Origin-Destination Trips and Skims Matrices

presented by
César A. Segovia, AICP
Senior Transportation Planner
AECOM

September 17, 2015
Today’s Webinar Content

• Matrix Estimation

• Data sources

• Florida Application
The essence of travel demand models (TDM) is to be able to represent where people begin and end their trips. That requires a spatial context (from point A to point B, on what specific routes, etc.) as well as time, (time of day, duration and delays of the trip, costs).
Time space dimensions

- Depending on the characteristics of the analysis, its scope, the geographic resolution and corresponding time, different techniques will be applied, which in turn will require more or less details in the data inputs.

- A macro-level model involves whole regions, usually at the metropolitan planning organization level, or a Department of Transportation district. Long term planning is the main purpose of these models.

- At the meso-level, the users are dealing with modeling just a sub-area of the macro-level to evaluate specific corridors.

- The micro-level of traffic analysis focus on specific facilities’ geometry and operations.
• Interestingly, data requirements could be as intensive in the macro level as it could be in the micro-level:
  • A macro level requires a big number of perhaps external trips origin-destination (OD) surveys plus the internal observed trips, including large number of traffic count stations.
  • The meso-level may require more detail data on the time distribution of those traffic counts, perhaps to sub-hour periods.
  • A micro-level model involves very detailed geometry of specific facilities, where the spatial-temporal data on delays is critical and have an effect on traffic flows and route choices.
Highway Assignment Levels

Region
Subarea
Tight Subarea
The Need for a Matrix Estimation

- Surveys can only collect data on a limited sample size. A TDM may have 2,000 traffic analysis zones (TAZ), which means there are 4 million cells in its corresponding OD table, while the network only has data on a 1,000 count stations.

- Example 1: Consider building an OD table for truck movements in Florida. There are only a few hundreds count stations with classified counts, plus GPS data from a sample of trucks.

- Example 2: Intelligent Transportation Systems (ITS) equipped facilities can provide detailed data on traffic counts, speeds, traffic flows at sub-hour level. An OD table from a regional model could be available through sub-area extraction, for peak periods. There is a need to match the regional travel patterns to satisfy the observed counts on the project facility.
• From NCHRP Report 765¹: *The idea is simple: find a reasonable OD table that will reproduce known traffic counts. On large networks, there are many different OD tables that will reproduce traffic counts with equal quality, so there is a need for additional information to help choose an OD table. Most algorithms available today supplement the traffic counts with a “seed” OD table that is a best-guess approximation of the desired result. A seed OD table may be one that has been observed in the past, one that has been observed recently but imprecisely, or one developed from principles of driver behavior.*

The OD matrix estimation (ODME) reads an initial, or seeded, trip matrix from which to modify based upon an iterative process of evaluating the ratios of assigned trips to counts on links, modifying the OD trips, and the reassignment of the modified trip matrix until the user-specified number of passes is reached.

Also required as input are the minimum path files, one per iteration, which are generated during the equilibrium highway loading; and this process is continued until a statistically acceptable trip table and assignment is generated.

When a time-variant OD matrix estimation is needed, as in a dynamic traffic assignment (DTA) modeling framework, then the ODME can also be made dynamic. In this case, a dynamic OD table gives the number of trips between each origin zone and each destination zone that start at a particular time. Dynamic OD tables require dynamic traffic counts and a dynamic seed OD table.
When developing OD matrices from counts (as it is usually the case), a good fit to the screenlines used is necessary, but it may not be sufficient. As stated previously, the ODME is mathematically under-determined (more unknowns than knowns), which means that it could have many solutions; i.e., it could be that many OD matrices could perfectly match the counts observed on the links.

The seed matrix used as a starting point (in the ODME iterative process) has a great influence on the final solution; i.e., technically, one can use a seed of 1’s and iterate from there to reach to an acceptable match to the existing counts. However, it may not represent the true OD matrix underlying those counts.

The availability of good seed data, or a high confidence in the seed OD matrix being used, plus a good fit to the counts, and how well the resulting assignment trip lengths match the seed matrix trip lengths are all required elements for a good estimation.

Big data (coming from Bluetooth or GPS devices) can provide help finding another source for these seed matrices. This combination of sources will lead to a carefully developed OD matrix that have enough resemblance to the true OD patterns in the study area.

Good Calibration is Essential

- Traffic flow model, demand, assignment
  — Sequential and iterative approach

<table>
<thead>
<tr>
<th>Goodness of Fit Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>113</td>
</tr>
<tr>
<td>% RMSE</td>
<td>13.6</td>
</tr>
<tr>
<td>MAE</td>
<td>77</td>
</tr>
<tr>
<td>R squared</td>
<td>0.976</td>
</tr>
<tr>
<td>GEH &lt;5</td>
<td>87</td>
</tr>
<tr>
<td>GEH &lt;10</td>
<td>100</td>
</tr>
</tbody>
</table>

Multi-Resolution Research. Presented at the Model Task Force by Dr. M. Hadi (FIU) -March 2014
Origin-Destination Trips and Skims Matrices

