Travel Model Validation Practices Peer Exchange White Paper

December 18, 2008

Helping Agencies Improve Their Planning Analysis Techniques

TMIP

Travel Model Improvement Program

Travel Model Validation Practices Peer Exchange White Paper

Original: June 10, 2008
Final: December 18, 2008

Prepared by:
Cambridge Systematics, Inc.

Prepared for:

Federal Highway Administration

Table of Contents

List of Tables

List of Exhibits

List of Appendices

1. Peer Exchange Model Validation Resource Paper
2. ARC Activity-Based Model Validation Presentation for May 9, 2008 Model Validation Practices Peer Exchange
3. DRCOG Activity-Based Model Validation Handout for May 9, 2008 Model Validation Practices Peer Exchange

Travel Model Validation Practices
Peer Exchange White Paper

■ Introduction

Purpose of Peer Exchange on Travel Model Validation Practices

A Peer Exchange on Travel Model Validation Practices was held in Washington, D.C. on May 9, 2008. The general objective of the Peer Exchange was to provide background information and guidance for the improvement of travel model validation practices. The exchange was sponsored by the Federal Highway Administration (FHWA) Travel Model Improvement Program (TMIP) in response to the recent Transportation Research Board (TRB) Special Report 288, *Metropolitan Travel Forecasting, Current Practice and Future Direction* (*SR 288*) [1] which outlined model validation issues in the summary (*emphasis added*):

> **Validation Errors:** Validating the ability of a model to predict future behavior requires comparing its predictions with information other than that used in estimating the model. Perceived problems with model validation include *insufficient emphasis and effort* focused on the validation phase, the *unavailability of accurate and current data* for validation purposes, and the *lack of necessary documentation*. The survey of [Metropolitan Planning Organizations] MPOs conducted for this study found that validation is hampered by a *dearth of independent data sources*.

An unrelated TMIP e-mail exchange in early 2008 on "Correlation Coefficient versus Coefficient of Determination" also illustrates issues regarding the role of statistics in model validation. The initial question posed in the e-mail exchange was whether R or R^2 of modeled versus observed traffic counts should be used for validation. The original exchange referenced the document, *Calibration and Adjustment of System Planning Models* (FHWA-ED-90-015), dated December 1990, which suggested that "'correlation coefficient' is defined as an acceptable target for model accuracy; that being typically 'greater than 0.88' (page 35)." The original e-mail generated 13 responses, many of which noted that an R or R^2 value of 0.88 was an arbitrary value or achieving such a standard was neither necessary nor sufficient in establishing the validity of a travel model.

Federal Transit Administration (FTA) efforts to improve model validation for New Starts project ridership forecasts also bring a renewed focus on current validation practices. The FTA emphasizes the use of models that are tested rigorously against current transit

---

[1] Special Report 288, *Metropolitan Travel Forecasting, Current Practice and Future Direction*, Transportation Research Board, 2007.

ridership patterns. This is designed to ensure that model forecasts serve as a useful basis for quantifying and understanding the user benefits from proposed New Starts projects that are competing for limited discretionary Federal funds. As presented in the FTA New Starts Workshop in St. Louis in 2007, the implications of careful calibration and validation methods are threefold: first, they require better current data; second, they call for a focus on transit markets; and third, they require better tests and standards[2].

The FHWA is working to improve modeling and travel forecasting procedures through several efforts being conducted under the TMIP. One effort performed in parallel to the Model Validation Practices Peer Exchange was the development of a webinar on Project Level Forecasting.[3] The webinar took place on September 25, 2008, several months after the Peer Exchange. Procedures documented in National Cooperative Highway Research Program (NCHRP) Report 255, *Highway Traffic Data for Urbanized Area Project Planning and Design*[4], were discussed. That 25 year old report is still often used as a reference for travel model validation and for adjusting travel forecasts for roadway project planning and design. Since the webinar included a discussion of the need and desire for updating NCHRP Report 255, coordination efforts are required to ensure that any updated recommendations regarding model validation and the development of project level forecasts based on the Peer Exchange are complimentary and used in the correct contexts.

With the above in mind, it is hoped that the ultimate product of the Peer Exchange on Travel Model Validation Practices will be practical advice and priorities to agencies on:

- Types of model validation to be considered, including temporal considerations (forecasting and backcasting), common sense examination (telling a coherent story), and validation of each model component in addition to the model as a whole;
- Factors (e.g., data) to be considered for proper mathematical and statistics tests;
- The need for and value of local or national validation standards for specific validation tests;
- Prioritization of model validation efforts considering scarcity of resources; and
- Proper documentation of model validation efforts.

---

[2] *Session 5, Calibration/Validation*, Federal Transportation Administration, http://www.fta.dot.gov/planning/newstarts/planning_environment_7276.html. Accessed September 2, 2008.

[3] *Project Level Forecasting*, a TMIP Webinar offered on September 25, 2008, summary materials available at http://tmip.fhwa.dot.gov/discussions/webinars/archive/09252008_plf/. Accessed October 3, 2008.

[4] NCHRP Report 255, *Highway Traffic Data for Urbanized Area Project Planning and Design*, Transportation Research Board, 1982.

Peer Exchange Participants

Representatives from the Federal Government, state Departments of Transportation, regional councils of governments, academia, and the consulting industry were invited to participate in the Peer Exchange (Table 1). All invited participants were familiar with travel model theory, travel model validation, and travel model application for planning purposes.

Table 1. May 9, 2008 Peer Exchange Participants

Attendee	Agency/Firm
Public Agencies	
Charles Baber	Baltimore Metropolitan Council
Greg Giaimo	Ohio Department of Transportation
Bruce Griesenbeck	Sacramento Area Council of Governments (SACOG)
Vladimir Livshits	Maricopa Association of Governments (MAG)
Arash Mirzaei	North Central Texas Council of Governments (NCTCOG)
Guy Rousseau	Atlanta Regional Commission (ARC)
Erik Sabina	Denver Regional Council of Governments (DRCOG)
Dick Walker	Portland METRO
Shuming Yan	Washington State Department of Transportation
Academia	
Alan Horowitz	University of Wisconsin – Milwaukee
Frank Koppelman	Northwestern University
Consultants	
Bill Davidson	PB Americas
Paul Hershkowitz	Wilbur Smith Associates
Ken Kaltenbach	Corradino Group
David Schmitt	AECOM
Federal Agencies	
Ken Cervenka	FTA
Fred Ducca	FHWA
Brian Gardner	FHWA
Supin Yoder	FHWA
Peer Exchange Staffing	
Daniel Goldfarb	Cambridge Systematics
David Kurth	Cambridge Systematics
Laura McWethy	Cambridge Systematics
Tom Rossi	Cambridge Systematics
Sarah Sun	FHWA
Penelope Weinberger	Texas Transportation Institute

Purpose of White Paper

This white paper summarizes the May 9 Peer Exchange. It is intended to provide a framework for a planned update to the TMIP *Model Validation and Reasonableness Checking Manual* using the comments and recommendations from Peer Exchange participants. To that end, this white paper includes examples of practices to clarify, demonstrate, and underscore the comments and recommendations of Peer Exchange participants.[5]

■ Current Travel Model Validation Practices

Review of Current Practices

Resource Paper Summary

A brief summary of commonly referenced documents regarding travel model validation was provided to Peer Exchange participants prior to the Peer Exchange, including *TRB Special Report 288*, the TMIP *Model Validation and Reasonableness Checking Manual*, the Ohio Department of Transportation (ODOT) *Traffic Assignment Procedures Manual*[6], and *NCHRP Report 365*[7] (see Appendix 1 for the entire resource paper). The resource paper summarized current validation practices from over 40 MPOs. Based on conversations with agencies and reviews of agencies' model documentation, it concluded that few models have been extensively validated or validated using independent data other than traffic or transit boarding counts. Most MPOs focused on validating the trip assignment process without extensively addressing other model components. For MPOs that did validate model components preceding the trip assignment steps, many used the same data for model validation that were used to estimate the models.

There were few commonly agreed upon standards for error ranges other than benchmarks suggested by FHWA, which were frequently misinterpreted as standards. Common traffic assignment validation practices included comparisons of screenline, cordon line, and link-specific volumes to traffic counts, and regional assignment statistics such as R^2, percent root mean squared error, and regional vehicle-miles of travel.

Current Practices Reported by Peer Exchange Participants

Several Peer Exchange participants were asked to prepare presentations regarding model validation practices employed by their MPOs (see Appendix 2). Although time

---

[5] Barton-Aschman Associates, Inc. and Cambridge Systematics, Inc., *Model Validation and Reasonableness Checking Manual*, prepared for Travel Model Improvement Program, 1997.

[6] Giaimo, Gregory, *Travel Demand Forecasting Manual 1 – Traffic Assignment Procedures*, Ohio Department of Transportation, Division of Planning, Office of Technical Services, August 2001.

[7] NCHRP Report 365, *Travel Estimation Techniques for Urban Planning*, Transportation Research Board, 1998.

constraints precluded formal presentations, the participants used their current practices to illustrate points made during the assessment of current model validation practices.

Peer Exchange Assessment of Current Practices

The first portion of the Peer Exchange discussions focused on assessing current travel model validation practices, including why validation is performed, and how validation results are reported. The discussion was allowed to flow in the direction of participants interest.

One concern regarding models was the opinion apparently held by many planners and decision-makers that "A model is validated if it gets the money." Peer Exchange participants noted that this validation goal can result in an unwarranted focus on attaining standards associated with validation. Agencies' real world planning and prioritization efforts require that models be demonstrated to be "valid" in order to be germane. When reviewing the results of travel models, decision-makers and the general public often look for some measure of model accuracy or an assurance that the model provides reliable forecasts. This assurance is frequently provided by comparing highway and transit assignment results with observed traffic volumes or transit boardings.

There was some agreement that setting validation standards for matching traffic counts, transit boardings, and screenline crossings can be a double-edged sword. While standards can be used to help determine relative model accuracy, they also can encourage over-manipulation to meet the standards. This can be especially true if project rankings or construction funds are based on absolute values rather than relative results. While almost any travel model can be manipulated to attain a specified validation standard, it is important to emphasize the use of appropriate methods to meet the standard. Methods used to achieve a reasonable match between modeled and observed traffic volumes can be as important as the reasonableness of the match itself. Therefore, model validation should focus on the acceptability of modeling practices in addition to attaining specified standards. A model validation that matches specified trip assignment standards within a reasonable range using valid modeling procedures is better that a model that matches observed volumes with a tighter tolerance using questionable modeling procedures.

While matching traffic counts and transit volumes was deemed necessary by some of the participants as a way of validating models for decision-makers, others believed the focus on reproducing observed traffic and transit volumes perpetuates the "myths" of the predictive power of models and the relevancy of model results. Current validation practices focus on traffic count data which are not absolute. Traffic flows vary from day to day and season to season; matching a set of estimated daily or peak-period traffic flows more closely than their known variability should not be expected. Reasonably matching estimated traffic flows is only one measure of the validity of a model. Equally important are the validity of input socioeconomic data and networks, the reasonability of model parameters, and the validity of the modeling process preceding trip assignment. Furthermore, travel model validation should focus on the sensitivities of the models to variables and policies that affect traffic.

Model validation and the modeling techniques used to obtain specified validation standards are context specific. For example, NCHRP 255 includes procedures for adjusting regional travel forecast results to achieve validation standards for project level forecasts. The techniques are well established and have been used in many regions. However, the techniques are designed for project level forecasts and are not generally applicable for regional modeling.

Validation guidelines should consider the "business processes"[8] used by agencies requiring travel forecasts. Examples of business processes range from evaluating alternative land use policies using gross measures such as changes in daily vehicle-miles of travel (VMT), to evaluating air quality conformity compliance using more detailed measures such as time-of-day specific VMT and speeds, to detailed project planning requiring the use design-hour volume estimates. The impact of business practice on validation was summarized by one participant as, "How valid is a 24-hour traffic assignment validation when people [management] are asking the length of the peak period?"

Business process also can be a consideration for agencies with similar charges. All MPOs may use model results to help define their Regional Transportation Plans or their Transportation Improvement Programs (TIP). However, many small MPOs focus primarily on short-term capital-intensive changes to their transportation system, such as major intersection upgrades, capacity improvements to arterials, or new arterial construction while large MPOs focus on longer term system changes. The business process of smaller MPOs may require a focus on accurate short-term forecasts while reasonable long-term model sensitivities may be required to address the business process concerns for larger MPOs.

The differences in business process may be seen in the design of models for different sized MPOs. The Lincoln, Nebraska MPO travel model uses a simple mode split procedure with static nonauto shares based on 2001 National Household Travel Survey (NHTS) data. By design, the Lincoln MPO uses their model to focus on roadway improvements. In contrast, large MPOs have invested in detailed mode choice models or activity-based models to respond to larger system issues such as major roadway capacity improvements, New Starts analyses, and congestion pricing.

There was recognition that model validation needs can change according to the intended use of the model. For example, consider an agency that currently employs a travel model validated using relatively low and stable travel costs, and is tasked with evaluating a new congestion pricing policy. While travel cost may be considered in the model, it may be insufficient in effectively evaluating the policy's effects. Likewise, models may have been developed using assumptions that are inconsistent with new issues or policies. For example, model assumptions could include using free-flow auto travel times for path-building and path-skimming required as input to trip distribution and mode choice models. While the models would produce forecasts, they would be insensitive to increasing traffic congestion or congestion pricing.

---

[8] Business process refers to the charter, charge, or focus of the agency using the results.

Testing a model's sensitivity to new issues or policies might lead to the finding that the model is of limited usefulness for analyzing the impacts of the new issues or policies. This could lead to the suggestion that a new or different model is necessary. A finding that a model is not appropriate for testing certain issues or policies does not necessarily invalidate the model for use within specified constraints (e.g., relatively stable fuel prices). The following quote describes this situation:

> "Essentially, all models are wrong, but some are useful…the practical question is how wrong do they have to be to not be useful."[9]

Documentation of model limitations uncovered during model validation must be carefully written to explain where and when the model is valid.

Outlining the required levels of effort and expectations for model validation are not typically built into project scopes and work plans. This often results in the budgeting of insufficient time and resources for validation, and validations focused on reproducing the only data readily available: original data used to estimate the models, traffic counts, and transit boarding counts. The Peer Exchange participants identified the need for future guidelines to document the resources required (both time and money) and to discuss the efficient use of those resources for model validation.

Outside contractors and peer review panels have been effectively used to examine travel models to ensure they are proper for the intended analyses. While a peer review represents good practice, it is not always an effective tool in judging the validity of a travel model. Most peer reviewers are not provided with the "hands-on" experience typically required to uncover modeling issues. Rather, they typically receive documentation and summary reports for review; these items are generally inadequate for a full review of the travel model. Peer reviewers are subject to selection bias and subjectivity. In addition, peer reviews often take longer than anticipated. The most effective peer reviews are those that are initiated at the very beginning of the model development process so that the panel members are involved in creating the model development plan and making appropriate recommendations. Otherwise, a peer review can simply be a rubber stamp at the end of a model validation effort.

■ "Improved" Validation Practices

Resource Paper Summary

The resource paper for the Peer Exchange (see Appendix 1) provided background on several approaches for improved model validation practices. Common definitions of the

---

[9] George Box, Professor Emeritus of Statistics, University of Wisconsin, as quoted in *Project Traffic Forecasting, NCHRP 255 Review*, by Doug Laird, TMIP Webinar on Project Planning Forecasts, September 25, 2008.

term "validation" were presented along with extensions to the definition, which focused on explaining how or why travelers make their specific travel decisions. Examples of "improved" validation practices and issues to be addressed were presented as a starting point for discussion. These included collecting data to sufficiently test model estimates and results; validation tests that focus on the reasonability of model parameters; sensitivity tests that focus on the reasonableness of the model response to changes in socioeconomic conditions, transportation networks, and other changes; and backcasting or forecasting for temporal validation of travel models.

Summary of Improvements to Model Validation Practices Identified in Assessment of Current Validation Practices

The Assessment of Current Validation Practices identified several specific improvements to model validation practices as summarized below.

- The validity of input socioeconomic data and networks, the reasonability of model parameters, and the validity of the modeling process preceding trip assignment are equal in importance to matching traffic volumes.
- Travel model validation should focus on the sensitivities of the models to variables and policies that affect traffic.
- Validations should be performed to test the models' sensitivities to the issues and policies that were not important or considered when the models were originally estimated or calibrated.
- The use of appropriate modeling methods to meet specified validation standards should be emphasized.
- Validation guidelines should consider the business process used by the agency or agencies requiring travel forecasts.

The above points have been succinctly summarized in the TMIP *Shining a Light Inside the Black Box* webinar series and elsewhere as the need for the travel models to tell a coherent story. This modeling objective helps ensure that the various parameters, constants, coding conventions and other decision rules in the models reasonably describe travel behavior and can be used to logically explain the properties of models to the general public.

Peer Exchange Discussion Regarding Improved Model Validation Practices

Frank Koppelman summarized the general Peer Exchange discussions into the following definition for improved model validation at a level of detail needed to support public decision-making:

> **Steps to verify the ability of the model system to make reasonable predictions over a range of development patterns, transportation operations, and external factors.**

While the above is a succinct definition for improved model validation, comments made by Peer Exchange participants provided the philosophical underpinning and detail for the definition. The comments have been summarized under general themes, below. Where appropriate, examples have been added to illustrate the concepts.

Model Validation Must Be Context Specific

Model validation must be driven by the intended use of the models. Peer exchange participants emphasized that the uses of the models should be considered when determining the validation procedures. Examples of different contexts discussed by the Peer Exchange participants are discussed below.

Type of analysis being supported by the model, such as policy analysis or project planning. Model validation required for the analysis of a highway expansion will differ from model validation for policy analyses. Two basic modeling needs, highway system planning and New Starts applications, require the model results to satisfy more rigorous standards regarding their ability to match traffic counts or boarding counts. In this context, the focus of model validation must be whether the model is representing reality (Exhibits 1 and 2 provide illustrations). Proper model sensitivities are important for both project planning and policy testing. However, the need to reproduce observed roadway and transit volumes might be relaxed for models used for policy testing if the relaxation results in increased, but reasonable, model sensitivities.

Policy questions may appear suddenly and often lead to questions regarding model sensitivity. Often, policy questions were not considered when the model was developed, yet the model is expected to respond adequately in answering them. In such cases, model validations demonstrating appropriate sensitivities in response to different scenarios are important. Exhibit 3 illustrates a portion of the planned model sensitivity test for an activity-based model.

An example of the sudden appearance of policy level questions were those raised in the summer of 2008 in response to sudden increases in fuel prices. Such questions led to validation concerns regarding the use of fuel prices in models as well as the sensitivities of models to changes in fuel prices. Exhibit 4 is a portion of a New Starts e-mail list exchange regarding the approach one MPO planned in response to policy level questions regarding changes in fuel prices and their model’s fuel price sensitivity.

Exhibit 1. Summary from Ohio Department of Transportation Validation Guidelines

The information on traffic count error is used for an interesting initial check of the traffic assignment process. Specifically, assigned traffic volumes are compared to daily traffic counts for each link with a traffic count and the percent error is calculated. The proportion of links whose percent error exceeds the expected coefficient of variation (shown in Figure B.1) must be reported. Traffic assignments with **less than** 33 percent of the links exceeding the expected coefficient of variation curve are **regarded with caution by ODOT** due to the likelihood of artificial model or assignment adjustments to force agreement with counts.

