Phase 1 of the Statewide Model Update: Data Development
Statewide Model Update

Existing Statewide Model
• 2003 Base – 2030 Horizon
• Only Total Daily Traffic
• Limited Network Coverage
• Limited Sensitivity
  – Re-routing Only

New Statewide Model
• 2010 Base Year – 2040 Horizon Year
• Peak Hour and Daily Traffic
• Expanded Network Coverage
• New Sensitivity to:
  – Network changes
  – Induced demand
  – Alternative future land use scenarios
  – Population changes (aging, etc.)
• Version 3 – Commodity Flow Modeling
Statewide Model Update: Phases 1 & 2

Phase 1: Data Development (completed in April)
- New, Expanded Network
- New, More Detailed Zone System
- Obtain & Process Socioeconomic Data
- New Socioeconomic Forecasts
- Obtain & Process ATRI Truck GPS Data
- Combine NHTS & MPO Household Travel Survey Data

Phase 2: Model Development (begun in February)
- New Trip-based Model
- Time-of-Day Modeling (peak hour volumes)
- Destination Choice Models (greater accuracy)
- Possible Pivot-Point Structure (greater accuracy)
- Truck/Freight Modeling still being scoped
- Post-processing for Performance Measures (access to jobs, hospitals, etc.)
# Zone Size and Network Coverage

<table>
<thead>
<tr>
<th></th>
<th>Ohio</th>
<th>Iowa</th>
<th>Indiana</th>
<th>Tennessee v1</th>
<th>Tennessee v2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>11,500,000</td>
<td>3,100,000</td>
<td>6,500,000</td>
<td>6,500,000</td>
<td>6,500,000</td>
</tr>
<tr>
<td>Road Miles*</td>
<td>42,000</td>
<td>45,000</td>
<td>19,000</td>
<td>9,421</td>
<td>32,546</td>
</tr>
<tr>
<td>TAZ in state</td>
<td>3,660</td>
<td>1,866</td>
<td>4,690</td>
<td>1,222</td>
<td>3,293</td>
</tr>
<tr>
<td>Total TAZ</td>
<td>5,116</td>
<td>3,314</td>
<td>4,831</td>
<td>1,397</td>
<td>3,684</td>
</tr>
<tr>
<td>Pop / TAZ*</td>
<td>3,200</td>
<td>1,600</td>
<td>1,400</td>
<td>5,300</td>
<td>2,000</td>
</tr>
<tr>
<td>Acres / TAZ*</td>
<td>12.2</td>
<td>30.2</td>
<td>7.8</td>
<td>34.5</td>
<td>12.8</td>
</tr>
<tr>
<td>Miles / Acre</td>
<td>0.9</td>
<td>0.8</td>
<td>0.5</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Pop / Miles</td>
<td>270</td>
<td>70</td>
<td>340</td>
<td>690</td>
<td>200</td>
</tr>
<tr>
<td>Miles / TAZ</td>
<td>11.5</td>
<td>24.1</td>
<td>4.1</td>
<td>7.7</td>
<td>9.9</td>
</tr>
</tbody>
</table>

*in state

- New model has **triple** the network and zones
Version 2 Network and Zones
Version 2 Network and Zones
Network Development
Defining the Network

How do we decide what to include in the network?

**Old Model**
- Interstates & Principal Arterials

**New Model**
- Started with minimum criteria
  - Anything in the old model
  - Anything in the National Highway Planning Network (NHS, etc.)
  - All minor arterials
- Want network coverage one class lower than desired forecasts
- Began to look at TRIMS data, to consider volume thresholds, etc.
- Found the TRIMS Traffic layer (e.g., roads with TDOT traffic counts) has roughly the right level of network coverage
- Had to add ~100 links to minimum criteria
- May still add/remove a small number of roads in Phase 2 to ensure good loadings / balance with TAZ layer
Defining the Network

New vs. Old Network

TRIMS Traffic Layer
Network Topology: Connectivity & Routing

Model Requirements
• Connected and routable network

Options
• Connect TRIMS GIS layers – Not enough time & budget
• TN OIR E99 layer – Not ready until 2014
• TeleAtlas network – Chosen

Issues
• How to connect with TRIMS
• TRIMS missing ramps
• Elevation (Z) data
Network Attributes

Chosen Attributes

- Design
  - Direction
  - Divided
  - Access Control
  - Ramp
  - Lanes
  - TWTL
  - Reversible lane
- Lane width
- Shoulder width
- Terrain
- Water Xing
- RR Xing
- Speed Limit

- Administrative
  - Name
  - Functional Class
  - Ownership
  - County
  - State

- Intersection
  - Control Type

Missing Attributes

- Most attributes substantially complete from TRIMS
- 90% of roads missing speeds were rural minor arterials
- Speed was missing on roughly half of this class
- No volume / geographic pattern – ok to impute

Traffic

- AADT
- Peak hour %
- MU Truck %
- SU Truck %

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed Limit</td>
<td>15.94%</td>
</tr>
<tr>
<td>Divided</td>
<td>0.01%</td>
</tr>
<tr>
<td>Lanes</td>
<td>0.03%</td>
</tr>
<tr>
<td>AADT</td>
<td>0.05%</td>
</tr>
</tbody>
</table>
Network Development Process

Getting the TRIMS attributes on the routable TeleAtlas network…
  • Develop a Least-Common-Denominator (LCD) TRIMS line layer network with a nodes anywhere a chosen attribute changes
  • Get all the TRIMS attributes onto the single LCD layer
  • Simplify and reduce the LCD representation above if possible
  • Develop a common segmentation between LCD TRIMS and TeleAtlas
  • Pass the TRIMS attributes over onto the newly segmented TeleAtlas layer
  • Simplify the newly segmented TeleAtlas layer (remove unnecessary nodes)
TAZ Development
A Good Lookin’ TAZ

What should a TAZ look like? What makes for a good TAZ?

