater dearth of paradigms put it beyond my scope.

Uncertainty in forecasting

The primary reasons to undertake travel demand forecasts for proposed major facilities (new or expanded roads, new transit lines, new bus routes, etc.) are to: 'size' the facility in terms of capacity (width, lanes, buses, trains, seats etc.) per unit time; estimate the cost of project development; estimate revenue for toll-based projects; provide information for facility design such as pavements and supporting features such as nearby intersections; estimate socioeconomic and environmental impacts; and evaluate project costs versus benefits. This predict-and-provide method is intended to ensure that the built facility will have sufficient capacity to operate at the desired level of service over its intended lifespan, that its benefits will outweigh its costs, and that its impacts will be manageable.

The standard method of estimating travel demand for major proposals in urban areas, and for many proposals between urban regions, is the so-called 4-step method. The approach breaks the forecast problem into four computational stages: trip generation, distribution, mode split and assignment. For instance, the 4-step estimate for a forecast of average daily traffic (ADT) in year y on a proposed highway is:

$$
\begin{aligned}
ADT_y = \sum_{o,d,l} & (zonal\,origin\,and\,destination\,populations,\ employment,\ wealth,\ etc)_y \\
& *\ trip\,rates_y \qquad [Trip\,generation,\ step\,1] \\
& *\ distributed\,trips\,between\,O\,and\,D_y \qquad [Trip\,distribution,\ step\,2] \\
& *\ share\,using\,private\,vehicles_y \qquad [Mode\,choice,\ step\,3] \\
& *\ share\,using\,a\,particular\,road\,link\,l\,in\,year\,y. \qquad [Assignment,\ step\,4]
\end{aligned}
$$

The forecast traffic is then used in further computations to estimate impacts and design features. For instance, the number of (directional) lanes required to carry the predicted road traffic in year y is:

$$
\begin{aligned}
Lanes_y = & \ Next\,Highest\,Integer\,of: \\
& Average\,Daily\,Traffic_y * directional\,factor_y * peak\,hour\,factor_y * design\,hour\,factor_y \\
& / desirable\,level - of - service\,per - lane\,service\,volume_y
\end{aligned}
$$

Typically, each of the above terms is treated as deterministic, with a single value given the circumstances (e.g., Highway Capacity Manual 2010). But each term is also a function of many other factors and is highly variable in time and space. And each of these background factors is also is dependent on numerous additional assumptions, even though in most model applications they are treated as deterministic numbers not subject to variation or uncertainty. Usually, few calculations are made to determine range, variation or likelihood of occurrence of any of these terms. In road planning the problem is sometimes worked backwards, estimating the ADT needed to justify 4, 6, or 8 lanes, using typical values for the needed parameters (e.g., Florida Department of Transportation 2010). In the US most state department of transportation have similar methods for preliminary route planning. For more detailed corridor planning additional methods are also employed, but they also typically do not account for uncertainty.

There are several fundamental problems with this modeling structure. From an accuracy perspective, the method produces results that are increasingly viewed as inaccurate, perhaps 'just plain wrong', significantly biased toward over-statement, and not accurate enough for use as the basis for decisions involving large expenditures.

European and Australian experience

The magnitude and technical sources of these inaccuracies has been researched for some time. Beginning in the late 1970s the UK Department for Transport (United Kingdom) initiated studies to determine the accuracy of road traffic forecasting models. By 1988, the review had found 41 road projects that met their forecasts (i.e., actual traffic was within 20 % of predicted traffic), and 27 projects that had greater differences (Department for Transport (United Kingdom) 1988). Technical causes of these inaccuracies were attributed to overestimates of diversion to the new facility, underestimates of national traffic growth, failure to account for induced traffic, variations in regional growth, and subsequent changes in land use. Mackie and Preston (1998) identified numerous sources of error and bias in transport modeling, grouping them into those related to project objectives, definitions of alternatives, data errors, model structure and appraisal optimism. The Department for Transport (1998) began to revise its method of forecasting national road traffic in the mid 1990's, modifying the prior high-and-low forecasts (variations of GDP and fuel price, but without probability of occurrence) by adding a central fitted-on forecast intended to represent the most likely (highest probability of occurrence) outcome. By 2008 the Department for Transport (2008) had become more specific, requiring in its guidance for project development that forecasts of road traffic must include a 'core scenario' growth forecast and alternative scenarios above and below, with "*range of uncertainty equal to 2.5 %, times the square root of the number of years beyond the base year.*" However, there appears to be some question regarding how much this guidance is actually followed. The Department has recently expanded the comparative research to 55 projects, finding that about 90 % of forecasts were within 43 % of actual traffic (Department for Transport 2013).

The guidance further specifies that an Uncertainty Log be developed to enumerate and assess the likelihood of various events related to the tested scenarios, such as proposed developments, land use plans or zoning proposals, and infrastructure improvements. These findings are to be thoroughly reported in planning documents. Table 1 provides an outline of the classifications to be used in the Uncertainty Log.

A recent study by the European Court of Auditors (2013) investigated the accuracy of traffic and cost forecasts for 24 road investment projects in Germany, Greece, Poland and Spain. These projects were all non-toll, and included bypasses, road widenings, new 2-and 4-lane segments, new bridges, and upgraded 2-lane roads. Costs ranged from €5 million to over €650 million. The study found that while the projects partly delivered their intended purposes, assessment based on cost-benefit was impossible, and that "*Most of the audited projects were affected by inaccurate traffic forecasts*." Actual costs came in an average of 23 % higher than predicted. For the 19 projects with traffic forecasts, 5 were within 20 % of the actual traffic, 11 were between 20 and 50 % off, and 3 were greater than 50 % off. On average, the actual traffic was 15 % below the forecast traffic. However, these were mostly 1 year after opening counts.

Another national review of toll road traffic forecasts is on-going in Australia. In a widely cited paper, Li and Hensher (2010) reviewed the traffic forecasting accuracy of 14 major toll Australian toll routes in Sydney, Melbourne and Brisbane. They found overestimates averaging 40 % for five toll roads in the Sydney area. Using regression models relating the percentage error to various characteristics, they attributed most of the error to less toll road capacity (when opened, compared with forecast), elapsed time of operation (roads opened longer had higher traffic levels), time of construction (longer construction time delayed traffic growth and increased the error), toll road length (shorter roads

Table 1 Outline for the Uncertainty Log (DfT 2008)

Probability of the input	Status
Near certain: the outcome will happen or there is a high probability that it will happen.	Intent announced by proponent to regulatory agencies.
	Approved development proposals.
	Projects under construction.
More than likely: the outcome is likely to happen but there is some uncertainty.	Submission of planning or consent application imminent.
	Development application within the consent process.
Reasonably foreseeable: the outcome may happen, but there is significant uncertainty.	Identified within a development plan.
	Not directly associated with the transport strategy/scheme, but may occur if the strategy/scheme is implemented.
	Development conditional upon the transport strategy/scheme proceeding.
	Or, a committed policy goal, subject to tests (e.g. of deliverability) whose outcomes are subject to significant uncertainty.
Hypothetical: there is considerable uncertainty whether the outcome will ever happen.	Conjecture based upon currently available information.
	Discussed on a conceptual basis.
	One of a number of possible inputs in an initial consultation process.
	Or, a policy aspiration.

attracted less traffic), cash payment (modern no-cash payment increased traffic), and fixed/distance based tolling (fixed tolls reduced traffic). Li and Hensher's re-analysis of the US toll road data (see below) yielded findings similar to that of Australia. In a recent review of seven Australian toll roads opened since 2005, Bain (2013) found that in all seven, the counted traffic volumes were lower than the predicted volumes by 40–60 %. Reviews for Sweden, Norway (Welde and Odeck 2011), Spain (Vassallo 2007), the Netherlands (de Jong et al. 2007) and PPP projects worldwide (World Bank 2008) have similar findings. The Australian government (Australian Government 2012) subsequently cited "*inaccurate and over-optimistic*" traffic forecasts as a threat to investor confidence. Three lawsuits now underway challenge the forecasts for toll road traffic that subsequently came in significantly under projections (Bain 2013).

US experience

In contrast, in the US the focus has been largely on the accuracy of transit ridership forecasts. In the first major US review, the Federal Transit Administration (FTA) examined the accuracy of ridership and cost forecasts for 10 major fixed-guideway transit projects built between 1971 and 1987 (Pickrell 1989, 1992). The study found that transit ridership was substantially overestimated and costs substantially underestimated for all 10 projects. Ridership forecasts were over actual ridership by an average of 257 %, and costs were low by an average of 61 %. FTA updated the study in 2003 using an additional 19 cases; they found modest improvements in cost forecasting since the 1989 report, but still systematically underestimating actual costs by 17 % on average. Ridership forecasts had also improved since the 1989 study. In 2009 FTA updated the study again, this time with 18 fixed-guideway projects completed between 2003 and 2007 (FTA 2009, Session 2). Actual construction costs still exceeded the inflation-adjusted estimates developed in alternatives analysis by 40 %, and the actual ridership averaged just 75 % of predicted ridership. Curiously, FTA concluded that the cost forecasting had improved, but that ridership forecasting accuracy had not.

