2019-04-16

Workshop Summary: Progress in Improving Travel Forecasting Accuracy
Session 1061: Transportation Research Board Annual Meeting, Sunday January 13, 2019
Washington DC, Convention Center.

Introduction: *David T. Hartgen, The Hartgen Group*

The purpose of this Workshop is to document progress in improving the accuracy of travel model forecasts, and to suggest ways to account for uncertainty in traffic forecasts. The Workshop is sponsored by several TRB Committees: Travel Behavior and Values (ADB10), Travel Demand Forecasting (ADB40) and Transportation Planning Applications (ADB50). The support of these committees and TRB staff is gratefully appreciated.

Because of heavy snowfall and the ongoing government shutdown, several scheduled participants (including the organizer) were unable to attend. Many thanks are provided to numerous colleagues, noted below, for stepping up on short notice to ensure a successful Workshop. Thanks are also due to several audience members for their assistance in providing notes and impressions.

The uncertainty of travel forecasts has long been an issue in transportation planning. Although all forecasts of travel obviously contain uncertainty, its magnitude and causes have not been extensively studied until recently. However, beginning with the work of Martin Wachs and Don Pickrell in the 1980s and 1990s, and continuing with that of Bent Flyvbjerg and Robert Bain in the 1990s and 2000s, many studies have found considerable uncertainty in travel forecasts, to the extent that decisions based on them may be open to challenge. These studies have focused primarily on toll-road forecasts around the world, and transit forecasts in the US; the accuracy of US-based non-toll road forecasts has been unknown. To address this deficiency a number of states (e.g., Florida, Massachusetts, Michigan, Wisconsin, Minnesota, Ohio, Virginia) have reviewed their own prior forecasts and published findings. Additionally, federally-sponsored research (NCHRP 8-110) has recently completed a large national study of US-based traffic forecasts.

This Workshop pulls together these recent findings. In Hour 1, we heard from overseas and US experts regarding the results of recent research, in specific numerical form, for both transit and non-toll highway traffic forecasts. In Hour 2, the Workshop invited the audience to add their own experiences, largely from previously un-reported internal studies. In Hour 3, the Workshop broke into subgroups and developed suggestions for planners and modelers on how to deal with uncertainty in various contexts. This Report summarizes the findings of the Workshop, and provides background materials used by the participants. It is hoped that the Workshop will be repeated in future years, as additional information on this important topic become available. This summarizer (Hartgen) thanks the panelists, moderators and numerous commenters for their participation and input, without which this report would not have been possible.

Hour 1: Recent Progress*: Maren Outwater, RSG, presiding*

European Experience: *Kay Axhausen, ETH Zurich*

Awareness of uncertainty in traffic model forecasts has a long history in Europe, stemming from the seminal work of Flyvbjerg and colleagues in the 1990's. But recent attention has turned to understanding how model calibration – the matching of base-year model outputs with known traffic flows – can be improved, even before forecasts are attempted. This initial step is itself complicated by myriad assumptions and simplifications, including network detail, trip purpose, destination, mode, and assignment parameters, and external/freight flows. Even estimates of 'actual traffic' counts and traits are subject to uncertainty. While 'big data' sources may eventually lead to better estimates of flows, link speed, and travel times, various methods of calibration are expensive and often require numerous unverified assumptions.

Several recent efforts to improve calibration have been made in Hong Kong, Switzerland and Austria. In a Hong Kong study (Wong et. al., 2016), it was found that that model-based network travel speeds decrease as the density of network nodes increases, which may not occur in the real world. Therefore elasticity tests should be performed to determine how to limit this effect. In Switzerland (Rieser et. al., 2018), the development of a national model for external (between municipalities) flows and forecasts led researchers to recommend that all assumptions in model-based forecasts be fully documented, and that easy-to-understand quality measures be presented for a wide range of key assumptions. Austrian research (Sammer et. al., 2012) in discussion since 2012 also recommends full documentation, interval estimates, and easy-to-understand quality measures, e.g. pseudo-RSQ, but goes further to recommend full open access to model input data.

These studies also highlight likely future issues, including how to provide open access to essentially proprietary models, cost of model development, realistic expectations for accuracy, and cost versus benefits of more accurate forecasts.