Once the initial check has been completed, the ODOT manual specifies several validation tests:

- A review of network plots of modeled traffic volumes and traffic counts is recommended as the best check of a traffic assignment. No guidelines or standards are set for this test.
- Modeled to observed traffic volume RMS error is calculated for up to 18 volume groups. The model is deemed to have passed the RMS error test when the percent RMS errors for all volume groups are less than the standards specified in the manual. The manual suggests approaches to improve model results if the validation results are not acceptable.
- Modeled to observed vehicle-miles of travel (VMT) for the region, by functional and administrative class of the roadway, and by geographic ring and sector are determined. The manual states that the modeled regional VMT must be within 3 percent of the VMT based on traffic counts. Guidelines for maximum difference in VMT are set for functional and administrative classes of the roadway and for geographic ring and sector; the manual states that the modeled VMT should fall within the specified error ranges.

Modeled to observed traffic volumes crossing screenlines are determined. Guidelines for the maximum desirable deviation in screenline volumes are set in the manual.

Exhibit 2. Model Validation Excerpt (2002 Travel Demand Forecast Model Calibration Report for Ada and Canyon Counties, Idaho)

EXECUTIVE SUMMARY - COMPASS TRAVEL DEMAND MODEL QUICK FACTS

MODEL GEOGRAPHY (2002 CALIBRATION)

- Ada and Canyon Counties (a.k.a. Treasure Valley, Idaho)
- 480,000 people
- 2,100 centerline miles in the model network classified as a collector or higher
- 534 Traffic Analysis Zones (TAZs) - 346 in Ada County and 188 in Canyon County

METHODOLOGY

The 2002 calibration of COMPASS' travel demand model is a 3-Step model which includes a land use/trip generation module, a gravity based trip distribution, and a capacity constrained equilibrium traffic assignment process.

VALIDATION CRITERIA AND RESULTS

COMPASS' Transportation Model Advisory Committee (TMAC) approved the 2002 calibration of the 24-hour model on June 29, 2004. The peak hour model was subsequently approved on October 12, 2004. The following are the validation criteria and results for the 2002 calibration.

24-Hour Model						
Percent Root Mean Square Error (RMSE) by Functional Class				Percent Error by Functional Class		
Facility Type	%RMSE	MAX	Validation	Volume to Count % Difference	MAX	Validation
Interstate & Ramps	23.9%	< 40%	PASS	-0.2%	< 7%	PASS
Principal Arterials	22.1%	< 40%	PASS	-4.4%	< 10%	PASS
Minor Arterials	39.0%	< 40%	PASS	-11.5%	< 15%	PASS
Collectors	70.4%	< 40%		-16.8%	< 25%	PASS
Locals	82.2%	< 40%		11.6%	< 25%	PASS
Overall	**34.9%**	**< 40%**	**PASS**	-7.6%		
Without Locals	34.5%	< 40%	PASS	R-Squared = 0.90 Correlation Coefficient = 0.96 Screenlines = 85% passed		
Without Collectors and Locals	28.9%	< 40%	PASS			
Peak Hour Model						
Percent Root Mean Square Error (RMSE) by Functional Class				Percent Error by Functional Class		
Facility Type	%RMSE	MAX	Validation	Volume to Count % Difference	MAX	Validation
Interstate & Ramps	28.6%	< 40%	PASS	3.9%	< 7%	PASS
Principal Arterials	25.0%	< 40%	PASS	-1.2%	< 10%	PASS
Minor Arterials	40.9%	< 40%		-20.4%	< 15%	
Collectors	79.5%	< 40%		-25.7%	< 25%	
Locals	82.9%	< 40%		17.8%	< 25%	PASS
Overall	**37.8%**	**< 40%**	**PASS**	-9.5%		
Without Locals	37.4%	< 40%	PASS	R-Squared = 0.87 Correlation Coefficient = 0.94		
Without Collectors and Locals	31.4%	< 40%	PASS			

The 24-hour and peak hour model also went through a "Dynamic Validation" process. This process, documented by Fehr & Peers, evaluates how the model responds to changes in households and employment for selected TAZs. The COMPASS model responded appropriately.

2002 Travel Demand Forecast Model Calibration Report for Ada and Canyon Counties Page 3

Exhibit 3. Excerpt from DRCOG IRM Validation Plan

2007 Short Term Changes Sensitivity Test (Priority 2)

The 2007 transportation system involves several recent major changes: TREX, Southeast LRT, and North I-25 HOV/toll lanes. A 2007 model run would show how these system changes impact model results. 2007 data could be used to reveal how accurately the model predicted the impacts of the system changes. Due to the effort required to generate a synthetic population for the region, the 2005 population synthesis and employment estimates can be used as the basis for the 2007 model run. Of course, the transportation network will need to be updated to reflect 2007 conditions.

2035 RTP Long Term Changes Sensitivity Test (Priority 3)

The 2035 RTP model run will show the impact of long term changes on model outputs. The 2035 IRM model results will be compared to the 2035 regional trip-based model (COMPASS)* results.

FTA New Starts – 2035 TSM & Build (Priority 4)

Two runs of the IRM in accordance with FTA guidelines will be performed for this sensitivity test. This test would require first the run of the TSM to fix the trip table, and then a rerun of the model with the build alternative. This work should be performed for one of RTD's past or existing New Starts applications that has been modeled using COMPASS. FTA New Starts measures such as TSUB should be compared to previous results using the COMPASS forecasts.

One Parameter Sensitivity Test (Priority 5)

The sensitivity of the model to changes in individual variables can provide good insight to the overall sensitivity of the model. The sensitivity of individual model components has been tested by the estimation of the elasticity of those components. However, changes to a single parameter such as the increase of transit fares by 25 percent will provide insights of the sensitivity of the individual model components as well as the sensitivity of the overall IRM.

Source: Parsons Transportation Group, *DRCOG IRM Validation Plan Technical Memorandum*, Draft 2a, September 2007.

Note: The use of the name COMPASS for both the DRCOG trip-based model and Ada and Canyon Counties, Idaho model is coincidental.

Exhibit 4. Excerpt from New Starts E-mail List Exchange on Treatment of Auto Operating Costs in Models

We too have experienced pressure -- in this case, from our MPO board -- to adjust base and future auto operating costs. We have, however, mollified our board (for now) by saying five things to them, some of which have been mentioned by the first responders to this question. Here's what we've said, more or less.

1. We don't know any more about the connection between today's $4.00+ per gallon prices and 2030's real prices than we knew about the connection between last year's $2.00-to-3.00 per gallon prices and those in 2030. In other words, what we're seeing today should not necessarily compel us act hastily and assume increased future real gas prices.

2. It is true that most economists seem to be saying that higher real gas prices are here to stay. Even so, we don't know how technology, driven by market forces and regulation, will compensate for this in the longer-run; however, as other's have pointed out, there's certainly a good chance that it will compensate to the extent that operating costs per mile will stabilize.

3. From the mode choice model's perspective, relative prices among modes are important. Even if we fiddle with different future real gas prices, we must remember that we also have toll and parking charges to consider for auto modes, as well as fares for the transit mode. The rate at which the latter will increase over time, in real terms, will be driven by politics and economics together, and we don't know any more about the future trajectory of real fares than we do about that for auto operating costs.

4. When we do update base-year auto operating and other costs, we'll have to ensure that the model set remains calibrated and valid. Among other things, that will mean that we'll have to have a fairly good set of contemporary roadway volume and transit boarding counts to which to recalibrate, after updating the costs. It would furthermore behoove us to do some back-casting and/or other validation tests to ensure that the implied model set-level (not just mode choice-level) price elasticities yield reasonable real-world results. In any event, our most recent, comprehensive count set is for 2005/6, and given the time it will take to accumulate counts that coincide with the higher gas prices, we require the passage of some more time, and of course, the persistence of these higher prices, before we can mount a serious base-year model redo.

Even at that, everyone will have to understand that we'll be recalibrating and revalidating to a fairly short-run response to higher gas prices. Sure, overall MBTA ridership is up 10 or whatever percent over last year, but how much of that increase will be sustained, once some travelers have the opportunity to change vehicles, trip destinations, household locations, etc., in the mid- to long-run?

5. In the meantime, here's what we will do right now, so we told the board. We'll conduct some quick-and-dirty tests at a broad-brush level to see how the model set handles changes in gas prices. After all, it's been a long time since we zeroed in on that particular facet of the model set, and the household travel survey from which our modeled behavioral responses are derived, is quite dated. If, from this, we think that we have something valid to say, we'll then perform some tests, along the lines that Frank mentioned, in which we determine the extent to which changes in the relationships among real prices might affect ridership levels and benefits on some particular to-be-selected test project. This would certainly be consistent with FTA's guidance on assessing uncertainties in ridership forecasts.

There you have it. The cynic might say that all we've done is to highlight our ignorance and plead for more time. Maybe. But I really think it's important to be very cautious here, even while figuring out how to put some useful information out there for the decision makers. Caution is particularly warranted, in my view, when using traditional, hence limited, four-step models for this kind of thing.

Scope of analysis being supported, such as regional planning, systemwide planning, corridor planning, interchange justification reports, or site impact analyses. Systemwide planning and project level planning require different validation methods. For systemwide planning, validation should combine model sensitivity testing with criteria based on matching absolute numbers, such as independent estimates of VMT at an aggregate level. For project level planning, a tighter focus on matching absolute numbers might be most appropriate. For example, matching time-of-day traffic volumes and traffic speeds might be important validation measures for project level planning.

Planning horizon also is an important scope of analysis consideration. Typically, project level analyses are assigned short-term planning horizons while the systemwide analyses have long-term planning horizons. For the short term, criteria based on matching more detailed absolute numbers such as time-of-day traffic volumes become more important in validation efforts (Exhibit 1), since more detailed decisions regarding a project may be made based on the forecasts. Long-term planning usually focuses on more general goals and objectives regarding regions and the development of programs and allocation of resources to achieve those goals and objectives. Thus, validation should focus on model sensitivity to factors impacting travel decisions and traveler behavior (Exhibit 3).

A full range of types of "absolute numbers" may be considered as validation of travel models moves from long-term, regional planning to short-term, project level planning. For the long-term regional planning context, the absolute numbers considered include regional VMT, regional transit boardings, regional mode shares, and major screenline crossings or district-to-district flows. Model validations for short-term corridor planning, interchange justification reports, and site impact analyses focus on reproducing link specific traffic counts or detailed transit line boardings.

Temporal validations such as forecasting or backcasting can be important for systemwide model validations, particularly if sufficient time or transportation system changes exist between the years selected for the validation. The "2007 Short Term Changes Sensitivity Test" shown in Exhibit 3 is a pertinent illustration of a planned temporal validation process since the model is being developed using 1997 data. As described in the exhibit, there have been substantial changes to the transportation system in addition to the land use changes that occurred between 1997 and 2007. While the 2007 model run will be compared to observed data, the 2035 forecasts described in Exhibit 3 also provide information regarding model sensitivity.

While sensitivity testing is most often considered for long-term planning, it is also an appropriate validation test for short-term planning. Exhibit 5 is an excerpt from a model calibration report prepared for Ada and Canyon Counties, Idaho. The report describes the regional travel model validation results against observed data. The "dynamic validation" described in the report was performed at a small scale. Since the dynamic validation process does not compare model results to observed data, it may be useful to develop expectations of the results prior to applying the model for the scenario. Dynamic validation is relatively easy to perform and can provide valuable information regarding the reasonability of the model for both modelers and decision-makers.

Exhibit 5. "Dynamic" Model Validation Excerpt (2002 Travel Demand Forecast Model Calibration Report for Ada and Canyon Counties, Idaho)

DYNAMIC MODEL VALIDATION RESULTS

Dynamic model validation is a technique to evaluate a model's ability to respond appropriately to various demographic and network changes. Results of the dynamic validation tests are evaluated in terms of meeting an expected trend and magnitude in change. This technique is documented in a white paper by Fehr & Peers and serves as a guide on dynamic validation. COMPASS staff performed three dynamic validation tests on both the 24-hour and peak hour models. Two of the three tests involve evaluating the model's ability to forecast network impacts when the demographic data of a Traffic Analysis Zone (TAZ) are changed. The third test involved the model's response to the deletion of a major link in the regional roadway network. Both models responded appropriately to changes in households, jobs and the network.

24-HOUR MODEL

The 24-hour dynamic validation results are shown in Table 44 and Table 45 for Ada County TAZs 144 and 299. Table 46 and Table 47 summarize the results for Canyon County TAZs 475 and 566. All four TAZs responded as expected to the household and job changes. Figure 43 in Appendix D contains a map highlighting all four TAZs used for the dynamic validation.

Table 44

Results from TAZ 144 (Ada - Suburban)								
Baseline: 236 households				**Baseline: 19 Retail Jobs**				
Change in Households	Change in Vehicle Trips	Vehicle Trips per Household	Change in VMT	Change in Retail Jobs	Change in Vehicle Trips	Vehicle Trips per Retail Job	Change in VMT	VMT per Retail Job
0	0.00	7.31	0.00	0	0.00	20.40	0.00	164.21
-1	-3.91	7.31	-24.13	-1	-1.97	20.40	-9.07	164.21
-10	-54.26	7.31	-402.99	-10	-23.14	20.40	-179.45	164.23
-100	-526.37	7.31	-3,599.25	Non-applicable				
1	7.31	7.31	25.41	1	2.80	20.40	-15.44	164.21
10	49.91	7.31	2.41	10	26.96	20.40	111.97	164.19
100	530.28	7.31	3,415.34	100	258.68	20.37	1,697.84	163.98
1,000	5,262.83	7.30	36,434.54	1,000	2,562.56	20.12	15,855.75	161.90
5,000	26,354.60	7.25	187,525.56	5,000	12,809.78	19.09	84,627.45	153.43
10,000	52,721.12	7.20	380,311.43	10,000	25,617.21	17.97	170,888.47	144.14

Table 45

Results from TAZ 299 (Ada – Rural)								
Baseline: 34 households				**Baseline: 7 Retail Jobs**				
Change in Households	Change in Vehicle Trips	Vehicle Trips per Household	Change in VMT	Change in Retail Jobs	Change in Vehicle Trips	Vehicle Trips per Retail Job	Change in VMT	VMT per Retail Job
0	0.00	7.31	0.00	0	0.00	20.40	0.00	164.21
-1	-4.33	7.31	-93.57	-1	-1.48	20.40	-20.75	164.21
-10	-54.04	7.31	-614.51	Non-applicable				
1	6.54	7.31	66.01	1	2.99	20.40	33.64	164.21
10	51.94	7.31	255.76	10	27.22	20.40	293.29	164.19
100	540.91	7.31	5,494.44	100	257.56	20.37	2,791.30	164.00
1,000	5,381.43	7.30	56,559.80	1,000	2,563.19	20.12	29,830.62	162.12
5,000	26,933.90	7.26	293,003.01	5,000	12,807.61	19.09	148,962.79	154.37
10,000	53,886.61	7.21	586,994.77	10,000	25,618.39	17.97	285,941.85	145.71

2002 Travel Demand Forecast Model Calibration Report for Ada and Canyon Counties Page 43

Business process being supported, such as MPO, regional transit district, or air quality control district. The business process should be considered when determining validation guidelines. As defined previously, the business process relates to the charge of the agency using the travel forecasts. A primary objective for model validation is an improved tool to support decision-making. This requires that decision-makers and other users are satisfied that the model structure and results address their needs. Considering the business process provides a practical context to model validation.

Model Validation Must Consider the Entire Model

Model validation must consider the entire travel model structure and network design. A focus on matching traffic volumes or transit boardings is problematic since modeling problems occurring in the previous steps of the model are often ignored. Focus on the reproduction of observed traffic volumes or transit boardings has led to questionable modeling practices such as unwarranted link-specific speeds or capacities, or buses operating at higher than congested traffic speeds. Many years ago, one practitioner noted, "You can cover a multitude of modeling sins in traffic assignment." Indeed, a Texas Transportation Institute report, *A Sensitivity Evaluation of Traffic Assignment*, found little difference between the assignments of a synthetic trip table and a modeled trip table.[10] This lack of difference was attributed to the "power of the assignment process to mask major inaccuracies" in trip generation and trip distribution.

The model validation process should consider the input data and networks. Models are complex and require a substantial amount of input data for application. A simple model with 2000 zones and 10,000 highway links can easily have 100,000 individual data items. More complex models, such as the tour-based, IRM currently being developed by DRCOG, will include over 250 million data items to describe the regional population. Since data drive the model, quality control procedures used for developing input data should be part of model validation.

One example of quality control is the suggested FTA New Starts process for validating a coded transit network and the transit path-building process. One common method used for this validation is the assignment of observed transit trip tables built from transit on-board survey data to the coded transit network to compare the modeled to observed boardings. With the emerging use of Global Positioning Systems (GPS) in travel surveys, it may soon be possible to perform similar validations of roadway networks using observed path-choice data.

The use of targeted model adjustments, such as link-specific changes to improve traffic assignment validations, is counter-productive. The counter-productiveness is more pronounced as model components become increasingly integrated through the use of feed-back loops and other procedures designed to improve internal consistency. Through feed-back loops, the procedures used to calibrate an individual model component impact model components applied before and after that model component. For example, many

---

[10] *A Sensitivity Evaluation of Traffic Assignment*, Report #17-2, Texas Transportation Institute, 1975.

regions attempt to develop traffic assignment procedures that produce valid traffic speeds concurrently with valid traffic assignments. If the modeling process is used to forecast the impacts of congestion pricing, the methods used to produce valid travel speeds and traffic volumes are important to all steps of the modeling process.

The level of detail of the validation of each component might vary due to data availability or other issues. Model validation should consider the contribution of the individual model component validations to the overall model validation. For example, the validation of trip generation or trip distribution might warrant more emphasis for long term, strategic model uses than the validation of the trip assignment process. In contrast, models used for shorter-term project planning might warrant more resources being allocated to trip assignment validation. The proper allocation of resources to the validation process should consider the likely uses of the model and the payoffs of the validations of specific model components. Failure to validate specific model components due to unavailable data or resources should be documented for every model validation effort.

The Validity of Model Validation Data Must Be Considered

How well the validation data represent reality is a primary validation question. An example of this question is the veracity of traffic counts. Counts are often collected from multiple sources using multiple counting techniques. Counts may be stored as raw counts or factored counts, such as average annual daily traffic (AADT). Developing a validation dataset of average weekday traffic (AWDT) may be difficult due to the different sources, different counting methods (one-day, two-day, permanent traffic recorder), and reporting methods (raw axle counts, raw counts divided by average axle factors, AADT estimated from raw counts). The Southeast Michigan Council of Governments (SEMCOG) has addressed this issue by constructing a detailed traffic count database for the region (see Exhibit 6). Traffic counts from numerous sources are entered into the SEMCOG traffic count database, but each traffic count must pass validation criteria before it is stored as an unadjusted, 24-hour traffic count. Raw traffic counts by time-of-day (15-minute, hourly, etc.) also are stored providing a full 24-hour traffic count was performed. Storing the unadjusted traffic count data makes it possible to incorporate data from multiple years, days of the week, and seasons of the year into a consistent traffic count validation dataset.

Issues and concerns with existing land use data and the viability of those data may not be considered as often as concerns with traffic count data. Yet, land use data can be as variable as traffic counts, and errors in the land use data may ripple and compound throughout the modeling process. Existing trip-based models frequently use households stratified by socioeconomic strata such as income group and household size for each TAZ as input to the trip generation process. No readily available data exist to validate those input data. Census data showing the number of households by income group or by household size for small geographic areas may be acquired, but detailed estimates of households by income group and household size for small geographic areas is rarely available.