FTA does not have formal published guidelines for forecasting. Instead, FTA refers New Start sponsors to its recent workshops and draft policy Guidance (FTA 2009, 2013). The website and draft Guidance describe two basic methods of estimating fixed-guideway transit ridership: (1) a locally-based method using either a conventional 4-step model, an activity-based model, or a simplified data-driven incremental mode choice model of changes in travel times and costs for those trips that are within the influence area of a proposed project; and (2) a simplified national model called STOPS (Simplified Trips-on-Project Software) that (after completion of testing) sponsors will be able to use to prepare ridership forecasts for FTA New and Small Starts ratings purposes. The pre-cursor to STOPS was the ARRF (Aggregate Rail Ridership Forecasting) model that has been used by agencies for several years as a quality control comparison point to forecasts prepared from other methods. Both ARRF and STOPS are based on the experiences from regions with completed transit projects, and relates guideway ridership to census journey-to-work data, but STOPS is calibrated further with both BRT and rail projects and includes GTFS (General Transit Feed Specification) networks and peak period zone-to-zone auto travel times as inputs. Also, both methods encourage estimating opening-year patronage (FTA 2013), rather than focusing solely on long-range future estimates.

FTA also optionally permits (but does not require) proposals for New Start projects to address uncertainty in the forecasts of model inputs and other factors, suggesting a range of ridership forecasts (lower, most likely, upper). FTA also strongly recommends (but does not require) an 'incremental build-up' of the forecast for the locally preferred alternative, as well as a discussion of the specific sources of uncertainty ('upside' and 'downside') and their likelihood (FTA 2009, Session 10), similar to the Uncertainty Log used in the UK. FTA also requires before-after and predicted-actual comparisons of ridership.

However, the range of uncertainty may be too narrow. Reviewing 80 alternative total-travel scenarios from 50 large US regions, Bartholomew (2007) found that the on average the estimates of predicted regional vehicle-miles of travel varied only slightly (on average, about −2.3 % for an 11 % increase in regional density) from baseline forecasts, suggesting that the tests were not extreme. Whether transit sensitivity tests are broad enough is unknown.

Looking primarily at technical model accuracy, Chatterjee et al. (1997) reviewed the capability of traditional 4-step modeling techniques to provide increasingly detailed estimates of mobile source emissions resulting from transportation development and operations. They found that, in practice, some steps in the traditional modeling sequence (e.g., trip generation, distribution, and assignment) tended to decrease relative uncertainty (because they accumulate numerous estimates over a large numbers of cases), while others (particularly mode choice) tended to increase it (because they make estimates of small proportions). Nevertheless, the range of uncertainty in traditional traffic model forecasts was estimated to be on the order of 15–30 % for a major-road 20-year forecast, far wider than the detail needed for estimating emissions variations likely to be needed for air quality conformity tests, but considerably smaller than that estimated for toll roads, discussed below. The authors suggest ways in which current practice and analysis tools can be improved to increase the accuracy of their results. Bain (2011) reports an even wider variation for UK and other nation's projects, noting that "*at the 90 % confidence level, 5-year traffic forecasts for new roads are likely to have an accuracy of* ± *27.5 %*". In his review of US urban resurgence between 1970 and 2000, Wachs (2013) notes the failure of academics to see how transportation and urban resurgence are related, even though by 1975 virtually every large city in the US had a travel demand model and was using it for forecasting.

In a theoretical assessment of error propagation, Zhao and Kockelman (2002) used a simulation-driven 4-step model applied to the Dallas region to trace the uncertainty error through the 4-step process. They found that "*modeling error in effect grows* [from an average of ±30 % prior to generation to ±42 % in mode choice] *through the application of trip generation, trip distribution, and mode choice models… However, the final step of traffic assignment enjoys [an improvement] in uncertainty* [to an average of ±30 %] *(at the link-flow level), thanks to overlap in different trips' routings and mode and trip distribution choices across all travelers, along with congestion feedbacks (which moderate the presence of high link-demand values).*"

However, in contrast to the experience of other nations, only a very limited number of empirical studies have been undertaken regarding US road traffic forecasting accuracy. The US's National Cooperative Highway Research Program (NCHRP 2006) undertook a review of forecasting errors for 15 US toll roads opened between 1986 and 1999. They found, on average, that the actual traffic was 35 % below the predicted traffic. For revenue, only 3 of 26 toll roads reviewed were above revenue projections after 1 year, and only three were still above projections after 5 years. Referring to the study, Naess et al. (2006) found that 13 of 14 US toll roads had over-predicted traffic by an average of 42 %. The key factors affecting the accuracy of toll road forecasts were identified as demographic forecasts, travel characteristics, tolling culture (local familiarity), value of time and willingness to pay, personal preferences and reliability, ramp-up period, truck use, time choice modeling, model validation, peer review, risk assessment, and optimism bias. NCHRP concluded that "*the most successful [toll road traffic] forecasts generally had accurate or even conservative economic forecasts with moderate anticipated growth levels. These toll roads were built in corridors that were fully developed and where congestion already existed*" reflecting Bain's work.

In another very recent development, the New York Supreme Court (2013) permitted to go to trial a case involving alleged fraudulent action to conceal 'success fee' payments to a traffic forecasting company, and thus optimistic traffic forecasts, if five toll roads were ultimately purchased by an investor group. To the author's knowledge, this is the first lawsuit in the US that concerns the accuracy of traffic forecasts.

However, unlike the EU study, no US studies have empirically evaluated the accuracy of travel demand forecasts for *un-tolled* new roads, or for other types of actions such as road widenings which are commonly proposed in metropolitan regions. That issue focuses directly on the accuracy of travel demand models for toll-free proposals. Spielberg et al. (2007) reviewed the status of travel demand forecasting methods in US metropolitan areas, to identify "*shortcomings in travel forecasting models, obstacles to better practice, and actions needed to ensure the use of appropriate technical approaches*". They found that while the 4-step methods were generally adequate for analyzing major investment proposals, the 4-step methods did not adequately address the range of emerging policy options now being discussed, nor for financial analysis such as is needed for toll road evaluation. But they did not quantify the degree of uncertainty in travel demand forecasts. The study identified the primary weaknesses of current travel demand forecasting methods as lacking behavioral content, policy insensitiveness, and lacking dynamic conditions. This study recommended substantial funding increases for improved modeling (to $ 20 m annually), pooled fiscal resources from urbanized areas, and increased attention to activity modeling.

The report did not, however, call for new paradigms of travel or propose international efforts, nor did it address broader methodological or process issues.

The US federal government guidance for road traffic forecasting (Federal Highway Administration 2010, p. 23) has identified the major "*substantial uncertainties*" in traffic forecasts as *"population and employment forecasts, housing trends and costs, global and local economic conditions, other planned transportation improvements, time-of-day assumptions, parking prices, fuel prices, and long-term changes in vehicle technology"*. To deal with this form of uncertainty, the Guidance recommends an "*incremental buildup*" of the forecast, adding each element sequentially to the baseline traffic model until the full future forecast is made. This approach provides just one forecast for each alternative, and does not address the issue of uncertainty or inherent variation in each of the components of the forecast.

Recently the National Cooperative Highway Research Program (Donnelly et al. 2010) reviewed the advances in travel demand modeling, looking at activity-tour models, dynamic modeling, integrated land use, and freight. Regarding risk and uncertainty, the recommended approach is to conduct what-if scenarios that vary the most significant input assumptions about land use and system performance. The study did not address institutional issues such as non-local shares for major projects, lengthy planning timelines, regulations tying modeling to legal requirements, or optimism bias. NCHRP has also recently initiated a comprehensive assessment of very large and long-range strategic issues that might affect US transportation. The study is still underway, but the first report (NCHRP 2010) focused on broad trends like aging population, increased global trade, climate regulation, protectionism, and sustainability. The study has yet to assign probabilities to various events, let alone suggest how they might be accounted for in travel demand modeling. A third stream of recent research has focused on methods to incorporate uncertainty of travel time and the value of travel time into traffic modeling (Strategic Highway Research Program II 2013a, b). This research also recommends the use of scenarios (optimistic, average, and pessimistic) to deal with uncertainties in inputs.

More recently Chow et al. (2012) have noted the increased fracturing of travel demand modeling and its strained efforts to address emerging issues. They call for a broader educational structure based on fundamental flow principles rather than modal or topical issues, but do not address ethical or institutional issues. Work is also progressing on such issues as probabilistic estimates of travel time delays. For instance, Jaggi and Axhausen (2013) investigate how to incorporate dynamic loading by hour, variation within hour, and variation in capacity. They propose a probabilistic method of estimating travel time losses (delays) for each hour of the year, for different road classes, using the chance of a breakdown in traffic flow, as an extension of familiar flow-speed curves.