An FTA Forecast Reviewer Perspective on Forecasting Accuracy: *Ken Cervenka, Federal Transit Administration, remarks delivered by Gregory T. Giaimo, Ohio DOT.*

FTA cares about travel forecasts because they provide key information regarding investment decisions, and forecast accuracy increases public credibility in transit investments. Over the past three decades (in 1990, 2003, and 2008), FTA has issued three reports comparing predicted and actual ridership for major 'New Starts' investments. The first study, in 1990, found that 10 projects' actual ridership was, on average, only 42% of predicted ridership. The 2003 study, of 19 additional projects, found actual ridership 69% of predicted, while the 2008 study, of 18 additional projects, reported actual ridership at 74.5% of predicted. Since 2001 FTA has required before-and-after reports for individual projects; the 19 projects evaluated so far report, on average, actual ridership at 89% of predicted. This continuing series indicates that, on average, project ridership forecasts have become more accurate over time. A fourth review report is now underway.

Recommendations from these reviews include the following good-practice procedures:

1. When models are being constructed, pay attention to the model's ability to predict district-to-district person travel flows in an observable year, where data sources useful for validation include the Census-based journey-to-work flows (CTPP) and information from detailed transit rider surveys that enable examinations by purpose, mode of access, and socio-economic group.
2. Focus on the accuracy of existing and predicted model inputs, particularly networks and demographics. Be cautious about overly-optimistic assumptions.
3. Investigate why a 'calibrated' model may not appear to 'validate' well.
4. Study the plausibility of a forecast: is it realistic, given local parameters and similar implementations in other corridors or regions? What is the coherent story about the forecast?

FTA focuses heavily on the plausibility of current-year 'build' and 'no-build' forecasts, and how far-horizon forecasts, if prepared, compare to the current year. The local-area analysis should provide a sensible basis for the forecast, and avoid overly-optimistic projections. FTA is also supporting the use of STOPS (Simplified Trips-on-Project Software), which is sometimes used as an agency's primary ridership forecasting tool, and other times as part of a quality control process for making comparisons against local model-based applications.

A good predicted-actual analysis includes more than just a "percent error" comparison of weekday predicted and actual transit passenger trips on a project, and should dig into what a model output got right and wrong about changes in the region and corridor, district-to-district transit rider flows, and passenger activity at individual stations or station groups.

Transit Forecasting Accuracy: *David Schmitt, Connectics Transportation Group, remarks delivered by Kyeongsu Kim, CTG*

This presentation reports on a personal effort to track the accuracy of ridership forecasts for 140 major transit projects implemented over 5 decades. The effort is unrelated to the FTA's series described earlier. Its major sources are FTA's prior studies, public records from the TIGER and BUILD grant proposals, academic and practitioner accuracy reports, and transit-related websites. In total, this review contains 274 separate forecasts (~2 per project) and 415 separate estimates of actual ridership (~3 per project). Twenty-nine states are represented. 'Accuracy' is measured by the ratio of actual-to-forecasted ridership.

Overall results show that, on average, actual transit ridership was about 72% of forecasted ridership. Only 1/6 of projects under-forecasted their ridership, while about 1/3 of projects over-forecasted ridership by 50% or more. Perhaps surprisingly, more detailed findings show no difference in accuracy for various modes, CBD-serving vs non-CBD projects, starter vs expansion projects, or smaller vs larger projects (except for projects with ridership > 30K daily, which are less accurate). Numerous external assumptions were found to lead to over-forecasting ridership, particularly roadway congestion (100% of cases over-forecast), macro-economic conditions (87% of cases over-forecast), employment estimates (67% of cases over-forecast), supporting transit network (66% of cases over-forecast), project service levels (56% of cases over-forecast), and population estimates (53% of cases over-forecast). Clearly, input inaccuracy contributes significantly to forecast inaccuracy. Time-based plots indicate that overall accuracy has improved since 2007, a finding consistent with FTA's assessments. Project size has also declined, so over time, the 20-year rolling average error in ridership forecasts has dropped from about 40,000 boardings/day to about 5000 boardings/day.

The analysis concludes that transit forecast accuracy seems acceptable for capacity decisions, but is still optimistically biased for cost-benefit decisions. Inputs are a contributing cause of inaccuracy, but accuracy seems to be improving over time.