Exhibit 6. Description of SEMCOG Regional Traffic Count Database

SEMCOG *. . . Equipping local governments for the future*

Traffic Counts

The Regional Traffic Counts Database (RTCD) is SEMCOG's central repository for all traffic counts. Traffic counts stored in this database were collected and provided to SEMCOG by the Michigan Department of Transportation; by consultants specializing in traffic data collection; and by counties, cities, and villages in Southeast Michigan. Each count was taken during a continuous 24-hour period, beginning on the date indicated. Counts longer than 24 hours were divided into distinct 24-hour intervals and stored in the database as separate records. Query our Regional Traffic Counts Database.

All counts are unadjusted, meaning that no growth factors or seasonal, day-of-week, or axle correction factors have been applied. Please note that some counts are for one direction only (northbound, southbound, eastbound, or westbound), while others are two-way totals. Counts identified as "link" counts were taken approximately midway between the cross streets listed in the From Street and To Street columns; counts identified as "intersection" counts were taken near the cross street listed in the To Street column. SEMCOG periodically updates the database as new counts are received, so roads currently without counts available may have data in the future.

Source: http://www.semcog.org/TrafficCounts.aspx.

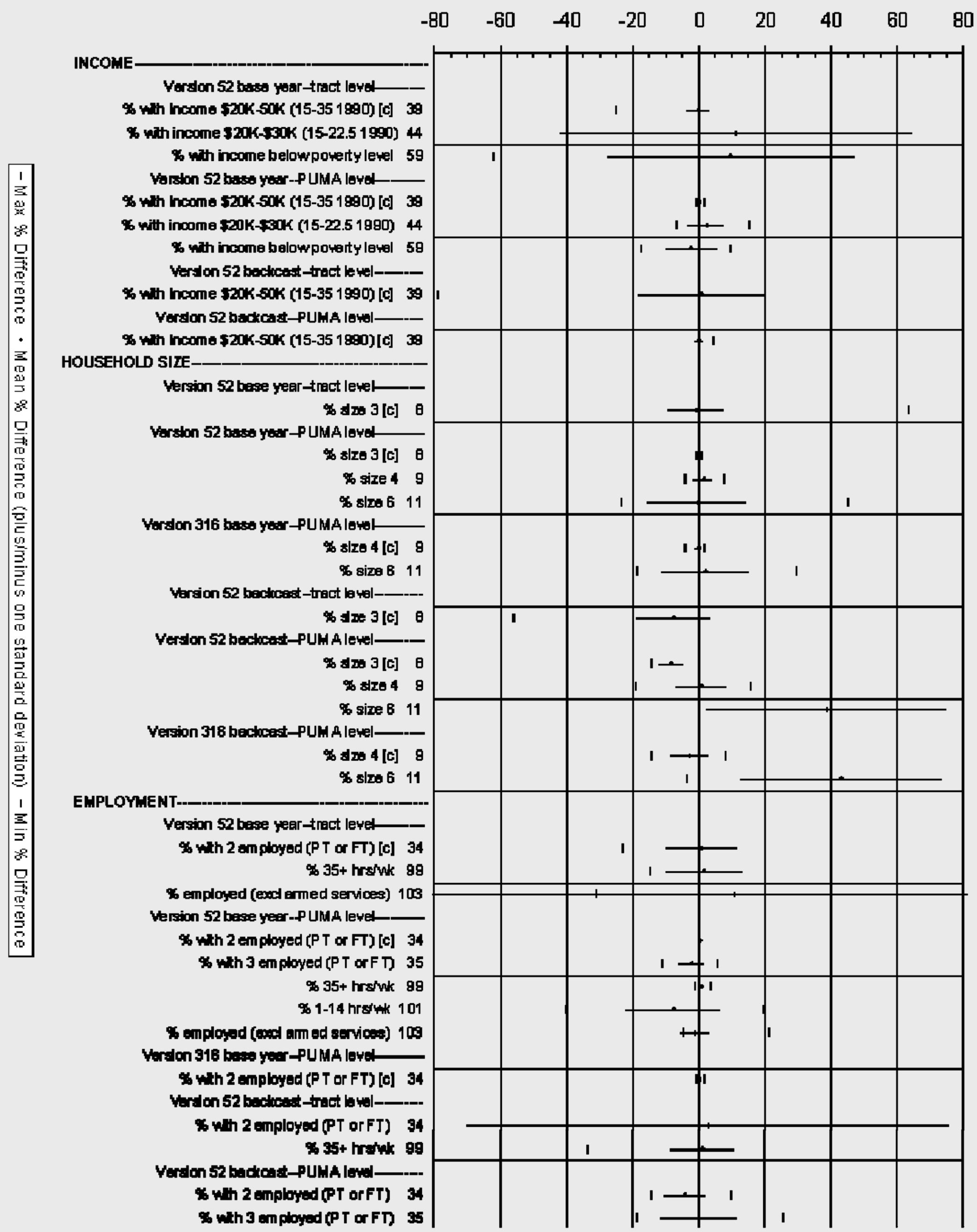
As regional models are increasingly disaggregated into smaller TAZs, parcels (e.g., SACOG SACSIM model), or points (e.g., DRCOG IRM), validation of data may become increasingly difficult or require innovative approaches. Exhibit 7 shows a portion of a table summarizing the validation results for the population synthesizer (PopSyn) developed for the ARC. Model validation was performed by backcasting to 1990 but was performed only for aggregate geographic levels such as tracts, PUMA, or counties.

The same data are often used for model estimation or calibration and model validation. Such a practice does not produce a quality validation. The practice simply replicates the same trends represented in model estimation or calibration and may provide a false sense of security.

Exhibit 7. Portion of ARC PopSyn Model Validation

Figure 1—Selected validation results

([c] indicates variable controlled at TAZ level; numbers after label correspond to variable number in complete validation output)



Source: John Bowman and Guy Rousseau, Validation of the Atlanta (ARC) Population Synthesizer (PopSyn), prepared for TRB Conference on Innovations in Travel Modeling, Austin, Texas, 2006.

Expectations and Results of the Validation Should Be Documented

Guidelines for model validation and validation documentation should be developed as part of the model development plan, and should include specifying a validation plan. As an example, an excerpt from the DRCOG IRM Validation Plan is shown in Exhibit 8. The DRCOG IRM validation plan was written after the model specification plan but prior to the estimation of the model.

A model validation plan should include a listing and assessment of the data available for model validation. Appropriate validation tests can be defined based on the listing and assessment of the available validation data. The listing and assessment can provide direction for the collection of addition validation data. An example of an assessment performed by SACOG is provided in Exhibit 9.

An accurate estimate of the resources needed for model validation is important to help ensure that the time, money, and resources allocated to model validation are used efficiently and effectively. Table 2 shows the results of an informal survey of the Peer Exchange participants regarding their perception of how current resources are allocated to the model development process and how they should be allocated. The following working definitions for model estimation, calibration, and validation were used by participants for their responses to the informal survey:

- **Estimation –** Using statistical analysis techniques and observed data to develop model parameters at a disaggregate level without bias or correction factors.

- **Calibration –** Development of constants and other adjustments to estimated or specified models in an effort to make the models replicate observed data for a base (calibration) year.

- **Validation –** Application of the models and comparison of the results against other data not used for estimation or calibration. Validation also includes evaluating the sensitivity of the model set to changes in input data or assumptions.

Table 2. Informal Survey of Peer Exchange Participants Regarding Allocation of Modeling Resources

Model Component	Typical Resource Allocation (Percent)[a]	Desired Resource Allocation (Percent)[a]
Data collection	30	39
Estimation	27	16
Calibration	19	17
Validation	15	17
Documentation	8	9

[a] Percents do not sum to 100 due to rounding.

Exhibit 8. Excerpt Detailing Planned Validation Tests from DRCOG IRM Validation Plan

Table 3.6 Daily Activity Pattern Model Validation Tests

Aggregation Level	Validation Measures	Expected Outcomes	Priority
Comparison of model parameters to other regions	• Comparison of model coefficients to: • Sacramento • San Francisco • Columbus	• No expectations; comparison only.	Level 1
Disaggregate	• Prediction success of modeled daily activity pattern choices against observed TBI estimation data	• Prediction success likely to be very low	Level 3
Aggregate	• Numbers or percents of residents making tours and intermediate stops by activity type: • For the region • By county • By household size and income group • By household size and auto ownership • By gender and age group • By employment status • By student status	• Compare modeled to expanded observed numbers or percents • Review for reasonable patterns	Level 2
	• Percent of "immobiles" (persons with no out of home activities during the day) by: • By household size and income group • By household size and auto ownership • By gender and age group • By employment status • By student status	• Compare to results summarized by Kay Axhausen (e.g. in *Transportation*, Volume 34, Number 1, January 2007 , pp. 107-128)	

Source: Parsons Transportation Group, *DRCOG IRM Validation Plan Technical Memorandum*, Draft 2a, September 2007.

Exhibit 9. Draft Validation Data Sources Assessment from SACOG (Portion of Entire Assessment Shown)

Monitoring Dataset	Comprehensive-ness	Currency	Consistency	Content	Geographic Scale	Travel Model Validation for…
Locally-Developed GIS						
SACOG Centerline GIS	Regional	Current (Annual Updates)	Well Defined Data Standards	Linear feature alignment; street addresses (by range); limited information on type of feature	Micro	Roadway alignment; potentially roadway distances; walk distances
Census Data						
Year 2000 Short-Form (STF1)	Regional	8 years old, decenial updates	Generally comparable, one decade to next. Thick sample (1:1)	Detailed aggregate demographics, cross-tabs	Very small area (census block)	Aggregate-by-area cross checks on population files and zonal datasets (pers/hh, workers/hh, hh income, auto ownership, etc.)
Year 2000 Long-Form (STF3, CTPP)	Regional	8 years old, decenial updates	Generally comparable, one decade to next. 1:5 sample	Detailed aggregate demographics w/cross tabs; journey-to-work travel data	Small area (block group)	Worker flows; home-workplace distance distribution
ACS	Currently, census places >65k in pop; by 2010, >20k	Rolling 3-year sample	Thin sample each year: 1:20	Detailed aggregate demographics w/cross tabs; journey-to-work travel data	???	TBA, once reported geography gets below places 65k and greater (2009 or 2010)
Travel Surveys						
SACOG Household Travel Survey	Regional	8 years old; 2010 update planned	Not comparable to 1991 survey; very thin sample (1:250)	Detailed disaggregate (person level) data; includes detailed demographic and trip-level information on all purposes (including non-work)	Micro (parcel/point)	Once expanded, limited/weak checks of tour frequency, home-to-tour-destination distance distribution, mode of travel by purpose and person type, etc.
2005 Transit On-Board Survey	All fixed-route operators	3 years old; no update planned	Not comparable to 1999 survey; 1:10 sample	Detailed disaggregate (passenger trip-segment level) data; includes some demographic and whole trip-level information.	Varies; mostly parcel/point locations	Aggregate checks on tour and trip mode choice; aggregate checks on transit assignment (boardings/trip, etc)
National Household Travel Survey	National	2001, update ongoing	Evolving survey instrument, but largely comparable for trend analysis	Reported detailed aggregate travel behaviour, cross-tabulated by demographics, area of residence, etc.	Some states have special add-ons; otherwise, national	Reasonable-ness checks on basic travel behavior (e.g. trips per person, per hh; VMT per person, per hh; etc.
Transportation Network/Supply						
Nat'l Transit Database	Selected fixed route operators	2006, w/annual updates	Generally comparable to prior years	Systemwide supply (revenue miles, revenue hours) by bus vs. rail; weekday vs. weekend values	Operator totals	Aggregate transit network stats by operator
HPMS	Regional	2006 available; annual updates	Generally comparable to prior years	Aggregate network supply (lane miles, centerline miles)	Jurisdiction, unincorporated remainder by county	For freeways, comparable to model network; includes all local streets (not included in model network), so lower level capacity classes not comparable to model.
Transportation System Utilization						
Traffic Counts	Regional, but very spotty locations	2005; 2008 ongoing	Comparable to some locations in 2000	Varies: all include typical weekday totals; most include hourly volumes by direction; quality varies, too: some counts are robust averages, some are single counts	Very spotty (e.g. about 1,000 locations, with some jurisdiction uncounted)	Aggregate traffic assignment validation, by: functional class of roadway, time period, link volume group
Transit Line Counts	All fixed-route operators	2005; 2008 ongoing	Comparable to 2000 data	Weekday averages for all in 2005	n/a	Aggregate transit assignment validation: daily boardings by line, operator, service type.
LRT Station Boardings	All LRT Stations	2005; 2008 ongoing	Comparable to 2000 data	Spring/fall weekday averages for 2005, by RT service period	By station	Aggregate station boardings, daily and by time period.
Park-and-Ride Lot Occupancy	All LRT lots; some other lots	2005; 2008 ongoing	Available back to ???	RT collects monthly; SACOG uses spring/fall	By station	Peak park-and-ride demand by station
HPMS	Regional	2006 available; annual updates	Generally comparable to prior years; questions as to the frequency of local jurisdiction counts/volumes	VMT aggregated to roadway class	Jurisdiction, unincorporated remainder by county	VMT by county
Nat'l Transit Database	Selected fixed route operators	2006, w/annual updates	Generally comparable to prior years	Systemwide ridership (boardings, passenger miles, passenger hours) by bus vs. rail; weekday vs. weekend values	Operator totals	Aggregate ridership statistics by operator

Table 2 suggests that Peer Exchange participants favored almost equal allocation of resources for model estimation, calibration, and validation. One reason for the allocation was provided by one participant's comment regarding his allocation. Specifically, while model estimation, calibration, and validation were presumed to be mutually exclusive steps for the informal survey, he viewed the three steps as an iterative process.

Table 2 suggests that the typical resource allocation to validation (15 percent) is reasonably close to the desired resource allocation (17 percent). The allocation of a fixed budget using the typical and desired allocations might provide a different picture. Table 3 shows the allocations of a hypothetical $3,000,000 budget for model development using the typical and desired resource allocations. As can be seen in Table 3, the desired resource allocation places less importance on model estimation and calibration, and more importance on all other model development efforts.

Table 3. Implied Budgets Based on Survey of Peer Exchange Participants Regarding Allocation of Modeling Resources

	Component Budgets Based on $3,000,000 Total Budget			
Model Component	**Typical Resource Allocation**	**Desired Resource Allocation**	**Difference**	**Percent Difference**
Data collection	$910,000	$1,190,000	+$280,000	31%
Estimation	$820,000	$490,000	-$330,000	-40%
Calibration	$580,000	$520,000	-$60,000	-10%
Validation	$450,000	$520,000	+$70,000	16%
Documentation	$240,000	$280,000	+$40,000	17%

Model documentation should explain expectations and limitations of the model to users of the models. As stated previously, models are complex and context specific. As such, they are not necessarily valid for all desired options and alternatives requested by an end user. Model documentation also may provide advice on when new models or new modeling techniques are necessary.

Recommendations for Improving Validation Practices

Specific recommendations regarding an improved *Model Validation and Reasonableness Checking Manual* were requested from Peer Exchange participants in addition to the direction provided by the general discussion of an improved model validation process. The following recommendations were provided in response to that charge:

- All elements of the current TMIP manual should be included in a revised manual
- A chapter on travel model validation priorities should be added to the manual

- "Trip-based" language should be removed from the manual
- The issues and impacts of error inherent in observed data should be discussed

Peer Exchange participants also were asked to specify primary and secondary model validation tests that would be appropriate to discuss in an updated *Model Validation and Reasonableness Check Manual*. Table 4 summarizes the recommended tests. The table does not provide details regarding the specific test procedures or minimum standards for the tests. Potential sources for observed data have been added to the table.

While the participants suggested that a revised manual avoid the use of trip-based language, there was some concern that this might be counter-productive since activity-based models currently are not within the reach of many MPOs. The recommended primary and secondary tests were specified using the trip-based context.

■ Concluding Remarks

Seven questions were presented to the Peer Exchange participants at the beginning of the general discussion. While specific responses to each of the individual questions were not explicitly obtained during the Peer Exchange, responses were developed in the initial draft of this White Paper based on review of the notes taken during the Peer Exchange. The initial draft of this White Paper was distributed to the Peer Exchange participants for review and comment. The responses to the seven questions as reviewed and edited by the Peer Exchange participants are listed below.

How important is it to match base-year observations?

While it is important to match base year observations for validation, simple matching of traffic counts, for instance, is not sufficient to establish the validity of a travel model. Quality model validation must test all steps of the travel model and also should test model sensitivity. In addition, model adjustments to correct problems need to reflect travel behavior rather than be based on simple arithmetic. This will help avoid repeating the problems in the future.

Any model validation and sensitivity testing must begin with good network design and network coding procedures in place. Without this solid basis for modeling, the best validation procedures and efforts are exercises in futility. The FTA's insistence on demonstrating the veracity of the coded transit networks and transit assignment procedures for New Starts analyses provides a good example of the importance of the proper representation of transportation networks on travel forecasts.

Table 4. Recommended Primary and Secondary Model Validation Tests

Model Component	Primary Tests	Secondary Tests	Potential Validation Data Sources
Networks/Zones	• Correct distances on links • Network topology including balance between roadway network detail and zone detail • Appropriateness of zone size given spatial distribution of population and employment • Network attributes (managed lanes, area types, speeds, capacities) • Network connectivity • Transit run times	• Intrazonal travel distances (model design issue) • Zone structure compatibility with transit analysis needs (model design issue) • Final quality control checks based on review by end users • Transit paths by mode on selected interchanges	• GIS center line files • Transit on-board or household survey data
Socioeconomic Data/Models	• Households by income or auto ownership • Jobs by employment sector by geographic location • Locations of special generators • Qualitative logic test on growth • Population by geographic area • Types and locations of group quarters • Frequency distribution of households and jobs (or household and job densities) by TAZ	• Dwelling units by geographic location or jurisdiction • Households and population by land use type and land use density categories • Historical zonal data trends and projections to identify "large" changes (e.g., in autos/ household from 1995 to 2005)	• Census SF-3 data • Quarterly Census of Employment and Wages (QCEW) • Private sources such as Dun & Bradstreet

Table 4. Recommended Primary and Secondary Model Validation Tests (continued)

Model Component	Primary Tests	Secondary Tests	Potential Validation Data Sources
Trip Generation	• Reasonableness check of trip rates versus other areas • Logic check of trip rate relationships	• Checks on proportions or rates of nonmotorized trips • Reasonableness check of tour rates • Cordon lines by homogeneous land use type	• NCHRP-365 (or subsequent manual) • Traffic counts (or intercept survey data) for cordon lines • Historic household survey data for region • NHTS (2001 or 2008)
Trip Distribution	• Trip length frequency distributions (time and distance) by market segments • Worker flows by district • District-to-district flows/desire lines • Intrazonal trips • External station volumes by vehicle class	• Area biases (psychological barrier – e.g., river) • Use of k-factors (Design Issue) • Comparison to roadside intercept OD surveys • Small market movements • Special groups/markets • Balancing methods	• CTPP data • ACS data • NCHRP-365 (or subsequent manual) • Traffic counts (or intercept survey data) for screenlines • Historic household survey data for region • NHTS (2001 or 2008)
Time-of-Day (TOD) of Travel	• TOD versus volume peaking • Speeds by TOD	• Cordon counts • Market segments by TOD	• Permanent traffic recorder data • NHTS (2001 or 2008) • Historic household survey data for region • Transit boarding count data

Table 4. Recommended Primary and Secondary Model Validation Tests (continued)

Model Component	Primary Tests	Secondary Tests	Potential Validation Data Sources
Mode Choice	• Mode shares (geographic level/ market segments) • Check magnitude of constants and reasonableness of parameters • District level flows • Sensitivity of parameters to LOS variables/elasticities	• Input variables • Mode split by screenlines • Frequency distributions of key variables • Reasonableness of structure • Market segments by transit service • Existence of "cliffs" • Disaggregate validation comparing modeled choice to observed choice for individual observations	• Traffic counts and transit (or intercept survey data) for screenlines • CTPP data • NCHRP-365 (or subsequent manual) • Transit on-board survey data • NHTS (2001 or 2008) • Household survey data (separate from data used for model estimation)
Transit Assignment	• Major station boardings • Bus line, transit corridor, screenline volumes • Park-and-Ride lot vehicle demand • Transfer rates	• Kiss-and-Ride demand • Transfer volumes at specific points • Load factors (peak points)	• Transit boarding counts • Transit on-board survey data • Special surveys (such as parking lot counts)
Traffic Assignment	• Assigned versus observed vehicles by screenline or cutline • Assigned versus observed vehicles speeds/times (or VHT) • Assigned versus observed vehicles (or VMT) by direction by time-of-day • Assigned versus observed vehicles (or VMT) by functional class • Assigned versus observed vehicles by vehicle class (e.g., passenger cars, single-unit trucks, combination trucks)	• Subhour volumes • Cordon lines volumes • Reasonable bounds on assignment parameters • Available assignment parameters versus required assignment parameters for policy analysis • Modeled versus observed route choice (based on data collected using GPS-equipped vehicles)	• Permanent traffic recorders • Traffic count files • HPMS data • Special speed surveys (possibly collected using GPS-equipped vehicles)

How close is "close enough?"