To summarize the current US situation, no systematic numerical assessment of accuracy in highway traffic forecasts has yet been conducted in the US. The Federal Highway Administration's Travel Model Reasonableness Checking Manual (Travel Model Improvement Program 2010) concentrates not on accuracy in forecasting but on model validation, that is testing to see if model results fit existing traffic count and VMT in the base year. According to the Manual, these tests should indicate that calibrated traffic estimates for observed routes are within 20 % of actual volumes for road volumes over 50,000 AADT, and within 30 % of actual volumes for road volumes between 10,000 AADT and 50,000 AADT. The discussion of sensitivity is about what parameters to use for congestion delay curves. In forecasting, the Manual indicates that the objective is *"not to achieve a close match between the forecast and base year results, but rather to ensure that the differences and trends are reasonable."* But there is no discussion of the required or known accuracy of travel demand forecasts, and no guidance similar to the requirements for sensitivity testing such at that in the UK—if that is actually followed.

Technical sources of inaccuracies in travel demand forecasts have been variously attributed to inaccurate demographic forecasts, particularly population, households and employment trends at the zone level, but also to major changes in pricing such as gasoline prices, rising vehicle availability to more households, changing lifestyles such as women in the workforce and immigration, and to unverified assumptions regarding the stability of household travel relationships. As such, they can be investigated by a variety of methods such as better data collection, use of scenarios, probability of occurrence and interaction among sources. Although time-consuming, this approach is relatively straightforward, analytically.

Ethical issues in forecasting

However, there are also broader pressures influencing the forecasting process itself. In early papers on this topic Wachs (1989, 1990) focused on the ethics of forecasting in transportation planning, and particularly the institutional environment that encourages, perhaps even demands, that forecasts be thinly veiled "*highly subjective exercises in advocacy*" for project justification rather than "*utilitarian*" objective assessments. In essence, forecasts are often *"deliberately designed to put certain projects in a favorable light and others at a disadvantage*". To combat this issue Wachs called for public education on advocacy versus objectivity, protection of those exposing abuses, strengthening ethics codes and stronger education programs for professionals. Altshuler and Luberoff (2003) also note the pervasive influence of federal funding in moving megaprojects forward.

In controversial work Flyvbjerg et al. (2003, 2006, 2009, 2013; Naess 2006; Cantarelli et al. 2010) reviewed the forecasting accuracy for 258 highway and transit megaprojects worldwide, built between 1927 and 1998. They found that for transit projects, actual costs were on average 45 % higher than estimated, but the actual ridership averaged just 61 % of predicted ridership. For highway projects, the cost forecasting error averaged about 20 % high, and traffic estimates averaged about ±20 %. They attributed these inaccuracies to methodological errors, input errors, personal behavior that was different than predicted, unexpected external changes, political actions, consultant bias, and appraisal bias of the promoter. Toll-road forecasts were found to have larger error than non-toll forecasts. Cost overruns for large projects were attributed primarily to "*political explanations*" rather than technical, economic or psychological reasons. This bias was attributed largely to implicit cooperation between government officials, contractors, consultants and politicians, rather than to errors in input or modeling. Using very strong language, they suggested that cost overruns seemed to be "*best explained by strategic misrepresentation, namely lying, with a view to getting projects started.*" The main cause was found to be *"[neglect of] risk and lack of accountability in the decision-making process.*" Flyvbjerg (2009) argued that "*a rapid increase in stimulus spending, combined with more investments in emerging economies, combined with more spending on information technology, is catapulting infrastructure investment from the frying pan into the fire"*. To combat megaproject uncertainty, the authors call for greater transparency, model performance specifications, due diligence in forecast reviews, tighter regulation of construction and financing, use of private financing for risk capital, and improved management based on state-owned-enterprise and build-operate-transfer business models. More recently Flyvbjerg has called for penalizing forecasters for producing "*significantly inaccurate and misleading forecasts*" (Forster

2012), but others have noted that inaccurate client data are more to blame for forecasting errors (Forster 2013).

Using a data set from Standard and Poor's on vetting proposed toll road projects, Bain(2009a, b, 2011) reviewed the traffic forecast accuracy of 104 highway projects from numerous countries. He found that after correcting for optimism bias about 90 % of actual traffic volumes were within ±43 % of their respective forecasts, but charts in the study show that the average estimated 20-year error was about +20 % that is, the average 20-year forecast traffic was 20 % higher than actual. Bain also cautioned against inflating forecasts as an "*act of deception*", and warned of sensitivity tests that are "*insensitive to project traffic or revenue*", and using *"pseudoscience to infer precision…..not supported by empirical evidence*". He also recommended the use of a traffic risk index to score the major technical and institutional risks for a specific project (Bain 2009a, b). Lemp and Kockelman (2009) provide a comprehensive summary of these studies.

Wachs, Bain and Flyvbjerg generally infer the existence of deception, lying or collusion from the cumulative evidence of groups of projects (for instance, a high proportion of projects with over-forecast of traffic or revenue), and from comparisons of project forecasts with those of peer projects. But of course this does not mean that a specific project forecast is necessarily biased. And Wachs also points out that it is usually not illegal to amend the assumptions of a forecast for political reasons. Legal disclaimers in consultant documents, liability insurance, and disclaimers in financial documents generally protect forecasters from *ex post facto* legal action.

The extent of this form of manipulation is not fully known. Some observers believe it to be quite common, perhaps even widespread, in toll road forecasting. Others suggest it is essentially an amoral 'victimless crime': the public gets the road and the contractors get a profit. Others suggest it is quite rare, perhaps even non-existent, in non-toll forecasting activities. Clearly, however, the recent rapid increase in toll-based financing of major projects along with concomitant risks for private investors has increased the visibility of the issue. Considerably more research would be needed to identify its current magnitude. The Australian court cases now moving forward, which are centrally concerned with ethics, may clarify this issue.

All of the above studies concentrate exclusively on large projects within urban areas that use the conventional 4-step modeling process and on large inter-regional projects using similar methods. There are no comparable studies of smaller actions such as transit service increments or pedestrian-bike studies, or a wide range of other policies, although periodic reviews of *ex post facto* travel response to such policies have been made and codified (e.g., Pratt 2013). Essentially, we know very little about the accuracy of such tools, but it is seems unlikely that they are any more accurate in forecasting than the 4-step process.

Increasing complexity and inadequate policy response

Another serious issue is that the range and content of policies that these models are expected to address is increasing, especially increasingly complex road pricing schemes. As noted above, these methods were initially developed to evaluate alternative major transportation actions such as new roads, major widenings and major new fixed-guideway transit proposals. But in addition to these issues, numerous other policy options are now under scrutiny. They include demand-management policies such as VMT-reduction policies, employer parking restrictions and carpool mandates, vehicle user pricing, congestion pricing flexible work schedules, carpooling and vehicle sharing; technological changes such as vehicle fuel efficiency, energy constraints and costs; transportation system actions such as capacity reductions ('road diets'), pedestrian-bike systems, transit improvements and transit-oriented development; transportation funding mechanisms such as tolling, and impact issues such as air quality, climate change, noise, induced travel, socioeconomic impacts, land use, urban form and equity issues. In the US federal requirements for modeling include "*estimating motor vehicle emissions (which depends on estimating speeds and traffic volumes by time of day), estimating new travel generated by adding new capacity, evaluating alternative land use policies, and estimating freight movement and non-motorized trips*" (Spielberg 2007), to name just a few. Many of these issues are handled post-model and thus incorporate whatever errors the models contain. Looming just ahead are topics like internet communication, social networks and media, self-driving cars, automated vehicle routing and immigration.

The primary methodological advancements that deal with some (but not all) of these issues include disaggregate (disutility) and revealed preference methods, stated preference methods, activity-tour models, multi-modal networks, advanced statistical methods, dynamic assignment, micro-simulation, integrated land use-transportation models, geographic information systems, internet-related applications, and social media data monitoring. These improvements are increasingly complex but have less verification and calibration, little or no external validation, and (so far) virtually no *ex post facto* testing of model accuracy. They have also become operable by only a limited number of specialists. Some observers go further, suggesting that travel demand modeling has split into largely isolated sub-disciplines that only a few individuals can successfully integrate (Chow et al. 2012) and sometimes fail to communicate with other specialists. Most standard texts do not discuss issues such as model complexity, accuracy, or practical modeling problems.

Further, some of these methodological advancements may not be real advancements at all, since they are primarily modifications of existing tools. While some models contain genuinely new elements, most are operated as bolted-on additions to the existing 4-step process. Although some observers might disagree, I think it is safe to say that the basic 4-step paradigm we use for travel demand forecasting has not changed substantially since its development in the 1950's. The box looks different, but the contents are remarkably similar. In short, we are asking the 4-step model to do far more than what it was initially designed for, a form of mission creep.