Traffic Forecasting Accuracy Assessment, Interim Findings from NCHRP 8-110*: Greg Erhardt and Jawad Hoque, University of Kentucky*

A significant knowledge gap in US travel demand modeling is the unknown accuracy of US urban road traffic forecasts. Recognizing this deficiency, the National Cooperative Highway Research Program (NCHRP) initiated a study, NCHRP 08-110, to analyze and improve the accuracy, reliability and utility of project-level traffic forecasts. This presentation reports on the interim findings of the study; a full report is expected in mid-2019.

To begin the study, detailed data on project traffic forecast accuracy was gathered from 6 states (FL, MA, MI, MN, OH, WI), and 4 European nations (Denmark, Sweden, Norway, United Kingdom). Data for each project/road segment was organized into 'forecast cards' documenting formats, data, and administrative information. Ultimately, the database contained 2300 projects with 16,000 road segments, of which about 1300 projects with 3900 segments had count information, location, traffic estimate and year, count year, etc.

The study approach utilized both 'Large-N' and 'Deep Dive' approaches. In the Large-N analysis, the goal was to determine the distribution of forecast errors, potential sources of bias, improvement if biases are corrected, and recommend procedural changes to better manage data in the future. 'Accuracy' was defined as the percent error between the actual count and the forecast volume, taking forecast volume as the base point, for both segments and projects. For 1291 projects, the 'mean percent error' was found to be -5.7%, the standard deviation of percent error about 25%, and the mean absolute percent error 17%. This means that, on average, a traffic forecast is likely to be higher than the later-in-time actual count by 5.7%. For 3912 road segments, the mean percent error was found to 0.6%, the standard deviation of percent error about 42%, and the mean absolute percent error 25%. Regressing actual count against forecast, the equation was found to be: *Actual = 37 + 0.94 Forecast*. Similar equations were developed for quintiles of the data, e.g. high (80^{th} percentile) and low (20^{th} percentile) error ranges. The analysis concluded that "95% of forecasts reviewed are accurate to within half a lane" [capacity]; traffic forecasts show a modest [upward] bias, with actual ADT about 6% lower than forecast ADT, and with mean absolute percent error of 25% for segments and 17% for projects. Further, traffic forecasts were found to be more accurate for higher-volume roads, higher functional classes, shorter time horizons, obtained from traffic models rather than trend-lines, actual unemployment rates close to forecast, and more recent forecasts.

In the Deep Dive analysis, the goal was to determine what aspects of traffic forecasts are likely to be inaccurate, and did these errors matter to the project decision. Detailed traffic forecasts were studied for projects in 6 cities (Eastown Road Extension in Lima, Ohio; Indian River Bridge repair in Palm City FL; Central Artery in Boston MA; Cynthiana Bypass in Cynthiana, KY; South Bay Expressway, San Diego, CA; I-41, Brown Co. WI). The study gathered background documents, including prior model runs; investigated potential sources of errors such as employment, population, auto ownership, fuel price, and network speeds; adjusted forecasts by elasticity analysis, and re-ran prior models with updated information. The Lima OH review indicated that "correcting input errors significantly improved forecasts", while The Palm City FL analysis found "very slight improvement" after similar corrections. For the Central Artery, initial forecasts were originally off by only 4% on existing links and 16% for new links, and were "slightly improved" after correcting input errors. For the Cynthiana Bypass, the key factor was a significant (43%) over-estimate of external (outside the region) traffic, which when corrected "reduced [{forecast] error significantly". For the South Bay Expressway, a national recession, decrease in [Mexican] border traffic and a toll increase were identified as key factors. For I-41, accuracy was improved after correcting for exogenous [external] population forecasts. The Deep Dive analysis concluded that the reasons for forecast inaccuracy are many, yet forecasts of population, employment and fuel price often contribute, as do external travel and travel speed assumptions. Better archiving of models and documentation are also needed.

The presentation concluded:

1. Use a travel model, along with professional judgement, when accuracy is a concern.
2. Understand the key travel markets using the project.
3. Use quartile regression models to estimate the range of uncertainty in the forecast and the accompanying risk.

4. Archive forecasts, documentation, and assumptions, ideally making the forecast reproducible later.
5. Use the results to improve your own model, particularly on its ability to predict change.
6. Providing a 'range of forecasts' is more likely to be 'right' than a single forecast.