"Close enough" depends on the intended use of the model being validated. Models used for project design or comparing alternative projects might require tight standards for model validation. In other cases, such as the evaluation of alternative transportation policies, the correct sensitivity of the model might outweigh the need for a close match of observed data. The varying uses and requirements of travel models has led some MPOs to develop advanced modeling techniques such as activity-based or tour-based models in an effort to respond to a wider range of questions. Alternatively, the varying uses and requirements of forecasts could lead to the development of multiple models for a region or multiple application approaches for a single model.

The "close enough" point of view outlined above must be weighed against economic realities affecting many DOTs and MPOs. Most users of the models and forecasts would like models that can respond to all issues and transportation options. Most DOTs and MPOs develop a single model for an area and use it to provide base forecasts for all analyses. The desire to use a single model might become even more prevalent as increasing infrastructure needs coupled with decreasing revenues result in shrinking modeling budgets. This calls for better guidance regarding good modeling and validation practices. Claiming acceptability for a model that fails to achieve specified standards for metrics such as percent RMSE, screenline crossings, and VMT ratios might seem irrational to a decision-maker if other agencies not using acceptable modeling procedures publish better validation results.

What steps are needed to ensure accurate validation data?

The need for better data was highlighted in the TMIP *Shining a Light Inside the Black Box* webinar series. Perspectives regarding how the real-world operates are driven by data. Data also drive how models can and should be informed. It stands to reason that more comprehensive and higher-quality data about travel demand, travel patterns, commercial vehicle movements, freight movements, external trips, performance of the transportation system, and volumes on transportation facilities and services are required for models. Data also drive the expanded validation and sensitivity tests discussed in the Peer Exchange.

A conscious, planned effort is required to ensure the collection of the data necessary to validate travel models. The data collection should focus on different types of data as well as the quality of the data. The validation data collected will be subject to sampling error regardless of the care exercised for data collection. For example, traffic counts and transit boardings vary on a daily basis. The collection of sufficient data to estimate means and standard deviations for each location or line is generally unreasonable (although, in effect, that is the purpose of factors used to estimate AADT or AWDT on roadways). Some of the drawbacks of sampling error for specific types of validation data (e.g., traffic counts or transit boarding counts) can be addressed by collecting multiple types of validation data.

Data collection should be initiated when new model estimations or updates are planned. The allocation of resources for collection of data should approach the allocation of resources for model estimation, calibration, and validation (combined).

How should model sensitivity be tested?

One of the primary methods for testing model sensitivity is forecasting or backcasting using the calibrated travel model. How well the forecasted or backcasted traffic volumes or transit boarding counts reproduce observed data should be tested.[11] When resources permit and data are available, temporal testing of individual model components should also be performed.

Model validation is primarily concerned with model sensitivity. Model calibration as defined earlier in this paper focuses on reproducing travel in a base year. Model validation focuses on the model's ability to reasonably respond to changes, typically by running the model for a different year.

One important aspect of sensitivity testing is the range of alternatives over which the model is valid. The limited ability of models to handle very large changes in behavior, transportation supply, or land use is rarely discussed. Large changes tend to overwhelm algorithms, coefficients, and constants designed to be sensitive within limited ranges. Even sensitivity and validation tests have limitations. Nevertheless, attempts should be made to determine the range of alternatives over which a model produces reasonable results.

Other tests to assess ability to forecast future travel?

Travel models should be continually assessed for their ability to forecast future travel. New issues or policies can appear quickly (e.g., the recent rapid increase in fuel prices). Travel models need to be assessed to determine whether they can credibly respond to the new issues or policies. The assessment may require innovative testing techniques. In addition, while outside actual validation, a good practice prior to the initiation of any model application would be a meeting or review session to clearly define model capabilities and model sensitivities.

While not specifically a validation test, the use of peer review panels throughout the model development and validation processes should be strongly considered. As noted under current practices, the most effective peer reviews are those that are initiated at the very beginning of the model development process. In these types of reviews, the panel members can make suggestions regarding model development and validation plans, model validation data needs, and model estimation, calibration, and validation results.

---

[11] Comparing forecasted results to observed data assumes that the travel model was calibrated for a previous year, such as 2005, and applied to a more recent year, such as 2008.

Models should be applied to prove that they can produce reasonable predictions of changes in travel both 1) between today and a future no-build scenario and 2) between a future no-build scenario and a realistic alternative. This must be accomplished through application of the model in full production mode; it cannot be asserted based on the quality of the model design and estimation, or base year calibration and validation. Findings from such applications are very helpful and typically highlight problems not evident in base year model validation runs.

Should risk analysis of future forecasts be performed?

The need to perform risk analysis of future forecasts is directly related to the decisions being made using the forecasts. In situations where scarce resources are being allocated based, at least in part, on the travel forecasts, risk analysis is a very important. New Starts analysis procedures provide a good example. The procedures require analysis of the components of the changes in transit ridership between a base year and the future both without and with the planned alternative. The risk analysis procedure provides the New Starts applicant and the FTA the information necessary to assess the sources of ridership (such as population growth and service improvements) for future year alternatives.

What is the role of validation documentation in raising model credibility?

Model validation results should be well documented in order to provide users of the travel forecasts the information they need to establish their confidence in the models. The model documentation should cover the limitations of the models as well as the capabilities of the models. If the model limitations and portions of the model that have not been validated are documented, users of the forecasts can reasonably assess the level of confidence they place in the forecasts. Understanding that a model cannot be used to test a specific issue or policy can, ultimately, lead to increased trust in the travel model when used for analyses for which it has been validated.

Validation documentation should also discuss the variables included in the model and how those variables influence the results. For example, mode choice documentation might note that auto operating costs are included in the model and that those costs represent items such as fuel costs, fuel efficiency, other out of pocket costs. This discussion might take place even if an explicit validation of model sensitivity to the variable has not been performed.

Appendix 1

Peer Exchange Model Validation Resource Paper

Peer Exchange on Travel Model Validation
(May 9, 2008)

Current Travel Forecasting Practices: Model Validation Resource Paper (Draft)

Prepared for:

FHWA Travel Model Improvement Program

Prepared by:

Cambridge Systematics, Inc.

May 6, 2008

Status of Travel Model Validation

Introduction

Purpose of Peer Exchange on Travel Model Validation

The objective of the Peer Exchange on Travel Model Validation Practices is to improve the state-of-the-practice in travel modeling. The FHWA Travel Model Improvement Program (TMIP) initiated the exchange in response to the recent TRB Special Report 288, *Metropolitan Travel Forecasting, Current Practice and Future Direction* (*SR 288*).[12] That report clearly outlines model validation issues in the summary (*emphasis added*):

> **Validation Errors:** Validating the ability of a model to predict future behavior requires comparing its predictions with information other than that used in estimating the model. Perceived problems with model validation include *insufficient emphasis and effort* focused on the validation phase, the *unavailability of accurate and current data* for validation purposes, and the *lack of necessary documentation*. The survey of MPOs conducted for this study found that validation is hampered by a *dearth of independent data sources*.

While it was quite unrelated to *SR 288*, a recent TMIP e-mail exchange on "Correlation Coefficient versus Coefficient of Determination" also points out issues regarding the purpose of model validation and the role of statistics in model validation. In summary, the initial question posed in the e-mail exchange was whether R or R^2 of modeled versus observed traffic counts should be used for validation. The original exchange referenced the document, *Calibration and Adjustment of System Planning Models* (FHWA-ED-90-015), dated December 1990, which suggested that "'correlation coefficient' is defined as an acceptable target for model accuracy; that being typically 'greater than 0.88' (page 35)." The original e-mail generated 13 responses with a number of the responses pointing out that 1) an R or R^2 value of 0.88 was an arbitrary value and 2) achieving such a standard was neither "necessary or sufficient" in establishing the validity of a travel model.

The efforts of the Federal Transit Administration (FTA) to improve model validation for ridership forecasts associated with New Starts projects also brings a renewed focus on model validation practices. The FTA emphasizes the use of models that are tested rigorously against current transit ridership patterns to ensure that model forecasts serve as a useful basis for quantifying and understanding user benefits from proposed New Starts projects. The implications of careful calibration and validation methods are threefold: first, they require better current data; second, they call for a focus on transit markets; and third, they require better tests and standards.

---

[12] Special Report 288, *Metropolitan Travel Forecasting, Current Practice and Future Direction*, Transportation Research Board, 2007.

With the above in mind, it is hoped that the product of this peer exchange will ultimately be to provide practical advice and priorities to agencies performing model validation, specifically:

- Types of model validation to be considered including temporal considerations (forecasting/backcasting), common sense examination ("telling a coherent story"), and validation of each model component in addition to the model as a whole;
- Factors (e.g., data) to be considered for proper mathematical/statistics tests;
- Prioritization of model validation efforts considering scarcity of resources; and
- Proper documentation of model validation efforts.

Purpose of Resource Paper

This resource paper is intended to set the stage for peer exchange discussions by providing brief summaries and analyses of typical current practices, possible "improved" practices, and the gap between the current and improved practices.

■ Current Travel Model Validation Practices

This section is not intended to be an exhaustive summary of suggested travel model validation techniques that have been documented or of all existing travel model validation practices currently used by MPOs and other agencies charged with maintaining travel models. Rather, it is intended to give a brief summary of the current state-of-the-practice regarding travel model validation.

Summary of Selected Reports

TRB Special Report 288

SR 288 found that the current state-of-the-practice has a lack of effort concentrated on the validation process, as well as an unavailability of independent data. The report details interviews undertaken with 16 MPOs, specifically with a category on validation. *SR 288* suggests that there are two superior methods for model validation: using a prior year model to forecast current travel and using a current year model to backcast travel for a prior year. Backcasting is only used by 13 percent of large MPOs, and 5 percent of all MPOs. Many MPOs use the same data for model validation as were used to estimate the models, and there are no commonly agreed upon standards for error ranges other than FHWA suggested benchmarks and FTA guidance. "Formal validation thresholds, when they exist, tend to be limited to achieving the percentage [root mean squared] RMS error thresholds for assigned link volumes compared with counts established in FHWA planning guidance."[13]

SR 288 reports that most agencies use four datasets for model validation: household surveys, Census Transportation Planning Package (CTPP) data, traffic count data, and on-board transit survey data. Using household survey data as validation data is problematic since most models are developed using the same data. CTPP data can be used to help validate work trip distribution, but the data are available for only the journey from home-to-work trip purpose. Traffic counts data are typically used by agencies for model validation. However, traffic count data often have varying coverage quality and "validate" only the end result of the modeling process, not the individual steps. Transit on-board survey data (and transit boarding data) are used to validate models in regions with detailed mode choice models. In regions applying for FTA Section 5309 New Starts funding, some relatively detailed requirements are being established for transit model validation using the results of recently collected transit on-board survey data. Transit boarding count data are subject to the same concerns as traffic count data.

SR 288 noted a few examples of exceptional validation practices. These included the use of an air passenger survey for model validation, use of aerial photos to validate socioeconomic forecasts, implementation of rigorous logic checks on input population and employment data, and review of county-to-county trips.

TMIP Model Validation and Reasonableness Checking Manual

The *TMIP Model Validation and Reasonableness Checking Manual* suggests validation tests that are used by many MPOs. There are two main categories of validation tests suggested in the manual:

1. Reasonableness checks which typically compare model results to known values, such as traffic counts; and

2. Sensitivity tests to determine the model's response to transportation system, socioeconomic, or policy changes.

The manual indicates that many validation processes focus on the end results of the travel model, such as matching the assigned traffic volumes to traffic counts. The manual recommends supplementing this typical overall model validation process with validations of individual model components to ensure that each part of the model is behaving correctly. Appendix A provides a summary of the various model component tests suggested in the manual.

Ohio Department of Transportation Traffic Assignment Procedures Manual

In 2001, the Ohio Department of Transportation (ODOT) produced a Traffic Assignment Procedures manual to serve as a basic primer for highway network construction, highway

[13] SRR 288, page 92.

path-building, and traffic assignment. The manual provides the following direction regarding travel model validation (emphasis is from the original document):

> Model validation is the process of comparing model results to an **independent** data source. Model validation is a complete topic in unto itself which is described adequately in the U.S. DOT documents *Calibration and Adjustment of System Planning Models* and *Model Validation and Reasonableness Checking Manual*. A key thing to point out from these manuals is that *model validation involves checks of every step of the travel demand model process. This document only seeks to describe validation checks of the final traffic assignment volumes, not those needed for intermediate model steps*. ODOT has not previously formulated any specific guidelines for validation of intermediate model components; however, there are some reporting conventions that have been followed in the past that can be obtained by looking at any of the previous model validations conducted by ODOT. In addition, *NCHRP Report 365* contains many default parameters that can be compared to the results of each step of the model process to determine the reasonableness of the results.[14]

Thus, the manual states that model validation covers every step of the modeling process but that procedures and standards are provided only for validation of traffic assignments. This is due, mainly, to the fact that traffic counts are the most readily available, independent, data source. The manual sets the standard that a base year model validation assignment should be run every 10 years corresponding to the decennial Census. Appendix B provides additional information regarding ODOT's suggested model validation process.

NCHRP Report 365, Travel Estimation Techniques for Urban Planning

By its very nature, *NCHRP Report 365* is not a validation manual. The report states that validation is usually performed at different levels of the model system, in order to get a picture of how the model is performing as a whole, as well as looking more closely at the facility level. The report suggests the following validation tests based on the traffic assignment results:[15]

- Screenlines (checks trip distribution as well as assignment);
- Cordon lines (CBD, for example, checks both trip generation and trip distribution);
- Cutlines for major corridors (checks assignment functions and link attributes);

---

[14] Giaimo, Gregory, *Traffic Assignment Procedures*, Ohio Department of Transportation, Division of Planning, Office of Technical Services, August 2001, page 27.

[15] *Travel Estimation Techniques for Urban Planning*, NCHRP Report 365, Transportation Research Board, 1998, page 96.

- Link-specific volumes; and
- Regional statistics (such as RMS error that produces statistics on assigned versus observed traffic by facility type and volume groups).

As noted in the ODOT *Traffic Assignment Procedures* manual, information contained in *NCHRP Report 365* is often used for comparison in the validation of individual model steps. Past work performed by Cambridge Systematics staff and information from colleagues in other consulting firms and public agencies support the ODOT assertion. Model estimation or calibration results are frequently compared to information listed in *NCHRP Report 365* for validation purposes. These comparisons can be for any of the model components.

FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards

The technical memorandum *FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards: Literature Review* was prepared by Cambridge Systematics for the Florida Department of Transportation. The technical memorandum lists an assessment of typical model validation practices used throughout the United States. This assessment found that most models use the validation information presented in the reports *Calibration and Adjustment of System Planning Models*, *TMIP Model Validation and Reasonableness Checking Manual*, and *NCHRP Report 365* as validation standards (rather than benchmarks as they were intended). States which have established their own standards include Michigan, Oregon, and Tennessee. Commonly reported statistics found by this assessment include percent of trips by trip purpose, terminal times by trip purpose, average trip length by trip purpose, auto occupancy rates by trip purpose, volume-over-count ratios by facility type or functional class and screenline, and RMS error by facility type or functional class and volume group.

Information from Reviews of Published Documentation

Cambridge Systematics staff has recently reviewed published information regarding modeling practices from a number of MPOs. Information regarding model validation practices was included in the reviews. The documentation for 39 of the MPOs provided at least partial information regarding their model validation practices.

Table 1 summarizes the year for the most recent model validation for the MPOs. Only two of the 39 MPOs performed their most recent model validation prior to 2000; about 50 percent of the 39 MPOs performed some level of model validation between 2005 and 2008. The information summarized in Table 1 suggests that agencies charged with travel modeling stay relatively up-to-date with some type of validation of their models.

Table 1. Most Recent Year for Model Validation for Selected MPOs

Year	MPO Size (Population)			
	Small: Under 200,000	Medium: 200,000 to 1,000,000	Large: Over 1,000,000	Total
Prior to 2000	–	1	1	2
2000	2	1	5	8
2001	–	–	1	1
2002	1	–	2	3
2003	–	1	–	1
2004	–	1	3	4
2005	1	5	2	8
2006	–	2	3	5
2007	1	1	4	6
2008	–	–	1	1
Total	**5**	**12**	**22**	**39**

Table 2 summarizes the model components that were discussed in the various model validation documents reviewed. Validation of the traffic assignment process was, by far, the most frequently validated model component. However, several caveats must be considered in conjunction with this observation:

- Information regarding only the most recent model validation may have been provided. Agencies might perform regular or semi-regular overall model validations based on regularly collected traffic count data to ensure that the model is performing reasonably. Thus, model components not listed in the most recent model validation, such as trip generation, may have been validated when they were originally estimated or calibrated.

- Common validation tests for model components such as trip generation or trip distribution are typically based on data collected much less frequently (and typically at higher cost) than traffic count data. "Validations" against original model estimation data may not have been reported by some agencies.

Documentation for only two of the agencies included explicit detail regarding validations of their travel models for multiple years. A third agency (PSRC) indicated in a telephone interview that temporal validation of models estimated using year 2000 household survey data currently was being performed using the data from a 2006 household survey. However, as noted above, some agencies (e.g., SEMCOG) validate their models through a comparison of traffic assignment results to traffic counts on a semi-regular basis. A semi-regular model validation process also might be considered temporal validation of travel models.