In summary, the complexity of current models means that few modelers are proficient enough to understand the uncertainties, and many users have no way to assess model validity and have to basically trust the outputs. This leads to an impression of unwarranted accuracy in forecasting and false precision in estimates.

Widening gap between theory and practice

Many observers have also noted the widening gap between model theory and practice, which seemed to accelerate in the early 1970s with the advent of federal research money and advanced statistical tools and model structures. Many model advancements are not even used in practice, instead getting bogged down in the minutia of model results and trivial findings. And when the newer methods, for instance activity models and dynamic assignment, are placed in practice, it is alongside or as add-ons to existing 4-step processes (Jones 2012). Few regional planning offices have the expertise to operate these tools, leading by default to applications by consultants and academics if they occur at all—and of course the results might not be more accurate. Ironically the US's continuing investment in model improvement efforts (Travel Model Improvement Program 2010), particularly its investment in TRANSIMS, may have accelerated modeling advancements but resulted in lagging implementation, particularly in smaller regions that find the advancements cost-ineffective. As Jones notes (2012) the time lag gap and complexity gap between research and practice seems to be widening, and in fact is not new: as early as 1990 Pas noted the widening gap, and the various reports of the International Association for Travel Behavior Research (IATBR 2013) going back to its first conference in 1973 (TRB 1973) each also note the theory–practice gap.

In the background

Travel demand forecasting does not exist in a vacuum, either in theory or in practice. In the background are a host of additional political, regulatory, fiscal, geographic and institutional circumstances that affect model accuracy and forecast errors.

Regional planning

The history of travel demand models is well documented in a variety of papers, including Boyce and Williams (2005); Boyce and Williams (forthcoming, 2013); Ran (2012), Shuldiner and Shuldiner (2013) in this issue of *Transportation,* Pas (1990) and Jones (2012). These reviews trace the development of modeling methodology and modeler views of travel but generally place less attention on institutional considerations. In the US, requirements for long-range travel demand forecasting are actually *mandated i*n the 1962 Federal Aid Highway Act, which required 'comprehensive, continuing, and coordinated' (3-C) transportation planning in metropolitan regions with populations above 50,000 as a condition for federal financial assistance. The rules accompanying the Act laid out criteria for planning, including 20-year demographic forecasts by zone, multi-modal forecasts (quite a new exercise), economic analysis of alternative investment plans, and many other requirements. These requirements initially came from the apparent successes of early transportation studies in Chicago, Detroit, San Juan, Pittsburgh and other regions, which relied heavily on the capabilities of (then) 'modern high speed computers'. During the 1960s and 1970's consultants and academics exported this technology to virtually every major city in the world. The 4-step process has gone through several stages of refinement since then. Ran et al. (2012) classify these periods largely by the availability of communication technology, and Boyce and Williams (2005) classify them by mathematical complexity and consistency, rather than by topical focus. Recent regulations (for instance, requirements for conformity with state air quality plans, noise and water quality impacts, encouraging time extensions to 40 years rather than the original 20, congestion management and environmental justice) have substantially increased the 4-step model's complexity and stretched its credibility, some say to the breaking point. Some of these exercises, particularly EIS requirements for US projects, require or encourage project-specific point-based estimates of future travel demand by time of day and speed, a clearly unrealistic goal.

Some observers have recently criticized the present long-range planning process as fundamentally flawed, with too many unachievable goals, feel-good scoring to evaluate projects, failure to challenge critical underlying assumptions, and general lack of critical thought. In a review of 75 long range plans for US metropolitan regions, O'Toole (2008) found that *"nearly all [long range transportation plans] included population and growth and traffic, but few describe[d] how reliable the travel forecasts might be, [and] no plan reported that planners did any sensitivity analyses to deal with questionable assumptions and forecasts."* In reviewing the long range plans for 23 large US regions this author (Hartgen et al. 2008) found that very few plans were realistic about the impact of projected growth on congestion, often forecasting much lower road speeds as congestion increased, even with large plan expenditures. Virtually all plans used the 4-step process for traffic forecasting, but none made forecasts for pedestrian or bike use, or of car-sharing, regional pricing scenarios, work-schedule policies, or other options. Only one region (Los Angeles) had estimated the impact of CO_2 control policies. The review also noted remarkably similar content and format across plans that rigidly parroted federal guidelines, suggesting that they had become standardized 'process' documents driven by federally-directed formats and contents, intended to ensure self-certification for continued federal funding, but then put on the shelf until the next 'five year plan' update.

Many non-capital policies reviewed in the long range plans are typically not easily amenable to analysis with the 4-step process, but also are not analyzed using other methods. These include regional policies for pedestrian-bike networks, and some bus-based transit networks. Policies that are geographically limited, such as transit-oriented development, are also typically not studied with these tools. On the other hand, some topics such as air quality conformity are over-modeled relative to impact. The author's study (Hartgen et al. 2008) found that the typical US region predicted 50–80 % reduction in emissions just from vehicle turnover, but the emissions impact of new projects was typically just 0.1–0.25 % of regional emissions, way below the modeling accuracy threshold. Another topic, environmental justice, is typically treated by analyzing where disadvantaged populations live along proposed highway projects, ignoring the increases in regional access to jobs or housing that such projects might bring. On the other side of the ledger are topics that are under-modeled, such as regional accessibility, induced travel, productivity and jobs, and economic development which are of great interest to localities (e.g., Worsley 2013) but rarely studied in long range plans.

Project planning

At the project level, there is also considerable regulatory and technical complexity. For project level forecasts, the relative magnitude of uncertainly of zonal population and employment projections increases exponentially with declining zone size and increasing forecast horizon (even for historically stable or filled zones that may be re-built to higher density). It is one thing to forecast, with some confidence, the likely population of an entire urban region, but it is much more difficult to forecast the traffic in and out of a small zone, let alone one influenced by a proposed new route (the so-called induced travel issue).

In the US, the analysis of projects has become substantially institutionalized. The Environmental Impact Statements required for major new highway projects by the National Environmental Policy Act of 1970 was initially intended to evaluate alternatives against a "no build" option, and develop a "preferred alternative" for later design. A brief review of Environment Impact Statements for major road projects reveals remarkable similarity in the structure and content of documents, each following the exact same format, even using similar words and density of material for traffic forecasts and induced travel. Technical supporting documents are also remarkably similar. Sensitivity analysis of forecasts for projects is unheard of. For transit projects, early ridership forecasts commonly used 'adjustments''(sometimes called 'bias coefficients') to account for un-measured transit features in modeling estimates, or included otherwise optimistic assumptions about operation and cost, but some sensitivity analysis is now undertaken, particularly in more recent documents. The underlying goal of both processes seems to be to ensure that the required steps are followed to comply with law and regulation, but also to justify the project and lock in federal dollars before other projects get them. There is virtually no attention to the experience of similar projects, either locally or in other regions (sometimes termed 'reference class' reviews) that might shed light on project worthiness.

Another serious problem is the overshadowing presence of large non-local (federal, state and private) matching funds for major projects. In the US, the federal share for Interstate projects is generally 90 %, and 80 % for other federal-aid projects, with the states (sometimes the private sector) picking up the bulk of the remainder. For major transit projects, the federal share is typically 50 %, and localities (sometimes states) fund the remainder. In both cases the process is intensely competitive, with sponsors vying for funding within each state (for roads) or between cities (for transit). Since these projects are generally financially infeasible without funding help from higher levels of government or the private sector, localities rely on federal, state or private sector funding to progress them. This biases the local benefit-cost ratio in favor of the project since the localities generally receive all or most of the benefits but contribute at most 20–50 % of the cost. (Sometimes, additional items are monetized to increase the benefit-cost ratio). Localities may have less interest in the forecast traffic but instead just want the project built. In essence, the non-local funding pushes forward projects that might not be built if localities had to pay for them entirely. This is particularly the case for projects with private-sector involvement, in which amorphous distant investors are seen as taking the risk, but it also applies to projects funded by formula allocations. Although there were certainly expressions of such concerns in the past (for instance, the Dulles Greenway), one might also argue that the whole issue of modeling uncertainty did not move to the fore until private-sector-financed megaprojects failed expectations of cost or demand.