Comments on Transit Forecasting Accuracy: *Maren Outwater, RSG Consulting*

The prior presentations indicate that the profession has improved the accuracy of transit ridership forecasts, but we are still outside acceptable bounds. Forecasts can be improved by the use of before-and-after studies, reduced bias in socioeconomic inputs, and sensitivity tests to determine the impact of forecast assumptions. However, it is likely that forecasts will remain optimistically biased, due to assumptions about socioeconomic growth, network, and pricing. The use of observed data and parallel models can improve understanding and accuracy of forecasts.

Input biases are well-known. These include assumptions about future local population and employment, and broader macroeconomic conditions. Tests of project forecasts for the near term will not alleviate this bias. Also, optimistic assumptions about the service levels of a particular project, particularly travel times, along with changes in the supporting transit network, competing transit networks, and local roadway congestion, all tend to produce optimistic forecasts. Better data on near-term local road congestion can be used to reduce this bias.

Project pricing, particularly assumptions about fares and auto fuel prices, are another source of potential bias. Such issues are extremely difficult to forecast, and analyze.

To improve accuracy and understanding in near-term forecasts, recommendations include limiting the inaccuracy of socioeconomic inputs, sensitivity testing to understand the impacts of various assumptions, use of recent observed data as a validation of assumptions, and use of different models and comparisons with other similar projects.

Hour 2: Additional Research: *Greg Giaimo, Ohio Department of Transportation, presiding*

In Hour 2 of the Workshop, the audience was invited to participate by two methods. First, audience members were invited to discuss their own experiences with evaluating the accuracy of travel forecasts. The gist of the discussions was that, for near-term forecasts, accuracy could be substantially improved by improving model calibration results, not just for total volumes but also for directional, time-of-day, peak-hour traffic, screen-line, external traffic, along with better information on actual network speeds and travel times. For longer-term forecasts, stronger assessment of population and employment projections was identified as the key input requiring more scrutiny. Also, better assessment of future networks (e.g., future road capacity and speeds) was suggested.

Second, audience members were asked to fill out a short information sheet documenting their experiences, for possible addition to the NCHRP data set. These are included in the attachments to the Workshop, and are summarized in the Table below. A total of 21 full responses were received, representing 7 local governments, 7 consulting groups, 5 universities, and 2 state governments. Generally respondents noted that review of traffic forecasts is quite commonplace, but is often informal; detailed comments suggest that most of the focus is to review the model's base-year calibration (primarily by comparing link volumes, not travel times, speeds, or time-of-day traffic).

Summary	Have you reviewed the	Have you presented a	Do you have a standard	Do you review forecasts	Validated model on

	accuracy of your travel forecasts?	range of forecasts to designers?	forecast reporting template?	made by others?	forecast of change?
Total Resp 21	Total resp 21	Total Resp 19	Total Resp 17	Total Resp 16	Total Resp 17
Local Govt 7	Yes 18	Yes 12	Yes 7	Yes 9	Yes 9
Consultant 7	No 3	No 6	No 8	No 6	No 6
University 5		NA 1	Both 2	Both 1	Both 2
State Govt 2					
Fed Govt 0					
Non-US Govt 0					

Generally, the responses indicated that travel forecasts are often reviewed later for accuracy, typically when a model was being updated, or when a major modeling effort was planned. However, these reviews are generally not publically documented. The most common method of doing this is by comparing link volume forecasts with later traffic counts, and (for calibration) comparing model volume estimates with link-based counts. Forecasts were often found to be high, relative to counts. About 2/3 of respondents used presentation of ranges for a traffic forecast, usually through scenarios intended to be ‘high’ and ‘low’ estimates of local growth or other future circumstance such as driverless-car technology or major fuel tax increases. These estimates put boundaries on the limits of uncertainty of the baseline forecast, but do not estimate the probability of those ranges occurring. Planners and modelers often suggest ranges in their forecasts, but project designers and decision-makers want single-point estimates, not ranges. Federal guidelines (e.g. FHWA, EPA) require single-number traffic forecasts for design purposes. Respondents indicated that their agencies use a variety of formats for reporting traffic estimates, which depend on the circumstances. For instance, a traffic forecast for a major employment center typically has a different format than one for a regional long-range plan, a major corridor study, or a major transit route. Most respondents also indicated that they have an internal process for reviewing forecasts made by others, and most said that they have tried to validate models based on change prediction. The most common method of that step is back-casting.