Table 2. Model Components Validated in Most Recent Model Validation for Selected MPOs

Agency	Model Component Validated					
	Temporal	Traffic Assignment	Transit Assignment	Trip Generation	Trip Distribution	Mode Choice
BMC (Baltimore)		●	●	●	●	●
MUMPO (Charlotte, North Carolina)		●	●			
DRCOG (Denver)	●	●	●			
GBNRTC (Buffalo)		●	●	●	●	●
MAG (Phoenix)		●	●	●	●	●
Met Council (Minneapolis-St. Paul)		●	●	●	●	●
MWCOG (Washington)	●	●			●	●
MARC (Kansas City)		●		●	●	●
MTC (San Francisco)		●	●	●	●	●
NCTCOG (Dallas-Ft. Worth)		●	●			
NYMTC (New York)				●	●	●
PSRC (Seattle)	●	●	●	●	●	●
RTC (Las Vegas)		●	●			
SACOG (Sacramento)		●	●	●	●	●
SANDAG (San Diego)		●	●			
SCAG (Los Angeles)		●	●			
Memphis Urban Area MPO (Tennessee)		●		●	●	●
SEWRPC (Southeastern Wisconsin)		●	●	●	●	●
SEMCOG (Southeast Michigan)		●	●	●	●	●
WFRC (Salt Lake City)		●	●		●	●
Des Moines Area MPO (Iowa)		●				
GUAMPO (Greensboro, North Carolina)		●				
Hillsborough County MPO (Florida)		●				
Kern COG (California)		●	●			
Lee County MPO (Florida)		●		●	●	
MRCOG (Albuquerque)		●			●	
NFRMPO (Fort Collins, Colorado)		●				
TCAG (Tulare County, California)		●				
Bel-O-Mar Regional Council (Wheeling, WV)		●			●	
CCRPC (Champaign County, Illinois)						
CPCTS (Columbus, Georgia)		●				
Grand Junction/Mesa County MPO (Colorado)		●				
SIMPCO (Sioux City, Iowa)		●				
COMPASS (Ada and Canyon Counties, Idaho)		●				
Lincoln MPO (Nebraska)		●				
OKI (Cincinnati, Ohio)		●	●	●	●	●
Total Number of MPOs Performing Test	**2**	**32**	**18**	**14**	**18**	**5**

Appendix C provides several examples of the documentation of model validation results. Several observations can be made regarding the documentation of the travel models:

- There is no common format for model validation. Some agencies publish extensive detail regarding model results both with and without comparisons to observed data. For example, the MTC Baycast model validation documentation includes 158 pages of tables summarizing the model results for the various model components. At the other extreme, it is sometimes impossible to find published documentation of model validation.

- Model validation ranges from a simple comparison of results to other independent data allowing the reader to draw their own conclusion regarding the validity of the models to a "pass/fail" comparison.

- Results summarized in the *TMIP Model Validation and Reasonableness Checking Manual*, NCHRP Report 365, the 2000 Census Transportation Planning Package, and the 2001 National Household Transportation Survey are frequently used to provide independent checks on trip generation and trip distribution results.

- The 1990 publication, *Calibration and Adjustment of System Planning Models*, is frequently cited when specific validation standards are established for a region.

- When validation standards are established for a region, especially for traffic assignment results, it is often difficult to determine the types of model adjustments used in the model to enable the model to match the standards.

Interviews with Selected Agencies

In order to provide a basis for discussion for the Peer Exchange on Model Validation, detailed information regarding model validation practices was requested from 10 agencies. Agencies not directly represented at the Peer Exchange were, for the most part, selected for the interviews. The interviews are summarized in Appendix D.

Summary of Current Travel Model Validation Practices

Travel model validation practices for over 40 agencies were reviewed for the preparation of this resource paper. While there are some notable exceptions, most agencies tend to focus on matching traffic assignment validation as the primary validation practice. This practice seems to be driven by several factors:

- **Availability of Data –** Traffic count data are frequently collected by local agencies. The count data are frequently collected in support of other programs such as state-sponsored traffic counting programs and HPMS VMT estimates.

- **Resource Availability –** Traffic count data are, in comparison to other independent data sources such as travel surveys, relatively inexpensive to collect.

- **Established Guidelines and Standards –** In many cases, published documents such as the 1990 FHWA report, *Calibration and Adjustment of System Planning Models*, have been used by agencies to establish their own standards for determining whether a regional travel model is valid. In many cases, the information and suggestions contained in *Calibration and Adjustment of System Planning Models*, or the *TMIP Model Validation and Reasonableness Checking Manual*, have been misinterpreted to be Federal standards.

Validation of individual travel model components is infrequent and typically characterized by comparisons to summaries of expanded data used for the calibration of the model components. As such, the typical practices for individual model components might be called "valibrations" since they do not actually validate against independently collected data. The "valibration" of individual model components is infrequent since the data necessary are collected infrequently. Unfortunately, the practice also is infrequent in terms of the number of agencies actually performing the tests (see Table 2).

Table 3 summarizes the typical standard and enhanced validation practices in use today based on the agencies reviewed for this resource paper. As mentioned above, the overall model validation process is frequently defined by comparisons to observed traffic count data and, in regions with transit components included in the regional model, to observed boarding count data.

Table 3. Typical Model Component Validation Techniques Currently in Use

Component	Standard Validation Practices	Enhanced Validation Practices
Land Use Model	• None	• Comparison of land use "forecast" based on past base year to observed land use for current year
Socioeconomic Data Inputs	• Comparison of regional demographic distributions to most recent census data	
Network Data Inputs	• Review of network plots for continuity of functional classes, number of lanes, speed limits, etc. • Review of connector links • Comparison of total transit route travel times to published time tables	• Comparison of congested or uncongested travel times for paths to observed speed runs • Assignment of observed transit trip tables to networks and comparison of boardings by route

Table 3. Typical Model Component Validation Techniques Currently in Use (continued)

Component	Standard Validation Practices	Enhanced Validation Practices
Trip Generation	• Comparison of average trip rates by purpose and overall to regional estimates from household survey data (typically used to estimate the model). • Comparison of average trip rates and proportions of trips by trip purpose to other regions (e.g., based on NCHRP Report 365).	• Application of trip generation models estimated using household survey data from previous year to results/ rates from current household survey.
Trip Distribution	• Comparison of average trip lengths by purpose and overall to regional estimates from household survey data (typically used to estimate the model). • Comparison of average trip lengths by trip purpose to other regions (e.g., based on NCHRP Report 365). • Comparison of district-to-district distributions of home-based work trips to 2000 CTPP data. • Screenline summaries.	• Application of trip distribution models estimated using household survey data from previous year to results/ rates from current household survey.
Mode Choice	• Comparison of trips and shares by mode to observed data from household survey data (typically used to estimate the model). • Comparison of trips and shares by mode to observed data from transit on-board survey (may or may not have been used for model estimation).	• Application of FTA guidelines for validating transit path-building and mode choice models for Section 5309 New Starts modeling.
Time-of-Day of Travel	• Comparison of trip or shares of trips by time-of-day by purpose and overall to regional estimates from household survey data (typically used to estimate the model).	• Comparison of model trips and VMT by time-of-day to peaking patterns based on traffic count data.
Trip Assignment	• Comparison of VMT by varying stratifications of the highway network to estimates of observed from traffic counts. • Percent RMS error, R2, percent deviation, etc., of modeled to observed count data by stratifications of the highway network. • Comparison of modeled to observed transit boardings by line, type of service, etc.	• Time-of-day estimates of VMT by varying strata. • Comparisons of assigned speeds as well as modeled traffic including speeds/travel times over identified routes.

Very few areas formally perform temporal validation of their travel models by using a model calibrated for a current base year to backcast to a previous year (for which observed household survey data or traffic count data are available). While formal temporal validation is infrequent, temporal validation has been performed in numerous regions using de facto methods. The de facto temporal validations have occurred when regions revalidate their models either on a semi-regular basis when new traffic count data are collected or when updated travel models are validated for different years than the validation year used for the parent travel model.

Formal sensitivity testing of travel models as a validation test is rare. This assertion ignores cases where sensitivities of model components and model parameters are tested during model estimation. The PSRC has, perhaps, performed one of the most rigorous sensitivity tests of a regional travel model in their assessment of the models performance for different tolling initiatives. PSRC also reports that they "validate" their future year forecasts for reasonability.

■ "Improved" Validation Practices

As stated in the introduction to this resource paper, the objective of the Peer Exchange on Travel Model Validation Practices is to improve the state-of-the-practice in travel modeling. Thus, the "improved" validation practices cannot be specified without the comments and ideas presented at the peer exchange. Nevertheless, it is worthwhile to identify some validation practices that go beyond the standard practice of comparing modeled traffic to traffic counts and modeled transit boardings to transit boarding counts.

Perhaps the first step in specifying "improved" validation practices should be to define the purpose of model validation. The *TMIP Model Validation and Reasonableness Checking Manual* defines model validation as:

> In order to test the ability of the model to predict future behavior, validation requires comparing the model predictions with information other than that used in estimating the model. This step is typically an iterative process linked to model calibration. It involves checking the model results against observed data and adjusting parameters until model results fall within an acceptable range of error. If the only way that a model will replicate observed data is through the use of unusual parameters and procedures or localized "quick-fixes," then it is unlikely that the model can reliably forecast future conditions.

The recent TMIP Webinar, *Shining a Light Inside the Black Box; Part 2 – Model Testing*, provided the following definition of traditional model validation:

> "forecasting" current travel patterns to demonstrate sufficient ability to reproduce highway counts and transit line volumes.

The above definitions of model validation suggest that a model that reproduces observed travel patterns within "an acceptable range of error" without the "use of unusual parameters and procedures or localized quick-fixes" for one point in time can be used for forecasting future travel. They also ignore the common-sense aspect of model validation regarding whether estimated or calibrated model parameters are logical and can be explained.

Finally, narrowly focusing solely on these traditional definitions of validation ignores model sensitivity and the establishment of ranges of alternatives within which the model can be expected to produce reasonable results. Indeed, just after providing the traditional definition of model validation, the *TMIP Model Validation and Reasonableness Checking Manual* defines model application as:

> Although the model may replicate base year conditions, the application of the model to future year conditions and policy options requires checking the reasonableness of projections, so there is a link between application and validation as well. The sensitivity of the models in response to system or policy changes is often the main issue in model application.

An alternative definition of model validation can be used to address the static nature of the traditional definition of model validation:

> The determination of the ranges of error within which the travel model components reproduce existing travel patterns and travel behavior coupled with the estimation of the sensitivity of the model to changes in the transportation system, demographics, and transportation policies.

The FTA has, effectively, used a corollary to the above definition for model validation in their guidelines for New Starts through their suggestion that "travel models should tell a coherent story." While this suggestion includes an element of need to reasonably reproduce current travel patterns, the focus is really on reasonably explaining how or why travelers make their specific travel decisions.

Documentation Describing "Improved" Model Validation Practices

TMIP Model Validation and Reasonableness Checking Manual

While the *TMIP Model Validation and Reasonableness Checking Manual* was the source of one of the traditional definitions of model validation, the model validation concepts described in the manual provide a number of improved model validation practices, including:

- Validation of individual model components to ensure that each component reasonably reproduces observed travel characteristics.
- Validation of the overall set of models to test the effects of compounding errors.
- Performance of reasonableness checks of model rates and parameters against observed values, parameters estimated in other regions, or secondary data sources for consistency. The models should be evaluated in terms of acceptable levels of error,

their ability to perform according to theoretical and logical expectations, and the consistency of model results with the assumptions used to generate them.

- Performance of sensitivity tests of model responses to transportation system, socioeconomic, or policy changes.

TRB Special Report 288

Special Report 288 emphasized two primary points regarding the improvement of travel model validation.

- Validation requires comparing the model output with information other than that used in estimating or calibrating the model. The report points out that truly validating many of the model components against independent data would require a data collection effort as large or larger than the data collection employed for model estimation. Due to the cost of such an effort, the report suggests forecasting or backcasting as a way to check the validity of models. The report points out, however, that the use of forecasting or backcasting to validate a model might be useful only for identifying surface problems with models.

- Sensitivity testing is key to checking the reasonableness of travel forecasts. The committee recommended tests in which model inputs and assumptions were varied to determine whether the changes in modeled results were realistic. The use of the FTA's Summit program was suggested as a tool that could be used in conjunction with the model sensitivity checking.

TMIP Shining the Light Webinar

As noted above, the recent TMIP webinar, *Shining a Light Inside the Black Box; Part 2 – Model Testing*, provided the following definition of traditional model validation:

> "forecasting" current travel patterns to demonstrate sufficient ability to reproduce highway counts and transit line volumes.

The webinar identified a number of problems with the traditional model validation approach including lack of resources reserved for model validation, insufficient data for verification of model estimates, validation efforts overly focused on reproduction of existing traffic volumes or transit line volumes, lack of attention to the forecasting impacts of model adjustments to match traffic counts, and insufficient documentation of validation results.

Two primary improvements to typical validation practices were suggested in the webinar:

1. Collect data to sufficiently test model estimates and results such as:
   – Detailed person demand/travel flow data;
   – Detailed freight demand/travel flow data;
   – Actual highway and transit speeds;

- Actual point-to-point travel times; and
- Volumes on facilities/services.

2. Perform more meaningful model tests including:
   - Expansion of model calibration/validation efforts;
   - Interpretation of models vis-à-vis traveler behavior;
   - Demonstration of reasonable predictions of changes in travel; and
   - Informative documentation of testing results and forecasting weaknesses.

FTA New Starts Procedures

FTA guidelines for travel forecasts produced for New Starts have been evolving over the past few years. While the FTA makes it clear that they approve travel forecasts, not travel models, their approval of travel forecasts require generally acceptable, documented travel modeling practices. In an FTA-sponsored New Starts workshop held in St. Louis in September 2007, FTA staff defined validation as a plausible description of current travel behavior, a plausible basis for a discussion of current conditions, and plausible forecasts of changes in transit ridership. The first two parts of the definition, the plausible description of current travel behavior and the plausible basis for a discussion of current conditions, are demonstrated by the following guidelines:

- Model parameters, especially for mode choice and transit path-building, should be within specified ranges based on reviews of travel models from throughout the country. If the mode choice model parameters are outside of the generally acceptable range, a compelling argument supporting the parameters should be provided.

- Model parameter relationships used for transit path-building should be consistent with the relationships of comparable parameters used for mode choice.

- Transit speeds for buses operating in mixed flow should be based on congested auto speeds.

- The veracity of the transit network should be demonstrated through the assignment of observed transit trip tables built from recently collected (i.e., within the past four years) transit on-board survey data. The boardings resulting from the transit trip assignment should reasonably reflect observed boardings by line and estimated transfer rates.

- Matching overall trip (boardings) targets is not sufficient. Models used for New Starts also should account for transit markets defined by trip purpose, socioeconomic class, production attraction locations, and transit access modes.

The last part of the definition, plausible forecasts of changes in transit ridership, has roots in sensitivity testing and uncertainty analysis. Starting in 2008, New Starts applications will require the following standard ridership forecasts, analyses, and summary reports:

- Future No-Build alternative versus "today;"
- Future TSM alternative versus No-Build alternative;
- Future Build alternative versus TSM alternative;
- Opening year Build alternative versus today; and
- Detailed analysis of transit user benefits accruing from changes in in-vehicle travel times resulting from a proposed project.

The above analyses are intended to provide detailed information regarding the sensitivity of the travel models and the sources for forecast changes in transit ridership.

Improved Validation Methods – Examples from New Models Currently Being Developed

In a white paper prepared for the 2006 Innovations in Travel Modeling Conference held in Austin, Texas, Pendyala and Bhat provided the following comments regarding travel model validation:

> There is no doubt that any model, whether an existing four-step travel demand model or a newer tour- or activity-based model, can be adjusted, refined, tweaked, and – if all else fails – hammered to replicate base year conditions. Thus, simply performing comparisons of base year outputs from four-step travel models and activity-based travel models alone (relative to base year travel patterns) is not adequate…the emphasis needs to be on capturing travel behavior patterns adequately from base year data, so that these behavioral patterns may be reasonably transferable in space and time.[16]

Pendyala and Bhat suggest that emerging tour- and activity-based models be assessed, or validated, by their ability to respond to a range of scenarios and policies of interest and that tests of the models responses. The following sections provide examples of validation tests that are being applied in three regions developing new activity/tour-based travel models.

Denver Regional Council of Governments (DRCOG)

DRCOG has prepared an ambitious model validation plan that focuses on three types of tests:[17]

---

[16] Pendyala, R and C. Bhat, *Validation And Assessment Of Activity-Based Travel Demand Modeling Systems*, prepared for Innovations in Travel Modeling Conference, Austin, Texas, 2006.

[17] Parsons Corporation, *DRCOG Model Validation Plan*, Draft 2a, September 2007.

1. *Validation tests* focus on the reasonability of model parameters and the ability of the models to reproduce observed traveler behavior, traffic volumes, and transit ridership;

2. *Sensitivity tests* focus on the reasonableness of the model response to changes in socioeconomic conditions, transportation networks and alternatives, changes in congestion, changes in pricing, etc.; and

3. *System integrity tests* focus on the ability of the models and model implementation software to perform "as advertised," as well as determining limits such as space and time requirements for model application.

The validation plan describes possible validation and sensitivity tests for each model component included in the modeling process. Table 4 is an example of the planned validation tests for the daily activity pattern model, which forecasts the numbers of tours and numbers of intermediate stops made on a given day by each resident of the region for each of seven activity purposes.

Table 4. Example Daily Activity Pattern Model Validation Tests Planned by DRCOG

Aggregation Level	Validation Measures	Expected Outcomes	Priority
Comparison of Model Parameters to Other Regions	Comparison of model coefficients to: • Sacramento; • San Francisco; and • Columbus.	• No expectations; comparison only.	Level 1
Disaggregate	Prediction success of modeled daily activity pattern choices against observed TBI estimation data.	• Prediction success likely to be very low.	Level 3
Aggregate	Numbers or percents of residents making tours and intermediate stops by activity type: • For the region; • By county; • By household size and income group; • By household size and auto ownership; • By gender and age group; • By employment status; and • By student status.	• Compare modeled to expanded observed numbers or percents. • Review for reasonable patterns.	Level 2

Table 4. Example Daily Activity Pattern Model Validation Tests Planned by DRCOG (continued)

Aggregation Level	Validation Measures	Expected Outcomes	Priority
	Percent of "immobiles" (persons with no out of home activities during the day) by: • By household size and income group; • By household size and auto ownership; • By gender and age group; • By employment status; and • By student status.	• Compare to results summarized by Kay Axhausen (e.g., in Transportation, Volume 34, Number 1, January 2007, pages 107-128).	

Planned sensitivity tests for the DRCOG models include:

- Temporal sensitivity test comparing 1997 to 2005 model results to each other, to observed traffic volumes and transit boardings, and to results from the four-step model;
- Short-term changes sensitivity test that compares 2007 model results to 2007 traffic volumes and transit boardings; several major transportation system changes were made between 2005 and 2007 (a major highway expansion in the Southeast I-25 corridor, opening of a Southeast light rail line, and conversion of North I-25 HOV lanes to HOV/Toll lanes);
- Long-term changes sensitivity test that compares the 2035 forecast results from the activity-based model to the 2035 forecasts from the four-step model;
- FTA New Starts sensitivity test that compares 2035 forecasts for the TSM and Build alternatives using the activity-based model to those obtained using the four-step model;
- "One parameter" sensitivity test to measure the impacts on travel forecasts of the change in one item such as an increase of transit fares by 25 percent; and
- Land use changes sensitivity test to measure the impact of a major site redevelopment.

Atlanta Regional Council (ARC)

The population synthesizer, PopSyn, developed for the ARC model is being adopted for use in several other areas (e.g., Denver, Sacramento, and the San Francisco Bay Area).