Major weaknesses in travel demand modeling

The fundamental weaknesses of currently used in travel demand modeling have been extensively enumerated by many observers. In summary, they are:

- Unrealistic model paradigm: It is well recognized that the 4-step modeling paradigm developed 50–60 years ago is only a computational convenience that is not behavioral and does not reflect how traveler decisions are actually made. In spite of widespread agreement concerning this weakness, comparatively little basic research has been conducted into how households actually make activity and travel decisions (Pas 1990; Talvitie 1997; Jones 2012; Strategic Highway Research Program II 2013a, b). Even though the potential use of new different paradigms is significant, most activity models begin with largely unverified assumptions regarding household decision-making, for instance synthesizing activity patterns for demographic classifications, role and resource allocation within and between households, social networks and assumptions regarding trip chaining. And since many theories can often fit a data set, mere calibration is not validation of theory, let alone forecasting capability.
- Questionable methodological advances. The major improvements to the 4-step process (for instance, attitudinal inputs, choice models, activity-tour methods, dynamic assignment, micro-simulation, and land use integration) have significantly increased

model complexity and cost, but (with the exception of choice models) have not produced significant advances in performance or understanding. Further, (again, with the exception of mode choice models) these improvements have been applied to only a few limited cases for larger regions, leaving smaller regions stuck with the old paradigm (An exception is the SHRP II effort 2013a, b, for Jacksonville FL and Burlington VT). They are also expensive and time-consuming to implement, with unknown incremental forecasting value, which make them probably not cost-effective for smaller regions and possibly not cost-effective for large regions either.

- Questionable accuracy. As noted above, the few rigorous tests of 4-step model accuracy have not been comforting, showing instead wide variations between predicted and actual future traffic. Although no formal research has been conducted, it is likely that most 20-year traffic forecast accuracies for toll-free road widenings are no closer than within +-30 %, but probably a much higher uncertainty range.
- Cross-sectional data: Data used to build and then validate 4-step models is invariably cross-sectional in time and therefore by definition contains a frozen view of travel behavior at the time of the travel survey. There are only a very few long-term longitudinal panel surveys of travel, notably in Seattle (now defunct) and the Netherlands (beginning again in 2013 after earlier waves in the 1980's), and just a few before-after surveys, for example tolling projects in Atlanta, Seattle, and Minneapolis. But even if one could forecast the life-stages of a future population, the stability of travel behavior must also be assumed going forward. To use a cross-sectional model in forecasting requires the assumption that the underlying behavioral relationships captured in the data are unchanged from the base year going forward, implying that whatever changes led up to the current behavior will then be suspended going forward, a logically untenable position.
- Coarse zone structure: Most 4-step models have a zone structure that is too coarse for project planning or for numerous non-infrastructure policies. This is particularly true for projects at the edges of regions, where zones are larger and growth rates tend to be faster. But even within urban areas the zone structure is too coarse for modeling choice behavior for many policies, and is a particularly serious problem for transit forecasting where walk-access to and from proposed transit stops is typically shorter than zone-to-zone centroid distances. Of course, zone size could be reduced to deal with this issue, but then socioeconomic forecasts are even more uncertain. New GPS methods and internet-based data tracking, such as cell-phone data, may reduce this problem, but the resulting data are less accessible and have less attached socioeconomic data.
- Inadequate sampling. Small sample sizes, for instance for choice-based models, are too small for generalization and do not gather information on specific O-D movements. Even large area-wide surveys rarely interview more than 2 % of regional households, so most O-D flow cells are empty. GPS-based data and cell-phone tracking data are more extensive but lack demographic and socioeconomic characteristics.
- Over-sampling. Occasionally models are built with so many observations that all model statistics are found to be highly significant, even though the model itself is miss-specified and is useless in forecasting. This is likely to become a more serious issue as big data sets from cell phones, for instance, are used to estimate corridor traffic characteristics.
- Misspecification. The presence of low goodness-of-fit measures, for instance low R-squares, is often a tip-off that the model is miss-specified. Many observers note that poorly specified variables and missing variables are a common problem in travel demand modeling, and that reliance on just time and cost for modeling is a significant

mistake. For instance, heavy reliance on travel time, cost and income variables in mode choice modeling reduces attention to such factors as household location and work-school travel needs, habit, modal bias, prior investment in vehicle ownership, socioeconomic contexts reliability, convenience, safety and other factors that are likely more important, but difficult to collect and model.

- Weak calibration and verification. Even when calibrated on existing data or 'verified' using hold-out samples, goodness-of-fit statistics often show weak model explanatory power. And of course tests of base-year fit are not indications of external validity, and are no guide of predictive performance.
- Trivial or nonsensical findings. Often, tests for policies show very little or no difference between the policy impacts and the no-build option, such differences generally being much smaller than the intrinsic uncertainty in input parameters. But as Alonso (1968) pointed out many years ago, error propagation mathematics indicates that the largest relative errors come from subtraction and division. And arithmetic's significant-figure rules tell us that the precision of a mathematical formula cannot be more than its least precise term. It is therefore illogical to assert that the difference between the results of two tests is more accurate than the accuracy of the tests themselves. Basically, the smaller the difference being observed, the stronger, not weaker, the underlying model has to be.
- Limited before-after testing. Very few modeling exercises are designed to conduct before-after studies to determine how well the model predicted traffic. As noted above, in the US only major transit investments have been subjected to such tests, and even those studies have not identified the specific causes of the variation.
- Limited use of similar experiences. Modeling efforts have generally used locally-driven data such as trip rates and trip lengths, and locally developed forecasts for such items as population and land use, with only limited comparisons of forecasts to the experiences of similar regions. So-called 'reference class forecasting' has rarely been used as a backdrop to locally-driven forecasts, although the Federal Transit Administration is increasing its use (2013).
- Limited policy relevance. Models often contain variables that are irrelevant or trivial with respect to the alternatives being studied. The typical variables at hand (generally time or cost) are poor descriptors of the features of many policies.

What to do?

In the US the single greatest knowledge gap in travel demand modeling is not the unknown elements of travel decision-making, great as that is, but the uncertainty of toll-free road traffic forecasts based on conventional 4-step travel forecasts. A sponsored effort to research, consolidate and report on the demonstrable accuracy of historical travel demand forecasts (similar to the EU study), and the likely sources of error, would be of immense help in quantifying the magnitude of this problem. This should also be undertaken for other forecasts that do not use the 4-step method.

In the meantime we can of course continue on as we have been, funding modeling improvements incrementally as with TMIP and other research funding streams, addressing specific shortcomings, and encouraging adoption as regions update their models. But this approach would leave large gaps in knowledge regarding the accuracy of our models, no formal mechanisms for quantifying or reducing the uncertainly of our methods or in addressing the increasingly visible biases in them, and no structured mechanism for improving our understanding of travel decisions. This approach would likely accelerate the continued shotgun funding of research with no overarching sense of needed knowledge and a continued widening gap between theory and practice. As the weaknesses in travel demand forecasting methods become more apparent, driven likely by investor demands, doubts about their use will increase and questions about their usefulness will likely proliferate. This would lead, in my view, to the slow but steady erosion of relevance.

Hubris

On the other hand, the generally positive experience of the US Federal Transit Administration to improve transit modeling accuracy, and the efforts in the UK to improve road traffic modeling accuracy, suggest that modest progress in the accuracy of travel demand models is possible with a targeted and directed effort. (It should be noted that neither effort has yet actually demonstrably improved forecasting accuracy.) More likely, recent reviews of model accuracy highlight weaknesses that probably cannot be fixed without new basic modeling paradigms. Other disciplines (biology, astrophysics, even highway research—the Strategic Highway Research Program) have developed broad long-term agendas to improve understanding, but travel demand forecasting targets short term research reflecting researchers' interests, not users' long-term goals.

If such a moon-shot effort were undertaken, it would have to be based on the conviction that we need to get a lot more serious about understanding travel. In this approach, a concerted effort would be made to improve the accuracy of travel demand modeling over the next 20–30 years, and account for emerging opportunities such as cloud computing and social media. Ran et al. (Ran 2012) provide a broad overview of how these emerging communication technologies might ultimately provide the underpinnings of integrated system models that could 'optimize' future individual travel choices. And Shuldiner and Shuldiner (2013 in this issue of *Transportation*) outline how social media may transform travel behavior.

However, to be successful, this approach cannot rest just on Big Data or more computing power. We also need an honest assessment of what we know, what we don't know, and what we need to know to address emerging policy issues. These gaps have been characterized as consumer responses to collaborative consumption options such as car clubs, the impact of information services, internet-driven substitution for travel, and transport-land use interactions (Polak 2013), but one might also add the economic impact of transport proposals (Worsley 2013), energy pricing, automated cars, and emerging lifestyles. Once we know what we need, we then need to formulate long-term research plans spanning perhaps 20–30 years to get what we need and implement it. This will not be easy and will require coordination among a wide range of stakeholders.