Data Sources
Data needs

• **Demand Data**
  – Origin-Destination Trip Tables from an existing model
  – Tables should be broken into different periods of the day: AM peak, PM peak, and Off-peak (Mid-day, Night)
• Tables need to maintain the directionality of OD trips
• For each period, the Tables need to contain sub-tables by Vehicle class (SOV, HOV, Light Truck, Heavy Truck) by toll choice (Toll, Non-Toll)

• **Supply side data**
  - Network data
  - Signal control data
Data Needs

- Counts, speed, occupancy, and travel times based on Intelligent Transportation Systems (ITS) traffic detectors
- Travel time and OD matching measurements based on automatic vehicle identification (AVI) or automatic vehicle location (AVL) or GPS technologies
- Advanced strategy data (incident, work zone, ML pricing, ramp metering rates, etc.)
- Turning movement counts
Data Sources

- SunGuide data (TSS, TVT data, incident, DMS, etc.)
- Central data warehouse (RITIS/STEWARD)
- Weather data
- Managed lane dynamic congestion pricing rates
- Work zones
- Crash data (CAR System and Signal4)
- Signal control and ramp metering
- 511 data
- Private sector data: Inrix, Here (formerly known as Navteq), AirSage, etc. Note: FDOT has HERE and INRIX data available for public agencies use.
- Statistics office data (FDOT)
- Transit agency data
- AVI data (Bluetooth, Wi-Fi, ETC)
- Connected vehicles
Data Sources: PTMS and ITS

- Whenever data is available, comparing Portable Traffic Monitoring Site (PTMS) and ITS count data may improve the reliability of the data. Manual counts of recorded videos could be conducted to determine the accuracy of the data. It was found in previous studies that the manual counts are closer to ITS data than the PTMS during the peak hours.

- The ITS data can be obtained from the Statewide Transportation Engineering Warehouse for Archived Regional Data (STEWARD). This system was developed as a proof-of-concept to centrally archive data from traffic management centers around Florida in a practical manner. The effort concentrated on archiving information from the SunGuide traffic sensor subsystem (TSS) and the travel time subsystem (TVT). The STEWARD database contains summaries of traffic volumes, speeds, occupancies, and travel times aggregated by 5-, 15-, and 60-minute periods, as requested by the user. Using a web-based interface, the user can specify date and time ranges and detector locations for which the data is needed. The user can also download all generated reports in comma-delimited formats, which can be easily imported into database management tools. Instructions to get access at:
  http://www.dot.state.fl.us/trafficoperations/ITS/RITIS.shtm
• Bluetooth Detector (Traffax) has 100-meter range and consecutive 5-second scan windows.
• “Discoverable” bluetooth devices are detected and time-stamped.
• Bluetooth devices have a unique MAC ID.
• Chronological MAC IDs are linked together to form “trips”.
• Data analysis provides study area speed profiles and origin-destination estimates. More on this later...

<table>
<thead>
<tr>
<th>MAC ID</th>
<th>Device</th>
<th>Date</th>
<th>Data Cleaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMPLE</td>
<td>M</td>
<td>9:36:11 PM</td>
<td>Removed</td>
</tr>
<tr>
<td>SAMPLE</td>
<td>M</td>
<td>9:36:16 PM</td>
<td>Removed</td>
</tr>
<tr>
<td>SAMPLE</td>
<td>M</td>
<td>9:36:21 PM</td>
<td>Removed</td>
</tr>
<tr>
<td>SAMPLE</td>
<td>M</td>
<td>9:36:26 PM</td>
<td>Retained</td>
</tr>
</tbody>
</table>
40 detectors deployed in the study area

Duplicate detectors placed to reduce error

Deployment designed primarily to sequester interchanges

Data collection occurred May 9 – 16, 2013

Estimated penetration rate ranges from 3%–9%
How to: From Bluetooth to OD matrix

This is the actual data from one bluetooth site

<table>
<thead>
<tr>
<th>ID</th>
<th>gld</th>
<th>Blueld</th>
<th>bDay</th>
<th>bTime</th>
<th>bDayTime</th>
<th>Siteld</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>96 00:07:04:E6:EB:43</td>
<td>8/21/2014 1:33:54 AM</td>
<td>8/21/2014 1:33:54 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>97 00:07:04:E6:EB:43</td>
<td>8/21/2014 1:34:03 AM</td>
<td>8/21/2014 1:34:03 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>109 00:1E:3D:35:D6:45</td>
<td>8/21/2014 1:35:07 AM</td>
<td>8/21/2014 1:35:07 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>110 00:1E:3D:35:D6:45</td>
<td>8/21/2014 1:35:12 AM</td>
<td>8/21/2014 1:35:12 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>111 00:1E:3D:35:D6:45</td>
<td>8/21/2014 1:35:17 AM</td>
<td>8/21/2014 1:35:17 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>112 00:1E:3D:35:D6:45</td>
<td>8/21/2014 1:35:22 AM</td>
<td>8/21/2014 1:35:22 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>113 00:1E:3D:35:D6:45</td>
<td>8/21/2014 1:35:27 AM</td>
<td>8/21/2014 1:35:27 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How to: From Bluetooth to OD matrix

Combining all detector sites’ readings to form a “trip”

This driver was not detected at Site #5. The data records show that it went from site #3 to site #7, “jumping” over site #5. A couple of records added to fill in the “gaps”.
Then, by adding up all “TripId”s from every origin to every destination...it is then possible to find the percent distribution of trips.
How to: From Bluetooth to OD matrix

Those percentages shown in a map for site #3
• Data from traffic detectors should be screened carefully to avoid inconsistencies; e.g., the addition of on-ramps and subtraction of off-ramps to estimate the expected volume for the station can be used as a reference when assessing the accuracy of the measurements.