John Bowman and Guy Rousseau reported on the extensive validation tests performed on the synthesizer at the 2006 Austin conference.[18] Several innovative concepts were used for the validation of PopSyn:

- A temporal validation was performed by using year 2000 as base year and backcasting to 1990. The synthesized 1990 population was compared against 1990 census data to test the ability of PopSyn to generate a synthetic population with limited forecast (actually, backcast) information.

- Validation was performed by calculating aggregate characteristics of the synthetic population and calculating the same characteristics directly from detailed census tables for four levels of geographic aggregation: tract, PUMA, County, supercounty. Validation "accuracy" for a population characteristic was obtained when the mean percentage differences between the synthesized and observed data for the characteristic was not statistically significantly different from zero.

- Validation "precision" also was measured. Precision refers to statistical variance; a variable with a large variance in the difference between the synthetic population and the census validation value is considered imprecise.

Sacramento Area Council of Governments (SACOG)

SACOG is in the process of developing and testing their activity-based model, SACSIM. Current validation tests and comparisons that are underway or planned include:

- Synthetic population distributions (household size by number of workers by income level by age) to available census distributions for various geographies;

- Auto ownership model results for percents of 0-auto households, percents of households with number of cars less than number of workers, and average vehicles per household to available census data by various geographies;

- Number of tours per day (0, 1, or 2+ tours per day) by tour purpose and person type to expanded household survey data;

- Intermediate number of stops per tour (1, 2, or 3+ stops per tour) by tour purpose and person type to expanded household survey data;

- Percent of households making no travel for the day to expanded household survey data and/or aggregate data from alternative sources;

- Time of arrival and departure distributions by purpose and person type to expanded household survey data;

---

[18] Bowman, John and Guy Rousseau, *Validation of the Atlanta (ARC) Population Synthesizer (PopSyn)*, prepared for the TRB Conference on Innovations in Travel Modeling, Austin, Texas, May 21-23, 2006.

- Duration of activity distributions by purpose and person type to expanded household survey data;
- Worker flows from place of residence to usual place of work, to census data for various geographies;
- Trip length frequency distributions by purpose to expanded household survey data;
- Work tour mode of travel to census data for various geographies;
- Trip mode of travel for non-work tours to expanded household survey data;
- Walk skim distances for model estimated parcel-to-parcel walk distance to GIS-based skim distances with a focus on short trips;
- Traffic assignment results to traffic counts by link capacity class, volume range, etc.;
- Transit assignment results to transit boarding counts for weekday line volumes and LRT station boardings;
- Daily VMT by county to HPMS VMT estimates;
- VMT due to travel component (household generated, commercial vehicle generated, and external travel generated).

In addition to the above validation tests, the following sensitivity tests are planned:

- Impact of auto operating cost changes on VMT, transit trips, and bike/walk trips;
- Impact of transit fare changes on VMT, transit trips, and bike/walk trips;
- Impact of age distribution changes on VMT, transit trips, and bike/walk trips.

Validation "Accuracy" versus Model Sensitivity

The following sections are intended to raise discussion points that should be addressed during the peer exchange discussions.

How important is it to match base year observations?

Some agencies have identified that maintaining model sensitivity is as important as matching base year data. For example, PSRC has decided that they could live with "less accurate" models or model components with respect to matching any specified validation guidelines in order to ensure that the models are reasonably sensitive to changes in the transportation system. The initial validation test specified by the Ohio DOT explicitly implies that about one-third of the modeled traffic on links with counts will be greater than a specified level of error. For New Starts modeling, the FTA has recommended against using district interchange-specific mode choice model constants as a means to match district-to-district mode shares.

How close is "close enough?"

Modelers and decision-makers have often relied on arbitrary criteria to measure whether or not a model is valid. For example, a correlation coefficient of 0.88 was used in *Calibration and Adjustment of System Planning Models* to illustrate a suggested model validation procedure. That value has been interpreted by some as a litmus test of model validity specified by the Federal government.

Many modelers recognize that matching arbitrary standards is neither necessary nor sufficient for determining the validity of a model. At the same time, there is recognition that a model that produces results bearing no resemblance to reality will not be trusted by decision-makers. Thus, there is some call for the establishment of some criteria or guidelines to use to measure the validity of travel models.

There could be some concern that the establishment of any criteria or standards could lead to the same issue of an arbitrary criterion being used as a litmus test of model validity. To avoid this situation, would it be reasonable to supplement or replace numerical or statistically based measures with nonnumerically based guidelines? Would a comprehensive validation template be a possible approach to improving model validation?

What steps should be taken to ensure accurate validation data?

The initial Ohio DOT validation test explicitly considers the accuracy of traffic count data. The validation of the ARC population synthesizer also explicitly considered the variation in observed data. However, other examples of the consideration of the accuracy of validation data are rare.

How should model sensitivity be tested?

Some model validation efforts are starting to consider model sensitivity. Some of this testing can be accomplished by temporal validation of models especially when a there has been a major transportation system or demographic change between the validation years. However, other aspects of model sensitivity must be validated by the "that seems reasonable" approach. Are there methods to establish reasonable levels of model sensitivity?

What other tests should be considered to enhance the model's ability to forecast future travel patterns?

Are there tests other than sensitivity testing that can be used to establish model credibility for future forecasts? Is it possible to perform a future year "validation" based on expert opinions regarding likely changes in travel patterns for different network, socioeconomic, or policy changes?

Should risk analysis of future forecasts be performed?

The FTA has established a procedure to analyze the components of change for New Starts. Should, or can, a comparable procedure be established for testing the reasonability of future year forecasts?

What role should the validation documentation play in raising the credibility of travel models and their results?

There is tremendous variation in travel model documentation. Documentation ranges from being nonexistent to multivolume treatises. What elements of documentation are necessary to help establish the credibility of a travel model and its likely value in forecasting future travel?

Outline of Peer Exchange White Paper

A white paper documenting the peer exchange will be prepared. The following is the tentative outline for the white paper:

1. Introduction

1.1 Purpose of the Peer Exchange

1.2 Participant List

2. Current Validation Practices

2.1 Review of Current Practices

2.1.1 Resource paper summary

2.1.2 Summarize information provided during peer exchange

2.2 Assessment of Current Practices

2.2.1 "How close is close enough?" (The role of standards and guidelines.)

2.2.2 "Is model sensitivity properly tested?"

2.2.3 "How good are the validation data?"

2.2.4 "Are current model validation practices properly fulfilling a need?"

3. Improved Validation Practices

3.1 Elements of Improved Practices

3.1.1 What are the goals/objectives for validation?

3.1.2 What tests should be used?

3.1.3 How to collect and assess the validation data?

3.1.4 What are validation priorities when resource constraints are an issue?

3.2 Analysis of Current versus Improved Validation Practices

3.3 Recommendations for Improving Validation Practices

3.3.1 Documentation, research, training needs?

3.3.2 What are the priorities?

3.3.3 Federal, TMIP, TRB, state, MPO, university, consultant roles

Appendix A – Model Validation Tests Suggested in *TMIP Model Validation and Reasonableness Checking Manual*

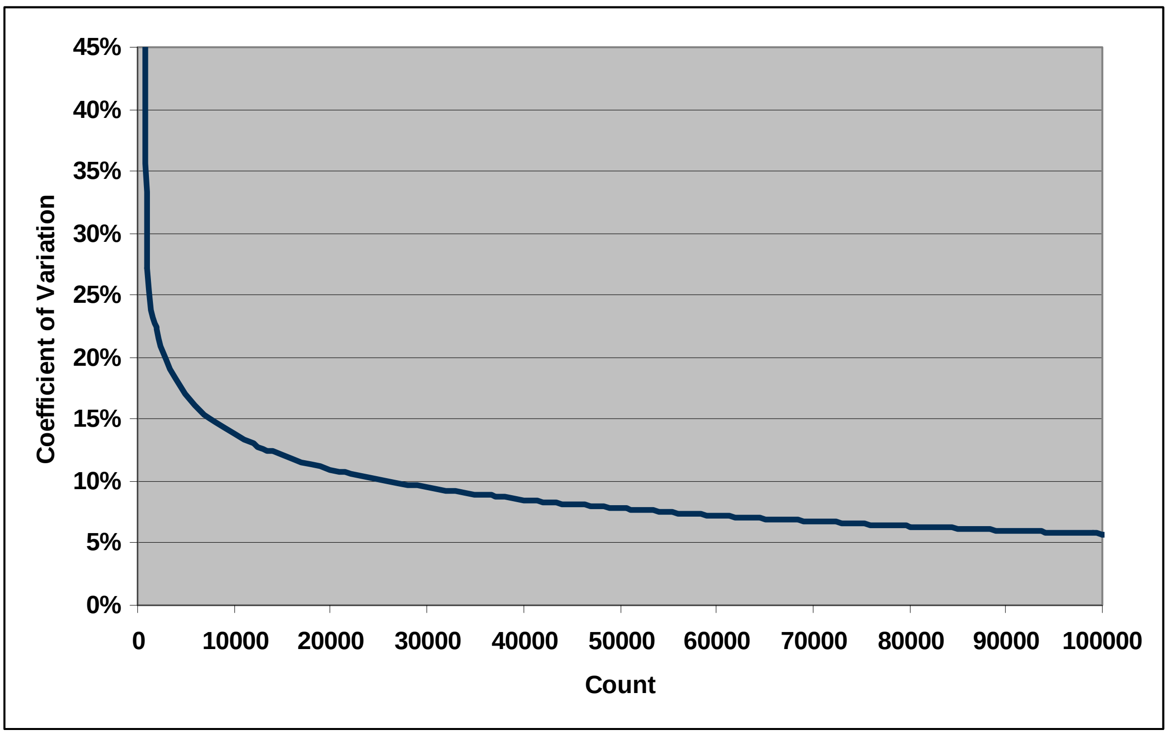
Model Component	Suggested Validation Tests
Model Input Data	• Aggregate totals against control totals by city/county/region. • Regional demographic characteristics such as average household size.
Socioeconomic Models	• Compare observed and estimated households by socioeconomic subgroup. • Calculate correlation of the shares of observed and estimated households and/or plotting the correlations at the district or census tract level to determine geographic biases.
Trip Production	• Calculate the total person trip productions per household or per capita. • Calculate total person trips by purpose. • Compare observed and estimated trips. • Calculate correlation of the shares of observed and estimated trips by purpose and/or plotting the correlations at the district or census tract level to determine geographic biases.
Trip Attraction	• Review home-based work person trip attractions per total employment, home-based school trips per school enrollment, and home-based shop trips per retail employment.
Trip Distribution	• Compare estimated and observed average trip lengths by purpose. • Compare estimated and observed trip lengths for trips produced and trips attracted by purpose and area type. • Plot estimated and observed trip length frequency distributions by purpose.
Mode Choice	• Compare modeled and observed mode selection probabilities by measures such as household, traveler, zone, and trip characteristics such as trip length (disaggregate validation). • Compare modeled and observed mode shares, transit ridership, traffic volumes, and auto occupancy rates (aggregate validation). • Compare modeled and observed home-based work trips on transit as a percentage of total transit trips. • Compare modeled and observed average auto occupancies to and from area types or major districts.
Time-of-day Model or Directional Split Factors	• Compare model time-of-day factors to those derived from NPTS and CTPP data by purpose, by mode, and by direction.

Model Component	Suggested Validation Tests
Highway Assignment	• Compare modeled to observed VMT and VHT for the region, per household, and per person. • Compare modeled to observed traffic volumes (by screenline and on links with counts). • Correlation (or R2), percent RMS error of modeled to observed traffic counts. • Assignment tests such as review of highway paths based on congested travel times, select link analyses, and assigning only through-trips to check paths.
Transit Assignment	• Check modeled versus observed boardings for the region, by mode/ submode, and by trip length. • Check transfer rates for reasonability.

Appendix B – Model Validation Tests Suggested in Ohio Department of Transportation Traffic Assignment Procedures Manual

The chapter of the manual describing validation checks and procedures is prefaced by a section on “Assignment Limitations.” The section also includes a description of sample error inherent in traffic counts with the caveat: “The amount of sampling error present in counts sets a lower bound on percent error of model results, below which it is unrealistic to expect the model results to obtain.” Figure B.1 shows the “Expected Coefficient of Variation in Daily Count Volume” from the *Guide to Urban Traffic Counting* referenced by ODOT.[19]

Figure B.1 Expected Coefficient of Variation in Daily Count Volume



Source: *Traffic Assignment Procedures*, ODOT, page 27.

The information on traffic count error is used for an interesting initial check of the traffic assignment process. Specifically, assigned traffic volumes are compared to daily traffic counts for each link with a traffic count and the percent error is calculated. The proportion of links whose percent error exceeds the expected coefficient of variation

[19] U.S. DOT, *Guide to Urban Traffic Counting*, 1981.

(shown in Figure B.1) must be reported. Traffic assignments with **less than** 33 percent of the links exceeding the expected coefficient of variation curve are **regarded with caution by ODOT** due to the likelihood of artificial model or assignment adjustments to force agreement with counts.

Once the initial check has been completed, the ODOT manual specifies several validation tests:

- A review of network plots of modeled traffic volumes and traffic counts is recommended as the best check of a traffic assignment. No guidelines or standards are set for this test.

- Modeled to observed traffic volume RMS error is calculated for up to 18 volume groups. The model is deemed to have passed the RMS error test when the percent RMS errors for all volume groups are less than the standards specified in the manual. The manual suggests approaches to improve model results if the validation results are not acceptable.

- Modeled to observed vehicle-miles of travel (VMT) for the region, by functional and administrative class of the roadway, and by geographic ring and sector are determined. The manual states that the modeled regional VMT must be within 3 percent of the VMT based on traffic counts. Guidelines for maximum difference in VMT are set for functional and administrative classes of the roadway and for geographic ring and sector; the manual states that the modeled VMT should fall within the specified error ranges.

- Modeled to observed traffic volumes crossing screenlines are determined. Guidelines for the maximum desirable deviation in screenline volumes are set in the manual.

Appendix C – Examples of Validation Documentation

■ Lincoln (Nebraska) MPO

The following excerpt is from the draft model documentation by Lima and Associates for the Lincoln MPO. While this excerpt was not explicitly in the validation section of the documentation, it provides a reasonability test against independent data that can be considered a validation test.

Total trips generated in this step are person trips and not vehicle/auto trips. Mode choice step of the Lincoln MPO model separates the auto trips from the non-auto trips. It is for this reason that the trip generation process estimates the total person trips. Table 16 lists the total number of trips generated by each trip purpose for the base year 2004. A comparison of the Lincoln MPO model trip percentages by purpose against the National Household Travel Survey results for "West-North-Central" zone validates the trip generation results of the model.

TABLE 16. TOTAL TRIPS BY PURPOSE FOR BASE YEAR 2004

Trip Purpose	Lincoln MPO	Percentage of Trips - Lincoln MPO	Percentage of Trips -West North Central*
Home-based Work	204,025	13.8%	10.9%
Home-based Shopping	302,640	20.5%	21.8%
Home-based Social/Recreational	168,871	11.4%	13.8%
Other Home-based	343,006	23.2%	20.5%
Not Home-based	459,816	31.1%	33.0%
TOTAL	1,478,358	100.0%	100.0%

*Source: National Household Travel Survey(2001), U.S Department of Transportation, Bureau of Transportation Statistics

The following excerpt from the validation section of the Lincoln MPO travel model documentation.

Percent Error of Traffic Assignment

The percent error of traffic assignment indicates the accuracy with which the transportation model replicates the actual traffic counts. Percent error is the difference between the assigned traffic volumes and the counted traffic volumes divided by the counted traffic volumes. Table 24 displays the percent error by functional classification for Lincoln MPO model.

TABLE 24. PERCENT ERROR BY FUNCTIONAL CLASSIFICATION

FUNCTIONAL CLASS	SUM OF COUNTS	SUM OF ASSIGN	NUMBER OF COUNTS	PERCENT ERROR	PERCENT ERROR TARGET*
Collector	267,981	214,851	51	-19.83%	25.0%
Interstate/freeway	352,200	349,309	21	-0.82%	7.0%
Major Arterial	1,291,935	1,327,761	58	2.77%	10.0%
Major Collector County	69,119	78,299	44	13.28%	25.0%
Minor Arterial	4,693,779	4,649,079	364	-0.95%	15.0%
Principal Arterial (Div)	927,313	962,444	76	3.79%	10.0%
Average Network Stats	7,602,327	7,581,743	614	-0.27%	5.0%

"Calibrating and adjustment of system planning models" December 1990, FHWA

■ Ada and Canyon Counties, Idaho

The following model validation excerpt is from the *2002 Travel Demand Forecast Model Calibration Report for Ada and Canyon Counties*, Report No. 09-2006, accepted by the Transportation Model Advisory Committee on June 22, 2006. It provides an example of a "pass/fail" type of validation:

STATIC MODEL VALIDATION RESULTS

The validation tests are the last step to developing travel demand forecast models and the results determine whether a model is ready for release. This section of the report focuses on the validation process and performance of both models.

24-HOUR MODEL

On May 25, 2004 COMPASS staff and TMAC agreed to use the following four specific static validation criteria for the 24-hour model:

- Minimum of 75% of screenlines should be within their maximum desirable deviation
- Minimum of 75% of roadway links should be within their maximum desirable deviation
- Model-wide correlation coefficient should be greater than 0.88
- Maximum acceptable root-mean-squared-error should not exceed 40%

These criteria come from various sources such as the TMIP Model Validation and Reasonableness Checking Manual, NCHRP Report 255 and Caltrans Travel Forecasting Guidelines. These same guidelines were used to calibrate the peak hour model and presented to TMAC in August 2004.

The following two tables, Table 39 and Table 40, summarize the performance of the regional 24-hour model per facility type. Overall, the model performs reasonably well and achieved a correlation coefficient of 0.96 and R-Squared of 0.90.

Table 39

Percent Root Mean Square Error (RMSE) by Functional Class			
Facility Type	%RMSE	MAX	Validation
Interstate & Ramps	23.9%	< 40%	PASS
Principal Arterials	22.1%	< 40%	PASS
Minor Arterials	39.0%	< 40%	PASS
Collectors	70.4%	< 40%	
Locals	82.2%	< 40%	
Overall	**34.9%**	**< 40%**	**PASS**
Without Locals	34.5%	< 40%	PASS
Without Collectors and Locals	28.9%	< 40%	PASS

Table 40

Model Volume to Actual Count Percent Difference by Functional Class			
Facility Type	Volume to Count % Difference	MAX	Validation
Interstate & Ramps	-0.2%	< 7%	PASS
Principal Arterials	-4.4%	< 10%	PASS
Minor Arterials	-11.5%	< 15%	PASS
Collectors	-16.8%	< 25%	PASS
Locals	11.6%	< 25%	PASS
Overall	**-7.6%**		

The Ada and Canyon Counties travel model validation also included a "dynamic" validation in which model inputs were changed and the results reviewed for reasonability:

Dynamic model validation is a technique to evaluate a model's ability to respond appropriately to various demographic and network changes. Results of the dynamic validation tests are evaluated in terms of meeting an expected trend and magnitude in change. This technique is documented in a white paper by Fehr & Peers and serves as a guide on dynamic validation. COMPASS staff performed three dynamic validation tests on both the 24-hour and peak hour models. Two of the three tests involve evaluating the model's ability to forecast network impacts when the demographic data of a Traffic Analysis Zone (TAZ) are changed. The third test involved the model's response to the deletion of a major link in the regional roadway network. Both models responded appropriately to changes in households, jobs and the network.