Recently the US's Strategic Highway Research Program II (2013a, b) has initiated a large ($4 m) project to improve modeling by integrating activity models with network and capacity information. The goal of the project is to build *a "dynamic integrated model [emphasizing] behavioral changes in use of highways in response to highway conditions. Methods proposed should address changes in demand such as micro-time of day choice (i.e. peak spreading) and route choice in response to adding lanes and in response to operational improvements such as ramp metering, signal coordination, freeway management, ITS, reversible lanes, HOV/HOT lanes, variable tolls, variable speed limits, and bottleneck improvements. Detailed, time-sensitive highway networks will be expected to* *include detailed and accurate highway operating characteristics to ensure that such policies can be adequately addressed"*. Rather than being a new paradigm, this effort essentially grafts activity modeling methods onto traffic assignment. And it might lead to failure, or (more likely) to limited use in practice. Extensions to incorporate social networks (e.g., Auld and Zhang 2013; Deutsch and Goulias 2013) substantially increase modeling complexity. The US effort to implement TRANSIMS (a individual-level micro-simulation network-based modeling system), spending very large sums over 20 years but (at this point) having no operational application and few regional tests, is a cautionary experience to be taken very seriously. Big Data and Big Research have a significant risk of Big Failure.

The following is my admittedly biased professional view of what might be needed in such an effort:

- Professional standards for the use of models and treatment of uncertainty.
  First, terms need to be better defined. This review found a variety of definitions for such terms as *base ridership/volume* (current year, year of original data, etc.), *forecast traffic* (e.g., traffic estimate used to make the construction decisions, post-decision estimates), *forecast year* (future year, years from opening, years from forecast, etc.), and *actual traffic* (e.g., traffic at one year after opening, 5 years, 10 years, 20 years, or by date, etc.). The definitions should also include methods for adjusting short day-long, week-long, or monthly counts to ADT using seasonality and axles. Although such variation is to be expected in emerging topics of study, if travel demand model forecasts are to be compared and evaluated, these basic rules for metrics need to be established.
  We need arrangements for monitoring and periodically reporting model accuracy. Perhaps an (international?) effort to consolidate and archive modeling forecasts, along the lines of the University of Minnesota's archive of travel surveys, or an extension of FHWA's archiving of long range transportation plans, could be developed cooperatively by professional organizations, universities and governments.
  Standards for expectations for model accuracy (not just validation), setting acceptable error ranges are also needed. How accurate should forecasts be, for various circumstances (context, mode, functional class, forecast year, etc.)? Perhaps 20–20 accuracy (±20 %, 20 years out) is close enough for a decision regarding total traffic for a proposed new road. But greater accuracy might be needed for designing pavement strength, or for decisions regarding toll financing. On the other hand, decisions regarding the number of lanes might be needed only to within ±50 % (Polzin 2013). Flyvbjerg (2013) calls for various professional associations to set such standards and the ethics for forecasting.
  Standards for the procedures for estimating forecasting uncertainty, in both ranges and scenarios, and their probability of occurrence, should also be developed, and reports should include uncertainty alongside every forecast (e.g., US National Association of Municipal Analysts 2005). Bain suggests the use of fan charts, a technique used in monetary policy (Britton et al. 1999), as a means for developing and describing the uncertainty of forecasts to non-experts. For Instance, Fig. 1 shows the 10-percentile (likelihood) ranges (high and low) of predicted transactions for a toll road forecast, by year. Since much of the discussion of this issue is based on toll-road development by large international corporations, international and private-sector organizational cooperation would seem to be essential in developing quantitative standards for model accuracy.

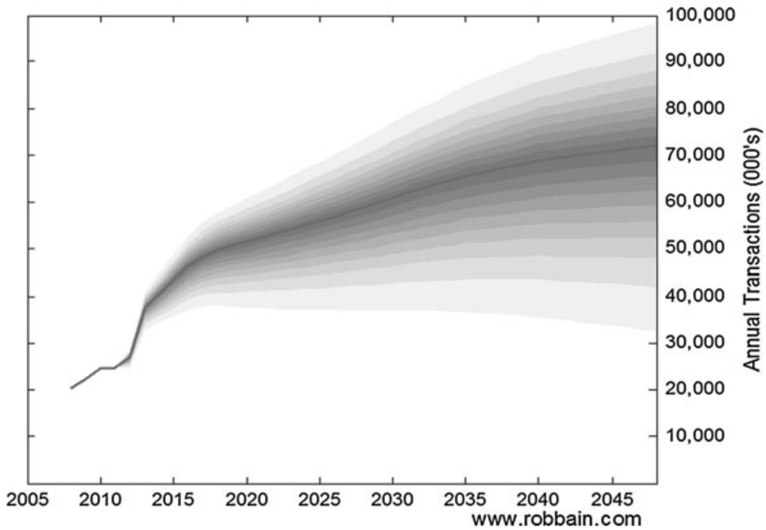

Fig. 1 An example of a traffic transactions fan chart

Additionally, we need more comprehensive and more direct education for both students and professionals regarding the ethics of forecasting. This is a topic that dates from at least 1989 (Wachs 1989) but is treated lightly in most curricula.

- Better modeling.
  The above steps set the stage for improved travel demand forecasting, but they do ensure it. We also need to substantially improve, in a coordinated fashion, our understanding and modeling of travel demand.
  A critical missing element in our current research is the lack of coordination among research agencies and institutions, both within and across nations. We need better organizational structures, not just to monitor and report on modeling 'successes' and 'failures' as noted above, but also to prioritize and develop improvements in a wide range of technical modeling issues. These should be international in scope.
  And we need better data. Virtually all of the information we have on travel behavior is cross-sectional, and does not track changes in behavior over time. We need to gather more real-time data to get at variations in travel, and use panel data as the rule, not the exception, in model development and validation. It will be a challenge to establish and fund such efforts, but the experience of long-term panels in other disciplines such as health, time use, and consumer surveys suggests that it can be done.
  Most important, we need better understanding, which is not likely to be achieved without coordinated research to understand travel behavior and household travel decision-making. This means developing a unified international research agenda, thorough conferences, meetings, and associations, to identify and recommend needed key research. The longer-term goal is to develop a holistic theory of household decision-making (attitudes, roles, activities, location and social networks, allocation of resources, travel, feedback to development, etc.) to guide model development. This means clarifying what is unknown that is needed, and establishing research programs to get it.

But knowledge is not enough. We need to get the results to practice. The focus should be on developing new paradigms for handling issues not adequately covered with current methods, particularly non-capital policies. Perhaps most controversial, the process should be user-driven to ensure relevance to policy. Buy-ins by major schools interested in participation will also be essential to reduce the remoteness of academic research from practice. Academic papers in journals and presentations at major practice-oriented organizations such as the Transportation Research Board should include a specific application-to-practice section outlining how to use the results in practice. Perhaps we also need to unhook tenure decisions from papers published, so that academic research will be better concentrated on coordinated goal-oriented research rather than on the incremental advancements now being reported.

This may require more money or at least more targeted funding. Potential sources include pooled funds (e.g., Spielberg 2007), private-sector funding of modeling research (e.g., Johnston 2012), or simply coordinated governmental research programs. There are numerous examples of this in various research communities, including the Transportation Research Board's Strategic Highway Research Programs. When there's a will, a way will be found.

- Better forecasting.

  As difficult as is better modeling, it is even more difficult to develop better travel demand forecasts, which rely not just on understanding behavior but also predicting it into uncertain circumstances.

  To do this, we need to develop better ways to quantify the uncertainty in land use, employment and demographic forecasts, particularly for sub-regional zones near proposed projects. We can conduct more research on model transferability, so that the circumstances under which modeling findings can be compared are more clearly understood. We can expand the link to practice through the use of elasticities and other devices. We can use Monte Carlo, stress tests (Lemp and Kockelman 2009) and reference class forecasting to develop realistic ranges and probabilities of outcomes, rather than single absolute numbers. We can develop ways to handle external events such as recessions and booms, political changes, or energy crises or breakthroughs. We can widely publicize caveats regarding model accuracy in project promotion and review material.

- Institutional improvements.

  As the above review suggests, a significant element of the malaise now being experienced in travel demand modeling is the institutional structure that drives it. Some of this is dependent on the increasingly complex tasks we are asking models to perform, but other elements are relatively simple to resolve. For instance, reserving a portion of modeling funds for evaluating subsequent model accuracy would quite rapidly establish information on short-term modeling accuracy. We can also take steps to lower optimism bias, for instance by increasing local stake in project funding and by evaluating proposal worthiness independently of the source of funds. A more extreme action might be to establish independent forecasting capability, separate from agencies, sponsors, or financers. Even further, we might include incentives and disincentives, or perhaps even penalties suggested by Flyvbjerg (Forster 2012), for forecast errors. However, the use of punishments might lead to firms exiting the business, and may have already worsened the case for better forecasts (Johnston 2012).