• Whenever data is available, comparing PTMS and ITS count data may improve the reliability of the data.

• For different purposes, a specific day or an average of all repetitive days may be used for calibration. Using the median day data may be better than using the averages, since the averages do not represent any of the real-world days.
Origin-Destination Trips and Skims Matrices

Florida Application
The specific objectives of this research were as follows:

1. Develop procedures for utilizing detailed data from multiple sources to calibrate the supply and demand sides of DTA models for use in modeling ML.
2. Develop procedures for the use of simulation-based DTA to model ML.
3. Assess the performance of the use of DTA compared to STA in ML modeling.
4. Test the ability of a DTA tool (Cube Avenue from Citilabs, Inc.), which is a strong candidate for use in Florida to model ML.
5. Demonstrate the application of the developed procedures to a ML implementation in Florida.

In this presentation, we are interested in the data and processes used to estimate the trip tables by class of vehicle and time periods.
Detector data collected from existing Intelligent Transportation System (ITS) operated by the regional traffic management center was critical for the demand estimation, model calibration, and validation in this effort. The corridor of interest is instrumented with microwave detectors every 0.3 to 0.5 mile that report volume, speed, and density measurements in 20-second intervals for each lane. This data was supplemented by measurements from the Portable Traffic Monitoring Sites (PTMS) ramp counts, from the Statistics Office of the Florida Department of Transportation (FDOT). The PTMS include 15-minute ramp counts for two or three days per year. No speed or classification data are provided. Ramp counts obtained from the PTMS and ramp metering detectors represent the total origin and destination demand on the linear network and are very useful in the demand estimation process.

In this study, it was found that there was an acceptable match between the used ITS and PTMS counts for the ramps. The PM peak on the mainline showed that the PTMS reported higher volumes, compared to ITS data. Manual counts of recorded videos were conducted to determine the accuracy of the data. It was found that the manual counts are closer to ITS data than the PTMS during the PM peak.
The initial source of demand in this research is a trip table for a peak period, extracted from a regional demand forecasting model (the Southeast Florida Regional Planning Model (SERPM) model in the case of the I-95 managed lane case study).

The next step is to convert the three-hour PM peak-period matrices obtained from the SERPM to 15-minute matrices using distribution factors that reflect the proportion of the trip tables for each 15 minutes of the day obtained, based on trip counts.

The next step is to adjust these matrices using the Cube Analyst static matrix estimation program. This matrix estimation process performs the estimation by considering a number of input parameters based on the static assignment of Cube Voyager. This process applies a maximum likelihood approach to optimize the trip tables based on the deviation between traffic counts and assignment results and initial (seed OD) matrices.
Application: I-95 ML, Static O-D ME
Application: I-95 ML, Dynamic O-D ME

• Instead of running Analyst with Highway assignment module for 12 consecutive periods of 15-minute intervals, Analyst Drive (the dynamic OD estimation tool) runs during the whole model period, coupled with the Cube Avenue assignment module.

• In the single Cube Avenue run, the model period is divided into 15-minute intervals. This procedure is supposed to be superior to static OD estimation, because Cube Avenue models the queues and queue spillbacks, and thus can capture the effects of congestion on subsequent time intervals.
Recommended improvements to the Analyst OD estimation process are listed as follows:

• Incorporating speed, density, and/or queue length in the objective function of Analyst.

• Allowing the user to specify lower and upper bounds for each OD pair cell (there is already a global parameter that is applied to all cells yet cannot be varied by cells).

• Allowing the user to better control the temporal variability of the results. In the Analyst optimization tool that is based on static assignment, the temporal variation cannot be controlled. The optimization process can achieve totally different local optimal solutions for sequential intervals, since the optimization does not guarantee achieving global optimal.

• Allowing the user to keep the proportionality between specific OD pairs (e.g., from all of the trips originated from I-95, with 30% directed to SR 836 and 30% destined to the Florida Turnpike).

• Incorporating zonal trip end, partial matrix, partial trips, and confidence matrix¹ in Analyst Drive for dynamic demand estimation.

1. Analyst Drive has this ability in the latest version
Questions are guaranteed in life; Answers aren't.
Contact info

Gabrielle.Matthews@dot.state.fl.us

Cesar.Segovia@dot.state.fl.us

Cesar.Segovia@aecom.com

Thank You!