24-HOUR MODEL

The 24-hour dynamic validation results are shown in Table 44 and Table 45 for Ada County TAZs 144 and 299. Table 46 and Table 47 summarize the results for Canyon County TAZs 475 and 566. All four TAZs responded as expected to the household and job changes. Figure 43 in Appendix D contains a map highlighting all four TAZs used for the dynamic validation.

Table 44

Results from TAZ 144 (Ada - Suburban)								
Baseline: 236 households				Baseline: 19 Retail Jobs				
Change in Households	Change in Vehicle Trips	Vehicle Trips per Household	Change in VMT	Change in Retail Jobs	Change in Vehicle Trips	Vehicle Trips per Retail Job	Change in VMT	VMT per Retail Job
0	0.00	7.31	0.00	0	0.00	20.40	0.00	164.21
-1	-3.91	7.31	-24.13	-1	-1.97	20.40	-9.07	164.21
-10	-54.26	7.31	-402.99	-10	-23.14	20.40	-179.45	164.23
-100	-526.37	7.31	-3,599.25	Non-applicable				
1	7.31	7.31	25.41	1	2.80	20.40	-15.44	164.21
10	49.91	7.31	2.41	10	26.96	20.40	111.97	164.19
100	530.28	7.31	3,415.34	100	258.68	20.37	1,697.84	163.98
1,000	5,262.83	7.30	36,434.54	1,000	2,562.56	20.12	15,855.75	161.90
5,000	26,354.60	7.25	187,525.56	5,000	12,809.78	19.09	84,627.45	153.43
10,000	52,721.12	7.20	380,311.43	10,000	25,617.21	17.97	170,888.47	144.14

Table 45

Results from TAZ 299 (Ada – Rural)								
Baseline: 34 households				Baseline: 7 Retail Jobs				
Change in Households	Change in Vehicle Trips	Vehicle Trips per Household	Change in VMT	Change in Retail Jobs	Change in Vehicle Trips	Vehicle Trips per Retail Job	Change in VMT	VMT per Retail Job
0	0.00	7.31	0.00	0	0.00	20.40	0.00	164.21
-1	-4.33	7.31	-93.57	-1	-1.48	20.40	-20.75	164.21
-10	-54.04	7.31	-614.51	Non-applicable				
1	6.54	7.31	66.01	1	2.99	20.40	33.64	164.21
10	51.94	7.31	255.76	10	27.22	20.40	293.29	164.19
100	540.91	7.31	5,494.44	100	257.56	20.37	2,791.30	164.00
1,000	5,381.43	7.30	56,559.80	1,000	2,563.19	20.12	29,830.62	162.12
5,000	26,933.90	7.26	293,003.01	5,000	12,807.61	19.09	148,962.79	154.37
10,000	53,886.61	7.21	586,994.77	10,000	25,618.39	17.97	285,941.85	145.71

A section of Eagle Road north of Fairview Avenue was deleted from the base year. Eagle Road is a primary north-south route and Eagle Road/Fairview Avenue is one of the busiest intersections in Ada County - ranked second in 2002 with over 5,500 vehicles in the peak hour. Figure 15 and Figure 16 are maps of the area where a network link was deleted. The model responded appropriately to the removal of a link by diverting trips to other parallel facilities.

Figure 15

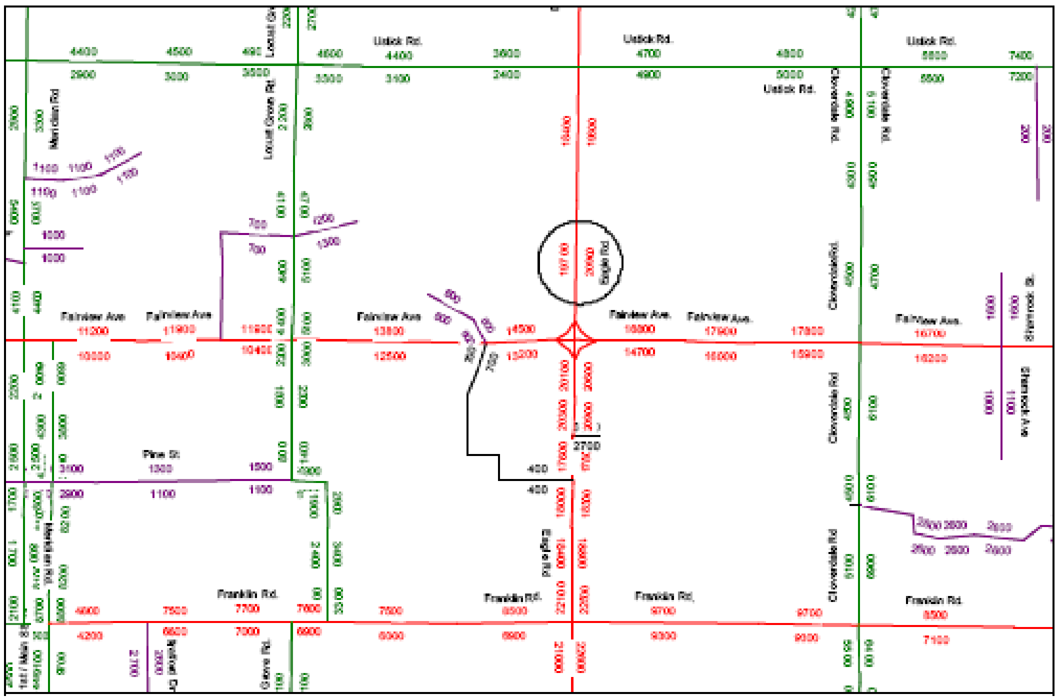
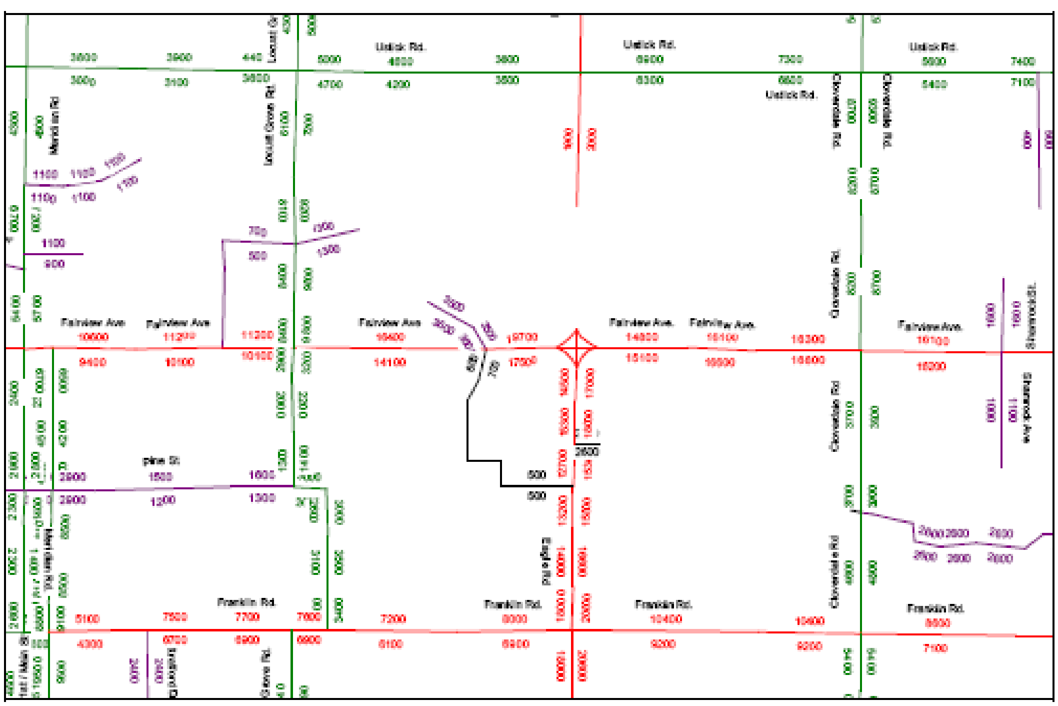


Figure 16

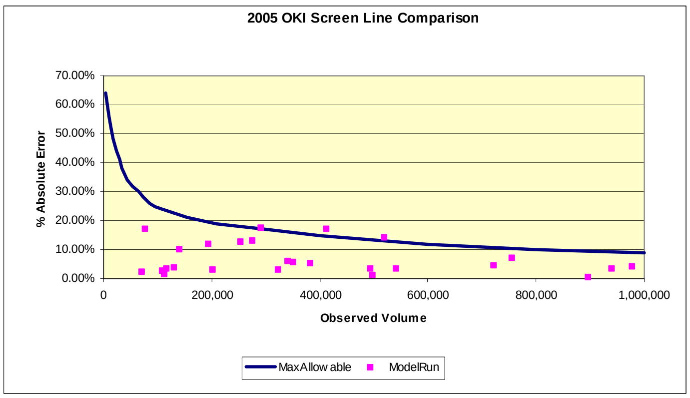


OKI Council of Governments (Cincinnati, Ohio)

The following slides have been extracted from a model validation presentation posted on the OKI COG web site. The slides provide examples of model validation results for several model components.

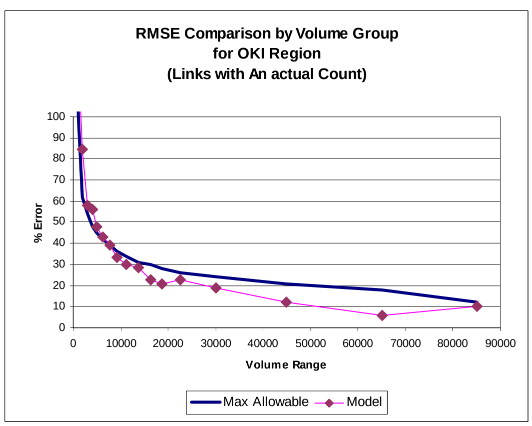
OKI/MVRPC Travel Demand Model Validation – Traffic Assignment (2 of 9)

Volume error % by screenline



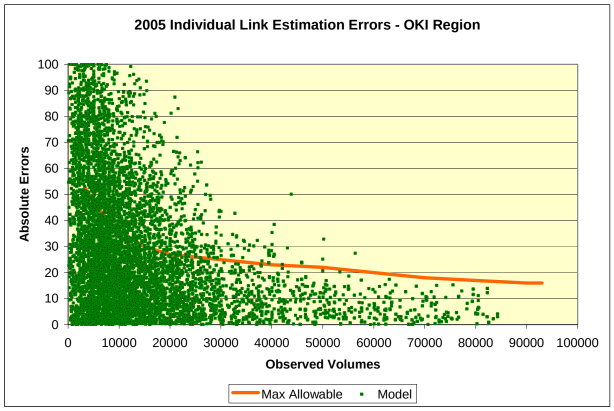
OKI/MVRPC Travel Demand Model Validation – Traffic Assignment (4 of 9)

RMSE comparison by volume group for links with actual counts



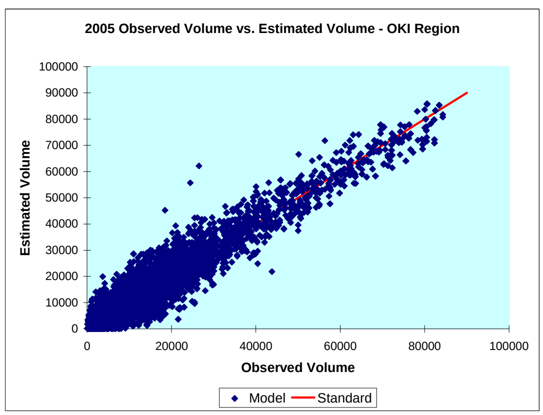
OKI/MVRPC Travel Demand Model Validation – Traffic Assignment (5 of 9)

Individual Link estimation error for all links except centroid connectors



OKI/MVRPC Travel Demand Model Validation – Traffic Assignment (6 of 9)

Link by link comparison between estimated volume and observed Volume for all links except centroid connectors



OKI/MVRPC Travel Demand Model Validation – Trip Generation (1 of 3)

Comparison of daily person trips per household with other metropolitan areas

Trip Purpose	2005 OKI model
HBW	2.018
HBU	0.072
HBO	4.606
HBSCH	0.030
NHB	2.481
Total	9.288

Trip Purpose	Houston 1985 Models	Dallas/Ft. Worth 1984 Travel Survey	Denver 1985 Travel Survey	San Francisco 1985 Travel Survey	Atlanta 1980 Travel Survey	Delaware Valley 1986 Travel Survey
HBW	1.71	2.29	1.96	1.89	1.95	2.27
HBNW	4.80	4.32	3.40	4.49	4.45	4.19
NHB	2.96	2.07	1.97	2.35	1.87	1.64
Total	9.47	8.68	7.33	8.71	8.27	8.10

Source: "Model Validation and Reasonableness Checking Manual", TMIP, FHWA, 1997

OKI/MVRPC Travel Demand Model Validation – Trip Generation (2 of 3)

Comparison of trip rate with national data

	2005 OKI	2001 Nation
Daily Person Trips per Household	9.29	10.49
Daily Person Trips per Person	3.79	4.06
Daily Vehicle Trips per Household	6.86	6.00
Daily Vehicle Trips per Person	2.80	2.32

Source of National data comes from 2001 National Household Travel Survey.

OKI/MVRPC Travel Demand Model Validation – Trip Distribution (1 of 2)

Comparison of travel time and travel distance with 1995 trip survey

Trip Purpose	Travel Distance (Miles)		Travel Time (Minutes)	
	1995 Observed	2005 Estimated	1995 Observed	2005 Estimated
Peak				
HBW	11.7	13.35	23.6	26.26
HBO	6	7.1	12.8	16.02
NHB	6	6.13	12	12.15
Offpeak				
HBW	10.2	10.23	15.1	14.67
HBO	6.2	7.01	10.3	11.08
NHB	5.6	5.59	8.8	8.41

OKI/MVRPC Travel Demand Model Validation – Trip Distribution (2 of 2)

Difference of county to county work trip flow between 2000 CTPP and 2005 model estimated (CTPP-Model estimate)

County	Butler	Clermont	Hamiton	Warren	Boone	Campbell	Kenton	Dearborn	Total Production
Butler	1.65%	-0.09%	-0.72%	-0.41%	-0.07%	-0.07%	-0.15%	-0.05%	0.09%
Clermont	-0.06%	0.60%	0.09%	-0.10%	-0.27%	-0.35%	-0.39%	0.00%	-0.46%
Hamiton	-0.71%	-0.34%	3.94%	-0.22%	0.16%	-0.17%	0.02%	-0.20%	2.46%
Warren	-0.58%	-0.14%	-0.74%	0.14%	-0.04%	-0.07%	-0.10%	-0.01%	-1.55%
Boone	0.02%	-0.03%	-0.04%	0.01%	-0.57%	-0.10%	-0.17%	-0.06%	-0.94%
Campbell	0.01%	-0.07%	-0.37%	-0.01%	-0.04%	0.89%	-0.16%	-0.01%	0.24%
Kenton	0.01%	-0.07%	-0.27%	-0.03%	-0.27%	-0.16%	1.09%	-0.03%	0.28%
Dearborn	0.01%	-0.01%	0.11%	0.00%	-0.31%	-0.03%	-0.14%	0.25%	-0.11%
Total Attration	0.35%	-0.14%	2.01%	-0.63%	-1.41%	-0.06%	0.01%	-0.10%	0.00%

OKI/MVRPC Travel Demand Model Validation – Modal Choice (1 of 4)

Comparison of modal split between 1995 survey and 2005 model estimate – home based work

Mode	Peak HBW		Offpeak HBW	
	1995 survey	2005 model estimated	1995 survey	2005 model estimated
Auto	96.3%	97.2%	97.5%	98.1%
Drive Alone	80.7%	81.5%	81.4%	81.8%
2 Person Auto	11.7%	11.7%	12.1%	12.3%
3+ Person Auto	3.9%	4.0%	3.9%	4.0%
Transit	3.7%	2.8%	2.5%	1.9%
Local Bus	3.4%	2.6%	2.5%	1.9%
Express Bus	0.3%	0.2%	0.0%	0.0%
Total	100.0%	100.0%	100.0%	100.0%

OKI/MVRPC Travel Demand Model Validation – Modal Choice (3 of 4)

Comparison of modal split between 1995 survey and 2005 model estimate – home based other

Mode	Peak HBO		Offpeak HBO	
	1995 survey	2005 model estimated	1995 survey	2005 model estimated
Auto	99.1%	99.4%	99.3%	99.5%
Drive Alone	34.6%	34.6%	42.3%	42.1%
2 Person Auto	33.4%	33.5%	34.3%	34.5%
3+ Person Auto	31.0%	31.2%	22.7%	22.9%
Transit	0.9%	0.6%	0.7%	0.5%
Local Bus	0.9%	0.6%	0.7%	0.5%
Express Bus	0.0%	0.0%	0.0%	0.0%
Total	100.0%	100.0%	100.0%	100.0%

Appendix D – Detailed Interview Summaries

■ Metroplan (Little Rock, Arkansas)

The most recent model validation for the Little Rock area was performed in 2004 using 2000 as the validation year. By local definition, validation focuses only on the performance of the highway assignment model while calibration refers to the process of estimating model variables for the other steps in the process. This definition is based on the concept that the validation demonstrates the level of confidence the user can place in the forecast assignment results.

The actual model validation process focuses on matching:

- VMT for the region, by jurisdiction, by facility type, and by area type;
- Screenline crossings for 32 screenlines; and
- Assignment statistics including RMS error and percent RMS error, and absolute and relative deviations of assigned to observed traffic.

Observed traffic count information for the validation is obtained from Highway Performance Monitoring System (HPMS) counts and summaries, and from traffic count data. The traffic counts cover facilities that account for about one-third of the total VMT in the region.

Validation criteria have been developed from available manuals and resources including the *TMIP Model Validation and Reasonableness Checking Manual*, *NCHRP Report 365, NCHRP Report 255*, and the 1990 FHWA report, *Calibration and Adjustment of System Planning Models*.

Metroplan's model validation efforts are driven by model software updates. The 2004 validation effort was a result of software platform change from TRANPLAN to TransCAD. Another validation effort driven by an update to TransCAD software is just being completed. The current validation effort includes the collection of additional traffic count data and the review and quality assurance of questionable count locations from the 2004 validation.

■ SACOG (Sacramento, California)

The four-step model for the Sacramento region, SACMET, was estimated in 1994, based on 1991 household survey data. An update of the auto ownership model was performed in 2000 based on 2000 household survey data. While SACOG documents the results of their validation tests, they do not set validation standards. This is based on the philosophy that they will not make "corrections" to the travel models simply for the sake of more closely matching an arbitrary standard. The following summarizes the validation tests for the various model components:

- **Trip Generation –** Trip rates by different socioeconomic groups were compared to those estimated using the 1991 household survey (i.e., the model estimation data). Trip rates also were compared to those for other regions.

- **Trip Distribution/Destination Choice –** Destination choice is used for the home-based work trip purpose and a gravity-based trip distribution model is used for the other trip purposes. The home-based work destination choice model results have been compared to 2000 CTPP worker flows and the trip length frequency distributions estimated from the 1991 household survey. The trip distribution models for the other trip purposes were validated by comparing the modeled trip length frequency distributions to the observed distributions from the 1991 household survey.

- **Mode Choice –** The primary validation of the mode choice model was the comparison of modeled to observed mode shares from the 1991 household survey data. The home-based work mode shares also were compared to the usual mode for commuting from the 2000 CTPP data.

- **Time-of-Day –** The SACMET model uses diurnal-direction split factors summarized directly form the 1991 household survey without adjustment. A comparison of modeled traffic flows by time-of-day were compared to observed counts by time-of-day.