Humility

The above approach is predicated on the belief that with a concerted coordinated effort the understanding of travel behavior and the accuracy of travel demand forecasts can be improved sufficiently for confident use in public and private transportation decision-making. Essentially it builds on the optimism of humans to better their environments through study, planning, and implementation of needed investments. But there are limits to such knowledge, and some things are not knowable with certainty. Given the uncertainties of its inputs, travel demand forecasts may be one of these. Travel, activities, demographics, land use and transportation investment are so complex that it may be simply not possible to usefully forecast future travel demand. As van Vuren (2013) cautions, *"We need to move away from the idea that models can solve problems and give the right answers. Models should be used to sharpen the questions and test different assumptions"*.

Accepting this uncertainty does not seem to have been a serious problem in the past. Travel demand forecasting as a craft is only 50–60 years old, yet thousands of projects were built in the three millennia before that. This is not to suggest that we should return to the King's Edict of project advancement or to complete reliance on the private sector to initiate major projects, but it does suggest that historically other mechanisms for decision-making that are not so data-driven have also produced major projects.

Other considerations might actually reduce the need for more accurate travel demand forecasts. Many projects are built to meet today's problems, not future demand, so the need for a long range forecast may not be so great. Slowing population growth and wide geographic variations in growth also suggest cautionary forecasts. Other justifications for projects (environmental, economic and social) are often as important as traditional user-based justifications (time savings, reliability, safety, and operating costs). And as noted above, traditional 4-step methods don't handle many current policies very well anyhow, so de-coupling the 4-step method from some policies might be sensible (Polzin 2013).

Rather the struggle to know what can't be known, it may be wiser and more fruitful to openly acknowledge the uncertainties of this business, and to build that uncertainty into our decision-making. This approach would contain the following elements:

- Highlight model limitations. Led by professionals, establish an international travel demand modeling organization with a clear mission to monitor and report modeling performance, highlight the limitations of forecasts, and improve modeling performance where that can be done. Several potential organizations already exist that could serve this mission.
- Through this organization, set and promote ethical standards for the conduct of travel demand modeling, which specifically identify modeling uncertainties and limitations, and ensure that all major forecasts adhere to these standards. This might follow the examples of other professions (accounting, engineering, etc.) that have successfully established standards of practice.
- Use due diligence methods to evaluate project forecasts, using statistically reliable benchmarks from similar projects to estimate the average of previous projects, compared with the proposed project (Flyvbjerg 2013).
- Evaluate the accuracy of travel demand models in a variety of settings, and publicize the findings. This could take the form of mutual fund ratings, government reviews, a private rate- my-forecast initiative, university-based assessments, or trade group evaluations.

- Admit we don't and can't know many elements needed for accurate travel demand forecasting, and recognize the limitations of our knowledge. Set realistic expectations.
- Substantially increase the recognition of forecast uncertainty by modelers, citizens and decision-makers. This can be accomplished though clear caveats for forecasts, use of ranges and probabilities, and such devices as fan charts.
- Substantially increase the transparency of travel demand forecasting, using clearer and simpler methods to describe the major techniques, automated game-like scenarios and the like.
- Convert forecasts from single point-based estimates to range-based with probability of outcome.
- Expand the use of scenario analysis that challenges baseline assumptions, particularly for scenarios that initially seem unlikely but are extreme.
- Significantly reduce the role of travel demand modeling in project decision-making. Repeal laws and regulations that contain model-driven estimates.
- Eliminate regulations that require point-forecast modeling or tie funding or compliance to modeling capability. Examples of these in the US are requirements to 'demonstrate' air quality conformity, congestion management regulations, long-range planning requirements, project-based noise modeling, and project-based environmental damage modeling.
- Weight model forecasts less with regard to other factors in project decision-making. Deliberately make less use of model results in cost-benefit assessment.
- Increase the local share of project funding to strengthen locality involvement in project decisions and get the hard questions asked when one's own money is on the line.
- Reduce or eliminate competitive grant funding which may contribute to optimistic forecasts that justify sponsor proposals.
- Improve education regarding uncertainty and the ethics of forecasting.

Various blends of these approaches could also be suggested. For instance, one might focus on several common elements that will be needed regardless of direction, such as better monitoring of modeling results, standards for modeling and forecasting, educational initiatives, and increased local share of funding, possibly a modest ramp-up of coordinated research, and stronger treatment of ethics.

Either of these approaches is quite different from what we are doing now, and frankly I am not too optimistic about either being adopted. To succeed, either would need the support of professionals, trade organizations, institutions and governments, politicians, academics, consultants, project developers and promoters, localities, and the public. Within all of these groups, there are strong vested interests for the status quo. Champions and advocates for each approach, or a blend of the two, should come forward.

Why is this important? The purpose of travel demand modeling and forecasting is to improve investment and policy decisions and the value of public dollars in an age of public austerity, by improving the accuracy and relevance of forecasting and analysis tools. These investments generally use taxpayer or client dollars, so professionals owe them the best estimates possible, along with recognition of uncertainty.

The essence of the scientific method is to observe, theorize, test to find discrepancies, and then modify the theory. The travel demand modeling community does a fair job of observing and theorizing, but we do a poor job of finding discrepancies and modifying our theories. Our fundamental modeling paradigm, the 4-step process, has not changed substantially in 60 years, and its accuracy is highly suspect. Some of us who have participated in this discipline are therefore concerned that the future of travel demand forecasting (if that itself can be predicted) is threatened by increasing skepticism. Will the discipline survive the next 50 years? What will it look like? Will travel demand forecasts be more accurate and more revered than today? Or will they increasingly be viewed as "*highly subjective exercises in advocacy*" (Wachs 1989), discredited by project reviewers?

Acknowledgments The author is indebted to many colleagues, but particularly to Robert Bain, Ken Cervenka, David Hyder, Ram Pendyala, Steven Polzin, Guy Rousseau, Howard Slavin and Martin Richards for comments relating to this paper. The author of course remains wholly responsible for the views expressed herein.

References

Alonso, W.: Predicting best with imperfect data. J. Am. Inst. Plan. **34**(4), 248–255 (1968)

Auld, J., Zhang, L.: Inter-personal interactions and constraints in travel behavior within households. Transportation **40–4**, 751–754 (2013)

Altshuler, A., Luberoff, D.: Mega-Projects: The Changing Politics of Urban Public Investment. Brookings Institute, Washington (2003)

Australian Government: Dept of Infrastructure and Transport. Public Consultation: Patronage Forecasting for PPP/Toll Roads. http://www.infrastructure.gov.au/infrastructure/public_consultations/patronage forecasting.aspx (2012). Accessed June 19, 2013

Bain, R.: Error and optimism bias in toll road traffic forecasts. Transportation, 36 (5), p. 469–482. http://www.robbain.com/Bain_Transportation_2009.pdf (2009)

Bain, R.: Big numbers win prizes. Project Finance International. www.pfie.com (2009)

Bain R.: On the reasonableness of traffic forecasts, Traffic Engineering and Control. www.tecmagazine.com (2011)

Bain R.: Toll Roads Big Trouble Down Under. Infrastructure Journal (2013)

Bartholomew, K.: Land use-transportation scenario planning: promise, and reality. Transportation **34**(4), 397–412 (2007)

Ran, B., Tavner, P.J., et al.: Perspectives on future transportation research: impact of intelligent transportation system technologies on next-generation transportation modeling. J. Intell Transp Syst: Technol, Plan Oper **16**(4), 226–242 (2012)

Boyce, D.E., Williams, H.C.W.L.: Urban travel forecasting in the USA and UK. In: Reggiani, A., Schintler, L. (eds.) Methods and Models in Transport and Telecommunications: Cross Atlantic Perspectives, pp. 25–44. Springer, Berlin (2005)

Boyce, D.C., Williams, H.C.W.L.: Forecasting Urban Travel—Past, Present, and Future. Edward Elgar Publishing Ltd., Cheltenham, UK (Forthcoming,2013)

Britton, E et al.: The inflation report projections—understanding the fan chart. Bank of England Conjunctural Assessment and Projections Division, Inflation Report, (1999)

Cantarelli, C et al.: Cost overruns in large-scale transportation infrastructure projects: explanations and their theoretical embeddedness. Transportation 10:1, p 5–18, Mendeley Publications (2010)

Chatterjee, A et al. Improving Transportation Data for Mobile Source Emissions Estimates, NCHRP Report 394, Transportation Research Board, Washington, DC. www.trb.org (1997)

Chow, J. Y. J., et al. Is transport modeling education too multi-disciplinary? a manifesto on the search for its evolving identity. Resource paper presented at the conference of the International Association for Travel Behavior Research, Toronto CA, (2012)

de Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., Hofman, F.: Uncertainty in traffic forecasts: literature review and new results for The Netherlands. Transportation **34**, 375–395 (2007)

Deutsch, K., Goulias, K.G.: Decision makers and socializers, social networks, and the role of individuals as participants. Transportation **40**(4), 755–772 (2013)

Department for Transport (United Kingdom): Traffic Forecast Errors Highlighted, Contract Journal (1988)