On balance, even though the audience survey was limited in response and scope, it seems to indicate that informal attention is being paid to model forecast accuracy, but that resistance to ranges by decision-makers and designers has slowed the formal implementation of uncertainty considerations. The ability of a model to provide an objective handle on traffic forecasts would seem to be an elemental requirement to using the tool to determine the likely change in travel for a potential project, based on need, capacity, and public benefits. Federal agencies could take the lead in encouraging states and localities to completely document and archive modeling efforts, so that later reviews of their models for accuracy can be undertaken before approving a model for subsequent use. The accuracy review should go further than a comparison of link counts, but also include review of corridor and area growth assumptions, speeds, directional and time-of-day traffic, screen-lines, peak-hour flows, truck traffic and external traffic where significant. If the forecast involves tolls, significant evaluation of key input parameters such as the value of time by trip purpose also need review. These steps would help state and local governments to prudently document, report, and use the inherent uncertainty in a traffic forecast as a positive feature, rather than a barrier to modeling improvements.

Hour 3: How to Proceed? *Maren Outwater, RSG, presiding*

In the third hour of the Workshop, the discussion focused on potential 'red flags' that might indicate a possible problem in a forecast, and on what actions need to be taken to improve the accuracy of traffic forecasts. The audience broke into three groups, to permit more focused discussion and full participation. The three group are:

Transit Ridership Forecasting
Smaller Region Highway Traffic Forecasting
Larger Region Highway Traffic Forecasting

Transit Ridership Forecasting: *Maren Outwater, RSG, facilitating*

The Transit Ridership Forecasting group recognized that in the US, significant progress has been made through FTA reviews and independent reviews such as Schmitt's, in quantifying the relative error in various ridership forecasts for major transit projects. Further, the overall accuracy of transit ridership forecasts has been substantially improved over the past several decades. The group also recognized FTA's effort, through its transit ridership evaluation tool STOPS, to concentrate on short-term (opening day and shortly thereafter) ridership estimates, rather than longer-range forecasts. To improve these estimates, the group recommended focusing on the accuracy of existing tools, such as STOPS, by using better estimates of existing district-to-district transit and auto flows, ensuring the accuracy of transit and auto network characteristics, particularly road congestion. For longer-range forecasts, the group recommended more attention to the overall plausibility of the forecasts and their underlying assumptions such as demographic changes in the corridor and region Also recommended were the use of 'ensemble modeling' (using a variety of models to converge on an estimate), conducting sensitivity tests of key input parameters, and developing open-source access to STOPS.

Such steps of course do not alleviate the need to develop more realistic longer-term forecasts, without which the overall benefit/cost estimate for a project cannot be calculated. They also do not address other specialized forecasting needs, such as route-level bus ridership, time-of-day or directional forecasts, station boarding and alighting, holiday or event forecasting, or related transit-access modes such as walk-up, park-and-ride, etc. These sub-classes of transit forecasts have not been extensively reviewed for accuracy.

Smaller Region Highway Traffic Forecasting*: John Miller, Virginia DOT, facilitating*

Although no specific definition of 'smaller region' was provided, the audience generally understood that (with some exceptions) smaller metropolitan regions are those with less than 200,000 urbanized area population, bus-focused transit services, less severe traffic congestion, slower or even negative growth rates, higher proportion of inter-city (external) travel, lacking toll road or managed lane systems, less sophisticated traffic models, and usually in attainment of air quality standards.

The group recognized the substantial work of the NCHRP research team for determining, for the first time, estimates of traffic forecast accuracy for US regions. The group also agreed with the NCHRP research team that traffic forecast errors are often due to errors in socioeconomic forecasts.

Regarding potential 'red flags' that might lead to a review of forecasts, several were suggested. One example would be a case in which a base-year model estimate of traffic is insufficiently close to observed counts (perhaps, as determined by NCHRP's error band chart in Report 765). Another case would be a large-scale predicted change in localized-area population or employment, such as from an announced new major industrial plant.