- **Highway Trip Assignment –** Daily and time-of-day validations for four time periods are performed. Typical assignment statistics, average link error, percent root mean squared error, and correlation, are summarized for the region by functional class and link volume range. Screenline comparisons also are performed. Aggregate modeled VMT by county has been summed for comparison to county-level HPMS data.

- **Transit Trip Assignment –** The transit trip assignment results are compared to weekday line boardings and, for the LRT system, daily boardings by station. In addition, modeled transfer rates have been compared to 2005/2006 transit on-board survey data.

While backcasting and forecasting are not explicitly performed for model validation, the SACMET model has been run for 2005 and the modeled traffic and transit assignment results have been compared to 2005 traffic count data and 2005/2006 transit boarding count data.

■ MWCOG (Washington, D.C.)

MWCOG's current regional travel model, Version 2.2, is based on household survey data collected in 1994. The model has been updated and validated to varying degrees approximately every other year. The trip distribution, mode choice, and highway assignment components of the Version 2.2 model have been validated.

The primary validation of the trip distribution model has been for the home-based work trip purpose. The home-based work trip distribution was compared to home-based work trip movements estimated from the 2000 Census Transportation Planning Package (CTPP) data. Formal validation of the distribution model for other trip purposes has not been performed; rather, those trip purposes have been validated based on a "secondary review."

The mode choice model also was validated against 2000 CTPP data for home-based work trips by comparing modeled to CTPP estimated transit and auto person trips by jurisdiction. Validation of mode choice other trip purposes was based on a comparison to rail and bus survey data. However, the date of the rail and bus survey(s) is unclear. The year 2000 model validation also included a review of regional transit mode shares and absolute numbers of transit trips by trip purpose.

The highway assignment for Version 2.2 of the model has been validated for 2000 by comparing VMT by state (Maryland, Virginia, and the District of Columbia) and for the region to VMT estimated using HPMS data. This validation excluded VMT on local roads. The model also was validated to 2005 data by comparing modeled VMT by jurisdiction along with screenline crossings for 38 screenlines. The 2005 traffic count data was available for 5,400 count locations.

As implied, some temporal validation of Version 2.2 of the MWCOG model has been performed. The existing validation compares model results for both the year 2000 and 2005. The validation also compared model results for years 2000, 2002, 2005, 2010, 2020, and 2030 to evaluate the performance of the model. All elements of the modeling process were compared for these years. Finally, the Version 2.2 model validation included sensitivity tests of transit fare changes as well as modifications to the highway network.

■ CAMPO (Austin, Texas)

CAMPO currently is in the middle of calibrating and validating a new base year (2005) model based on its most recent travel survey. The work is not expected to be completed until spring of 2009. The travel model used for the 2025 Long-Range Plan update was developed based on data collected in 1998 and 1999. The base (calibration/validation) year for the model was 1997. Information summarized below is based on the *1997 Base Year Travel Demand Model Calibration Summary for Updating the 2025 Long-Range Plan* (May 2000).

The primary validation of the 1997 travel model was performed by assigning the vehicle and transit trips to their respective networks and comparing the results to the 1997 vehicle counts and on-board transit ridership. Assigned VMT was compared to VMT estimated for the Highway Performance Monitoring System for the region and for each of the three counties comprising the CAMPO region. In addition, assigned VMT was compared to counted VMT for:

- Links stratified by facility type;
- Links stratified by area type within each of the three counties; and
- All links.

Finally, modeled screenline crossings were compared to counted crossings for 50 screenlines throughout the region.

CTPS (Boston, Massachusetts)

CTPS does not have a formalized validation report for the regional travel model. They do have a series of documents from corridor studies and New Starts applications where they look at speeds, travel times, and volumes for a base year.

The most recent validation has been to highway volumes, highway speeds and travel times, and transit boardings for 2005/2006. CTPS also relies on CTPP data for validation purposes. Specifically, the home-based work trip distribution and mode shares have been validated to the 2000 CTPP.

CTPS staff recognize the need to develop a validation report and formalize the validation process, but they have not had an opportunity to do this.

SEMCOG (Detroit, Michigan)

The SEMCOG regional travel model has been evolving over time. SEMCOG performs an annual validation of their model to ensure that it is reasonably reproducing current traffic counts; however, the validations also have been tied to improvements in the regional model or regional traffic count database. Validation of Version E-4 of the SEMCOG regional model was performed in 2007; Version E-5 of the model, which currently is under development and will be used for the 2035 plan, will be validated in 2008.

The trip generation and trip distribution components of Version E-4 of travel model were estimated and validated using 1994 household survey data. The mode choice component of the E-4 model was estimated using the 1994 household survey data but was updated and revalidated using results from a 2002 on-board transit survey. The time-of-day model was updated in Version E-3 of the model to better reproduce peaking patterns derived from 2002 traffic count data; the updates and validation procedures have been carried over into the E-4 version of the model.

Model validations focus primarily on reasonably reproducing observed traffic volumes. Validation criteria have been developed from the 1990 FHWA report, *Calibration and Adjustment of System Planning Models*.

SEMCOG standardized the traffic count data collection process used by local agencies in 2002 and has built and maintains an extensive traffic count database for the region. SEMCOG's standardization process includes the specification of acceptable counting procedures, the derivation of local factors to adjust 24-hour and 48-hour traffic counts to average annual daily traffic and average annual weekday traffic, and standard traffic count quality control procedures.

Temporal validation of the regional travel model has not been a primary focus of recent validation efforts. Nevertheless, the availability of traffic count data collected in a consistent manner over a number of years has provided a solid basis for a temporal check of the model. Interestingly, total VMT in the region has shown a decrease from 2002 to 2005 while version E-4 of the model, applied to both 2002 and 2005 socioeconomic and network data, has shown an increase in regional VMT. SEMCOG is waiting for processing of 2006 and 2007 traffic count data to be completed to verify whether the traffic count data will continue to show a downward trend in regional VMT or if the decrease in VMT was a temporary anomaly.

SEMCOG has identified the need for good, recent transit on-board survey data, commercial vehicle data, and external station data. The ability to allocate the resources necessary to collect the data continues to be an issue. SEMCOG staff identified the establishment of Federal guidelines as a means to focus management on the need to collect model estimation and validation data. For example, when FHWA field staff certifies the regional planning process without comment on the modeling process, management support for regional model improvements is decreased.

CAMPO (Raleigh, North Carolina)

The 2006 Triangle Regional Model currently is the validated model for the Raleigh, NC area. The 2006 model validation was driven by a model update performed in support of a New Starts application. Even though the model was updated for a New Starts application, the model documentation for the 2006 model focuses only on the trip distribution and highway assignment components of the model.

The trip distribution component for the 2006 model was validated by comparing average trip lengths by trip purpose to averages estimated from 1995 household survey data. This "validation" step was considered to be secondary and, thus, focused on the reasonability of the modeled average trip lengths in comparison to the 1995 data.

The primary validation focused on traffic assignment results. Modeled to estimated VMT by facility type and modeled to observed traffic count data were compared for facility types, screenlines, and cutlines. Assigned to observed traffic count comparisons included percent deviations, R^2, and RMS error by facility type. It is interesting to note that a target R^2 of 0.88 was set for the region based on the 1990 FHWA report, *Calibration and Adjustment of System Planning Models*.

■ PSRC (Seattle, Washington)

The PSRC model was last estimated in 2000 based on 1999 household survey data. The primary validation for that model was for the year 2000. The 2000 validation is the primary focus of this section. However, PSRC collected a new household travel survey (more than 4,000 households) in 2006 and currently is performing an innovative model validation based on those survey data. Specifically, PSRC staff are validating the model parameters estimated in the 2000 (using the 1999 survey data) using the 2006 household survey data.

The 2000 model validation covered all components of the travel model. The validation was performed by comparing aggregate model results to expanded 1999 household survey results. Specifically, the following validation tests were performed:

- **Trip Generation –** Regional average household trip rates.
- **Trip Distribution –** Average trip duration and length by purpose (in minutes) for the region, trip length frequency distributions by aggregated trip purposes, percent of intrazonal trips by purpose, and percent of trips by district and purpose.
- **Mode Choice –** Mode shares by purpose and mode, home-based work mode shares to the CBD, home-based college mode shares to the University of Washington, shares of total transit trips by trip purpose.
- **Time-of-Day Choice Model –** Proportions of trips by time-of-day and direction by trip purpose and mode of travel (drive alone, shared ride 2, and shared ride 3+); proportions of drive alone trips also compared by income group.
- **Traffic Assignment –** Traffic volumes and speeds aggregated from individual links to corridors, facility types, volume groups, area types, and time periods; 14 screenlines; VMT by facility type, volume group, and area type; VMT by time period (additional check on time-of-day choice model) and corridor; travel times by time period and corridor; total average daily traffic on link-by-link basis (scatterplot).
- **Transit Assignment –** Total boardings by operator and transit volumes on screenlines.
- **Truck Assignment –** Truck volumes on roadways stratified by freeways/expressways and principal arterials.
- **Regional Statistics –** Percent RMS error and R2 by functional class, volume group, and time-of-day.

PSRC has adopted UrbanSim as their land use modeling tool and currently are validating the model by simulating land use changes from 2000 to 2006 (using 2000 as a base condition). Validation is being performed for each of 15 different model components included in UrbanSim.

The emergence of tolling as an issue in the PSRC area has led to significant revalidation of the travel models to ensure that they have reasonable sensitivities to changes in travel costs. The validation has touched all components of the travel models. Much of the validation effort is based on reasonability testing of the model results based on different tolling schemes.

PSRC uses information published the *TMIP Model Validation and Reasonableness Checking Manual*, and the 1990 FHWA report, *Calibration and Adjustment of System Planning Models* to establish their model validation guidelines. At the same time, they are very aware that they must live with “less accurate” models or model components with respect to matching any specified validation guidelines in order to ensure that the models are reasonably sensitive to changes in the transportation system. Since PSRC has over 35 different validation measures and targets, they realize that they will not be able to match all of the targets equally well.

PSRC has identified currency and consistency of validation data as issues that they need to address. The currency of validation data is primarily impacted by the resources available to collect independent model validation data. The consistency of validation data results primarily from PSRC’s reliance on local agencies for traffic count data and the different standards and procedures used by local agencies in the data collection process.

PSRC staff indicated a desire for a national compilation of “best practices” in travel model validation. PSRC suggested that this manual could be similar to the TMIP Survey manual but should be updated regularly to ensure that it is up to date.

Appendix 2

ARC Activity-Based Model Validation Presentation for May 9, 2008 Model Validation Practices Peer Exchange

Planned Activity-Based Model Validation at the Atlanta Regional Commission

ARC Transportation Planning

Presented to:

TMIP Peer Exchange, Washington DC
May 9, 2008

Presented by:

Guy Rousseau
Modeling Manager
Atlanta Regional Commission
grousseau@atlantaregional.com

ARC ABM, Where Are We? The Incremental Approach

ARC Transportation Planning

- **Long-term** choice models estimated & implemented:
  - Auto ownership
  - Work destination choice
  - School destination choice
- **Short-term** models estimated & being implemented:
  - Selection of DAP (daily activity patterns)
  - Generation of tours
  - Destination choice for non-mandatory tours
  - TOD (Time-of-day) choice
  - Intermediate stop location choice
  - Tour / trip mode choice

ARC Transportation Planning

2008 ARC ABM Work Program

- Implement Short-term models:
  - Utility Expression Calculator spreadsheets
  - Logit model specification
  - Data input then read and interpreted by Java to run model
- Develop supporting Java classes ('jar files'):
  - Provide 'plumbing' to handle data structures
  - Monte Carlo simulation
  - Situational variables (previous choices made)
  - Program flow
  - Heuristics required for overall model system

ARC Transportation Planning

2008 ABM Implementation

- Coordinated DAP (daily activity patterns) for all households members choice model
- Fully joint travel/activity choice model (generation and participation sub-models)
- Tour destination choice model for all non-work travel
- Tour mode choice model for all non-work travel
- Tour TOD (time-of-day) choice model for all travel
- Stop-frequency choice model for all tour types
- Stop-location choice model for all tour types
- Trip departure choice model for all tour types, trip purposes, and trip placement in the tour chain
- Parking choice model for auto trips to CBD

Application Model Shell & Distributed Processing

- Stream-line application procedure:
  - Implement model feedback
  - Transit & highway network skimming
  - Distributed application programming for system-level **model calibration**
- Develop application model shell:
  - Population synthesizer
  - Core models
  - Auxiliary models (trucks, externals, special generators)
  - Network assignment & skimming procedures for highway and transit modes

2009: Calibration & Validation

- Develop structural calibration targets from:
  - 2001 household travel survey
  - Traffic counts
  - 2000 CTPP tables
  - Transit ridership data
- Perform model-by-model calibration runs
- Compare to targets (estimated VS observed)
- Adjust parameters through programmatic and manual procedures
- Have ABM up and running by end of 2009, but on a dual/parallel track with trip-based model, via a staged and gradual transition

ARC PopSyn Validation

- Maximize Quality of Synthetic Population, both for Base & Forecast Years
- Robust Validation Procedures
- "Back Cast" Validation To 1990
- Flexible Enough to Use Available ARC Land Use Forecasts

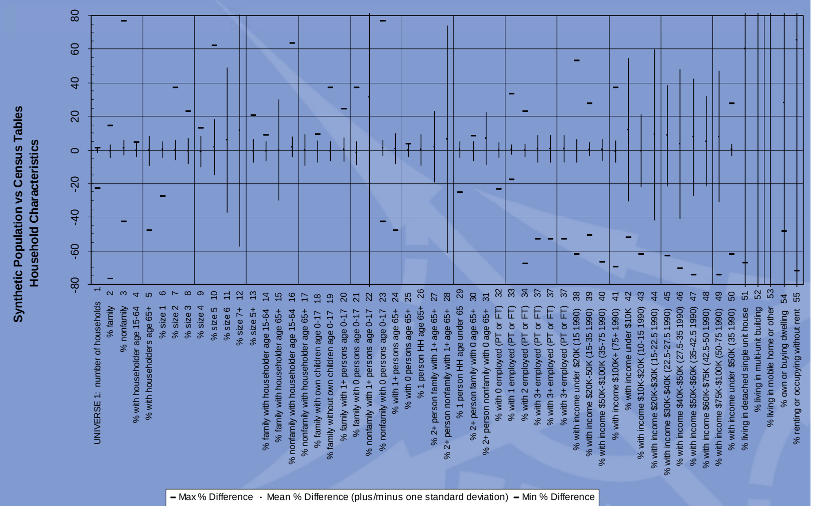
ARC PopSyn Validation

- JAVA Object Oriented Program
- Synthesizes Base Year Population From Census
- Incorporates Available Aggregate Population Forecasts Into Synthetic Forecast Year Population
- Validates Accuracy of Synthesized Member Characteristics at Multiple Levels of Demographic & Geographic Aggregations

Transportation Planning

ARC PopSyn Validator

Design: 316 HH, 20-County, 2000 Base Year, Census Tracts



Transportation Planning

ARC ABM Validation

- **Highway Assignment Validation - Speeds**
  - Obtain reasonable match of both speeds and traffic volumes on roadways
  - Compare modeled speeds to speed study results and/or observed speeds
  - Produce consistency with observed speeds by time-of-day

ARC ABM Validation

- **Trip Assignment Validation – Counts**
  - Compare modeled VMT and HPMS
  - Compare modeled volumes to traffic counts along screenlines
  - Compare modeled volume to traffic count RMSE by volume groups (30% RMSE max for total)
  - Transit assignments verified using boarding (un-linked) transit trip data and transit on-board survey

ABM Transit Validation

- Use maximum desirable deviation curves and link scatter plots
- Scatter plot of observed rail station boardings versus modeled, where modeled boardings would fall in line with observed boardings
- Use maximum desirable deviation formulas from highway assignment validation
- Stations would fall within an acceptable range if same criterion were used for station boardings as highway assignment volumes
- For regional bus boardings, modeled boardings would follow the observed boardings

ARC ABM Link to Transims?

TRANSIMS	Correspondence	ARC ABM
Population Synthesizer		Population Synthesizer
Activity Generator	?	Core Demand Models
Router	±	Conventional SUE or DTA
Traffic Microsimulator	±	Any Traffic Microsimulator
Feedback to Activity Generator through rules	?	Feedback to Core Models through LOS

3 Ways to use TRANSIMS

- The original 'full spirit':
  - Re-engineer feedback mechanism to account for adjustment of destinations & modes
- Hybrid (Portland METRO):
  - Regeneration of locations
  - Mode preferences
  - TOD congestion feedback
- Truncated:
  - Only Router and Microsimulator
  - Demand model (AB/4-step/fixed trip table)
  - ABM to produce sliced trip tables suitable for DTA / traffic microsimulation

ARC Transportation Planning

ARC ABM Link to PECAS?

- 2 Types of Linkages:
  - *Connection*: Feed PECAS land use model estimates of population & employment to ABM
  - *Integration*: Utilize labor flows from PECAS spatial Input-Output model to determine workplace location choice

ARC Transportation Planning

Workplace Location Choice

- Looking at examples from Ohio and Oregon Statewide Models
- Select employment location for every employed person in a household
- Segmentation by occupation & household type
- Selection probability is a function of labor flows forecasted by PECAS

Disaggregate Application for a Fully-Integrated Transport & Land Use Model

- PopSyn would use occupation sector control totals for workers based on needs of industries predicted in the spatial Input-Output PECAS model
- Each worker chooses a workplace TAZ based on the disaggregated probability
- Would allow for full consistency & integration between land use model and ABM, **but calibration can be intense**

Appendix 3

DRCOG Activity-Based Model Validation Handout for May 9, 2008 Model Validation Practices Peer Exchange

Calibrating/Validating Disaggregate Models: What do we Care About? How Far Can We Go?

The building of a disaggregate model may be likened to the purchase of a new pair of spectacles: you will see everything in much clearer focus, and some of it will be pretty ugly.

Confucius

New opportunities bring new problems.

- Troubles with specialty trip:
  - The logit curve does not produce zeros; and
  - Geocoding becomes more important than ever.
- Quotas: enforce them or not:
  - Match total workers with total jobs (by type); and
  - Match school enrollment.

Worrying more about things you could have worried about before, but did not.

- Neighborhood "morphology:" who is living there in the validation year?
- O-D transit assignment.

More detail in the model. More detail in the calibration/validation data?

- Transit time of day (on-board survey).
- Highway time of day (volume and speed).
- Transit trips by purpose (by submode, by line, etc.).
- Validate trip distribution by purpose, or just in aggregate?

Old problems in new clothing:

- Trip distribution – what was once one model is now several: usual workplace location; usual school location; workplace location type; tour primary location, intermediate stop generation; intermediate stop location.
- Time of day choice – ditto: time of day simulation; tour time day choice; trip time of day choice.
- Mode choice – ditto: tour mode choice; trip mode choice.

What more TRUE validation data can we get? How far can we go with it? And how much do we trust it?

- You can do a lot with transit on-board data.
- Equivalent “roadside” auto data? Example: toll users.
- We have highway counts by time of day: can we hope to match them?
- Counts are not perfect, so how hard do we try to match them?
- Is there such a thing as TRUE disaggregate validation data?

Some missing data:

- Household vacancies.
- Self and contract employed.

NOTICE

This report is being distributed through the Travel Model Improvement Program (TMIP).

FHWA-HEP-09-009

FHWA-HEP-09-009

Federal Highway Administration
TMIP
Travel Model Improvement Program