Department for Transport (United Kingdom). National Road Traffic Forecasts (Great Britain), 1997: Treatment of Uncertainty. Working Paper #5 (1998)

Department for Transport (United Kingdom): Transportation Analysis Guidance: Treatment of uncertainty in model forecasting. Section 3.15.5. http://www.dft.gov.uk/webtag/index.php (2013)

Donnelly, R. et al. Advanced Practices in Travel Forecasting: A Synthesis of Highway Practice. National Cooperative Highway Research Program Synthesis 406, Transportation Research Board. www.trb.org (2010)

European Court of Auditors, Are EU cohesion policy funds well spent on roads? Special Report No. 5, Luxembourg, Luxembourg. http://eca.europa.eu (2013)
Federal Highway Administration: Interim guidance on the application of travel and land use forecasting in NEPA. USDOT. www.fhwa.dot.gov (2010)
Federal Transit Administration: The predicted and actual impacts of New Starts projects: capital cost and ridership. Federal Transit Administration, Washington DC (2008)
Federal Transit Administration: Travel Forecasting for New Starts Proposals. Workshop PowerPoint presentations, Tampa Florida. http://www.fta.dot.gov/12304_9547.html (2009)
Federal Transit Administration. Proposed New Starts Policy Guidance. US Dept of Transportation. http://www.fta.dot.gov/documents/NewStartsPolicyGuidance.pdf (2013)
Florida Department of Transportation. Generalized annual average daily volumes for Florida's urbanized areas, Systems Planning Office, FDOT, Tallahassee, FL 32399-0450, October 4, (2010)
Flyvbjerg, B. et al. Megaprojects and risk: an anatomy of ambition. Cambridge University Press (2003)
Flyvbjerg, B., et al.: Inaccuracy in traffic forecasts. Transport Reviews **26**(1), 1–24 (2006)
Flyvbjerg, B.: Survival of the un-fittest: why the worst infrastructure gets built and what we can do about it. Oxford Rev Econ Policy **25**(3), 344–367 (2009)
Flyvbjerg, B. [The] majority of forecasters are fools or liars. BT Centre for Major Programme Management, Saïd Business School, University of Oxford. www.sbs.ox.ac.uk (2013)
Forster, A.: Punish consultants for faulty forecasts, says project expert. TransportXtra, Dec. 21. www.transportxtra.com (2012)
Forster, A.: Client data often to blame for forecasting errors, say planners. TransprtXtra, January 11. www.transportxtra.com (2013)
Hartgen, DT et al.: Are they ready? Mega-region growth and transportation investment. A report to the Urban Land Institute, Washington DC, March 7. www.hartgengroup.net (2008)
Highway Capacity Manual: Transportation Research Board. www.trb.org (2010)
International Association for Travel Behavior Research (IATBR). www.iatbr.weebly.com (2013)
Jaggi, B., Axhausen, K.: Comparison of approaches on the computation of travel time losses on high capacity, paper submitted to the Transportation Research board, 2014 Annual Meeting, version of June 16 (2013)
Johnston, D.: Traffic Forecasting for Australian Toll Roads. Comments to the Australian Road Agency. http://www.infrastructure.gov.au/infrastructure/public_consultations/patronage_forecasting_submissions/index.aspx (2012)
Jones, P.: The role of an evolving paradigm in shaping international transport research and policy agendas over the last 50 years. University of Oxford, UK, June 25 (2012)
Lemp, J. D., Kockelman, K. M.: Understanding and accommodating risk and uncertainty in toll road projects: a review of the literature. Transportation Research Record 2132, TRB, Washibngton DC, p. 106–12. www.trb.org (2009)
Li, Z., Hensher, D.: Toll roads in Australia: an overview of characteristics and accuracy of demand forecasts. Transp Rev **30**(5), 541–569 (2010)
Mackie, P., Preston, J.: Twenty-one sources of error and bias in transport project appraisal, Transport Policy, 5:1, p. 1–7 (1998)
Naess, P. Flyvbjerg, B., Buhl, S.: Do road planners produce more 'honest numbers' than rail planners? an analysis of accuracy in road-traffic forecasts in cities versus peripheral regions. Transport Reviews, 26(5), pp. 537–555 (2006)
National Association of Municipal Analysts (NFMA). Recommended Best Practices in Disclosure for Toll Road Financings. http://www.nfma.org/disclosure/rbp_toll_road.pdf (2005)
National Cooperative Highway Research Program, Driving forces influencing future freight flows, NCHRP Web Document 195. www.trb.org (2010)
National Cooperative Highway Research Program. Estimating toll road demand and revenue, Synthesis 364, Washington DC: Transportation Research. www.trb.org (2006)
O'Toole, R. Roadmap to gridlock: the failure of long-range metropolitan transportation planning. Cato Institute, Policy Analysis Paper 617. www.cato.org (2008)
Pas, E.: Is travel demand analysis and modelling in the doldrums? In: Jones, P.M. (ed.) Developments in Dynamic and Activity-Based Approaches to Travel Analysis, pp. 3–27. Oxford Studies in Transport, Aldershot (1990)
Pickrell, D. H.: Urban rail transit projects: forecast versus actual ridership and costs. Final Report, Urban Mass Transit Administration, Volpe Center, Boston MA (1989)
Pickrell, D.H.: A desire named streetcar: fantasy and fact in rail corridor planning. J. Am. Plan. Assoc. **58**(2), 158–176 (1992)

Polak, John.: New challenges in city transport modeling. Address to Modeling World, reported in Local Transport Today (2013)
Polzin, S. Prescription for the future. Private correspondence with the author (2013)
Pratt, RH. Traveler response to transportation systems changes. Report 95, National Transit Cooperative research Program, Transportation Research Board. Washington DC. www.trb.org (2013)
Shuldiner A and Shuldiner P. The measure of all things: reflections on changing conceptions of the individual In travel demand modeling. Transportation, Vol. 40.6 (2013)
Spielberg, F.: Metropolitan travel forecasting, Special Report 288, Transportation Research Board. www.trb.org (2007)
Stopher, P.R., Meyburg, A.H. (eds.): Second International Conference on Behavioural Travel Demand: Lexington Books. Lexington, Massachusetts (1976)
Strategic Highway Research Program II, Travel time reliability 2030: a reliability project brief. Transportation Research Board. www.trb.org (2013)
Strategic Highway Research Program II, Partnership to Develop an Integrated, Advanced Travel Demand Model and a Fine-grained, Time-Sensitive Network. Project C10(A and B). http://apps.trb.org/cmsfeed/trbnetprojectdisplay.asp?projectid=2829 (2013)
Supreme Court of the State of New York, New York County, Part 45, Syncora Guarantee Inc v. Alinda Capital Partners LLC et. al. Docket # 651258/2012, Motion Seq. No. 002. http://www.texasturf.org/2012-06-01-03-09-30/latest-news/public-private-partnerships/401-fraud-lawsuit-against-toll-giant-macquarie (2013)
Talvitie, A.: Things planners believe in, and things they deny. Transportation **24**(1), 1–31 (1997)
Transportation Research Board. Issues in Behavioural Demand Modeling and the Evaluation of Travel Time. Special Report 149, Transportation Research Board, Washington, D.C. www.trb.org (1973)
Travel Model Improvement Program. Travel model reasonableness checking manual, Second Edition, Federal Highway Administration. www.fhwa.dot.gov (2010)
van Vuren T. Be humble. Keynote address to German transport officials. As reported in Local Transport Today. www.transxtera.com (2013)
Vassallo, J. M.:Why traffic forecasts in PPP contracts are often overestimated? EIB University Research Sponsorship Program, EIB, Luxembourg. (Madrid: EIB, Polytechnic University of Madrid, Spain) (2007)
Wachs, M.: When planners lie with numbers. J. Am. Plan. Assoc. **55**, 476–479 (1989)
Wachs, M.: Ethics and advocacy in forecasting for public policy. Bus Prof Ethics J **9**, 1–2 (1990)
Wachs, M. Turning cities inside out: transportation and the resurgence of downtowns in North America. Transportation 40.6, (draft of 6-18/13) (2013)
Welde, M., Odeck, J.: Do planners get it right? the accuracy of travel demand forecasting in Norway. Eur. J. Transp. Infrastructure Res. Issue **11**(1), 80–95 (2011)
World Bank. Successes and failures of PPP projects (The World Bank, Europe and Central Asia Region) (2008)
Worsley, J. Multiple ways to model transport's impact on the wider economy. Remarks to Modeling World. As reported in Local Transport Today, (2013)
Zhao, Y., Kockelman, K.M.: The propagation of uncertainty through travel demand models: an exploratory analysis. Annals Reg Sci **36**(1), 145–163 (2002)

Author Biography

David T. Hartgen is Emeritus Professor of Transportation at University of North Carolina at Charlotte, President of The Hartgen Group, and US Co-Editor of *Transportation*.