Smaller regions generally do not have sufficient expert staff to accurately estimate growth rates for corridors or districts/zones, nor do they have the financial resources needed to hire contractors to do the work. Traffic forecasting for smaller regions (and for major projects) is often conducted by state DOTs, consultants, or others rather than local MPO staffs. Further, the region may have experienced few, if any, major road widenings or additions in recent years, so few cases are available for review even if staff were available. The mix of projects also tends to be focused on minor widenings, intersection treatments, and similar projects for which model-based traffic forecasts are unavailable. These circumstances may worsen if models become more sophisticated and use more finely-detailed data such as cell-phone origins and destinations by time-of-day. Further, improved traffic forecast accuracy may not be cost-effective, particularly if conformity-based or NEPA modeling requirements for in-attainment regions are unchanged or relaxed. Even if errors in forecasts are found, if relative across all alternatives the findings may be problematic. As a result, rarely is the question asked, “So, how accurate have our 20-year forecasts been up to this point?”

No one can argue against improving highway traffic forecasts in general. For smaller regions, however, the techniques required to make these improvements need to be commensurate with capabilities of the region's modeling staff. Further, the resources required to make these improvements, in terms of staff time or out of pocket costs, need to match the expected benefits of more accurate modeling.

Larger Region Highway Traffic Forecasting: *Greg Erhardt, University of Kentucky, facilitating*

As noted above, larger metropolitan regions are generally understood to be those with urbanized area populations greater than 200,000 persons, more extensive transit systems (sometimes with fixed route services), modest or even rapid growth rates, more severe traffic congestion, toll road or managed-lane systems, lower proportion of intercity (external) traffic, more sophisticated traffic and transit modeling, and (sometimes) in prior non-attainment for air quality standards (‘maintenance’ status). They also typically have larger MPO staffs and planning budgets, and regularly do their own traffic modeling.

The group discussed numerous issues, including trade-offs between spending more resources on modeling refinements versus the increase in accuracy from, say, better determining trip purpose.

The group recommended a significant improvement in better and more objective forecast input data, particularly for population, employment, AADT estimates and model inputs such as network congestion, travel times, trip generation rates, district-to-district flows, modal and route diversion parameters, value of time, toll rates, etc. Forecasts of regional population should be prepared by independent groups such as state agencies, as is done for example in Washington State, whereas local governments are thought best to determine sub-regional growth trends. These actions would reduce, but not eliminate, the tendency of local governments to over-forecast future growth.

The use of ranges for forecasts might be counter-productive since most traffic forecasts are adjusted, post-model, for corridor balancing or for micro-simulation of traffic flow. Requiring the use of ranges in traffic forecasts would substantially increase the cost and effort needed for project forecasts. However, better understanding of the assumptions that underlie post-processing, as suggested in NCHRP 765, is also needed. As an alternative, the group recommended scenario and elasticity testing, to determine the likely impact of alternative growth assumptions (although this would not determine the likelihood of various future scenarios).

Regarding cost-effectiveness of improving accuracy, the group noted that improving accuracy while ideal in theory, may not always be worth the effort. The assumption here seemed to be that, traffic forecasts were generally accurate enough, particularly for short-range forecasts, and that longer-range forecasts, for instance for build-no-build decisions, comparative differences would be sufficient.

However, the group understood the need for improved accuracy, if for no other reason than to improve public confidence in major decisions. It therefore suggested standardized documentation and archiving of traffic forecasts, so that later studies of accuracy could be made. Documentation should include not just traffic counts and AADTs, but also OD movements, speeds and directional flows, all of which are more possible now using cell-phone data and other sources.

Standards for accuracy should also be developed, for instance using NCHRP 765's estimates for permissible error in calibration and forecasts. But standards need to recognize local context, particularly the magnitude of traffic volume and the time-line horizon. For instance, a forecast for a major high-volume road upon opening would be expected to be more accurate than a forecast for 20 years in the future. For very large regions with major transit capacity issues, the use of tools such as Fast-Trips is useful in modeling capacity-related transit forecasts. Systematic and regular (5, 10, and 20 years) reviews of prior socioeconomic forecasts and traffic forecasts for major projects as well as regional trends should be used to establish differences between observed and predicted values. This helps to establish reference cases that can form the basis for further model improvements. Alternative forecasting methods, such as reference class analysis of similar projects in other regions, should also be used to develop reasonableness checks on model forecasts.