**Traditionally**
- Zone boundaries conform to the network
- And other boundaries, maybe
- And homogenous land use, maybe

**Travel Sheds**
- Zones as catchment areas around network
- Borrowed from hydrology
- First used for TAZ in NW 20+ years ago
- Increasingly common in statewide models
- Clearer relationship to the network, less ambiguity about loading points / centroid connectors
- Better able to represent distinct rural and small urban zones
- Take other boundaries more seriously
What are the statewide TAZ made of?

**Urban**
- Aggregations of MPO zones (except conform to 2010 Census geography)
- Mix of traditional and travel sheds

**Rural / Small Town**
- Aggregations of Census blocks
- Less traditional, mostly travel sheds
Putting the Puzzle Pieces Together

How do you group MPO zones / Census blocks into SWM TAZ?

Two step process

• First identify boundaries TAZ should not cross
  - County boundaries
  - Place boundaries (loose)
  - Major Rivers
  - Freeways
  - Railroads
  - Major ridgelines / slopes

• Then within the areas defined by these boundaries, group building blocks (blocks/MPO TAZ) into travel sheds around largest (non-freeway) facilities
  - Estimate number of desired zones and choose corresponding number of corridors, starting with highest AADT
  - Successively buffer around each corridor in increments (0.5 mi) until all blocks/MPO TAZ are assigned to a travel shed
  - Manually review and clean
Easy Criteria

Counties

Places
Harder Criteria

Slopes, Ridgelines and Water Features

- Plenty of water layers, but how to define “major”
- No canned “ridgeline” layers
- Created ridgelines by processing DEMs
- Tried to define “internal” criteria
- Instead, used visual inspection against TeleAtlas, looking for network gaps
- Found slopes more a barrier than ridges in some areas
Types of Zones

Six Categories
1. TN Rural
2. TN MPO
3. MPO outside TN
4. Single county outside TN
5. Multi-county outside TN
6. External stations
Socioeconomic Data Development
Demographics

**Decennial Census**
- Population, Households, Children, Seniors
- Block level data

**American Communities Survey**
- Workers, Vehicles, Income
- Block group level available
- Disaggregate to blocks proportionally to households
## Employment Categories

### Industry Categories

- Using standard 20 two digit NAICS categories for data development to support commodity flow modeling in Version 3

<table>
<thead>
<tr>
<th>NAICS Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Agriculture, Forestry, Fishing and Hunting</td>
</tr>
<tr>
<td>21</td>
<td>Mining, Quarrying, and Oil and Gas Extraction</td>
</tr>
<tr>
<td>22</td>
<td>Utilities</td>
</tr>
<tr>
<td>23</td>
<td>Construction</td>
</tr>
<tr>
<td>31-33</td>
<td>Manufacturing (31, 32, 33)</td>
</tr>
<tr>
<td>42</td>
<td>Wholesale Trade</td>
</tr>
<tr>
<td>44-45</td>
<td>Retail Trade (44 &amp; 45)</td>
</tr>
<tr>
<td>48-49</td>
<td>Transportation and Warehousing (48 &amp; 49)</td>
</tr>
<tr>
<td>51</td>
<td>Information</td>
</tr>
<tr>
<td>52</td>
<td>Finance and Insurance</td>
</tr>
<tr>
<td>53</td>
<td>Real Estate and Rental and Leasing</td>
</tr>
<tr>
<td>54</td>
<td>Professional, Scientific and Technical Services</td>
</tr>
<tr>
<td>55</td>
<td>Management of Companies and Enterprises</td>
</tr>
<tr>
<td>56</td>
<td>Administrative and Support and Waste Management and Remediation Services</td>
</tr>
<tr>
<td>61</td>
<td>Educational Services</td>
</tr>
<tr>
<td>62</td>
<td>Health Care and Social Assistance</td>
</tr>
<tr>
<td>71</td>
<td>Arts, Entertainment, and Recreation</td>
</tr>
<tr>
<td>72</td>
<td>Accommodation and Food Services</td>
</tr>
<tr>
<td>81</td>
<td>Other Services, except Public Administration</td>
</tr>
<tr>
<td>92</td>
<td>Public Administration</td>
</tr>
</tbody>
</table>
Employment Data Sources

**MPOs**
- Total employment estimate at MPO TAZ level
- Not full two-digit NAICS breakouts

**InfoGroup**
- Purchased data for all Tennessee – available to MPOs
- Individual business with lat, long locations
- Based on phone surveys, aggregated data

**LEHD**
- Freely available federal data
- Employment by NAICS category by Census block
- Based on administrative (tax) records, with some ‘fuzziness’ for privacy

**BEA**
- Freely available federal data
- Total employment by NAICS category at County level

**Woods & Poole**
- Purchased data including employment forecasts consistent with BEA
Urban Employment Allocation Process

Iterative Proportional Fitting (IPF)

- For each county:
- Scale MPO total employment by TAZ to 2010 BEA (if necessary)
- Apply average of InfoGroup and LEHD 2-digit NAICS breakout within each zone to create seed distribution of employment by TAZ by industry
- Apply IPF to seed distribution
- Resulting employment must
  - Respect MPO TAZ total employment (scaled to BEA)
  - Match county level BEA totals by industry
- Some limited manual cleaning/adjustment was necessary
Using InfoGroup and LEHD Together

**Cleaning**
- Compare differences and correlations look for outliers

**Combining**
- Both InfoGroup and LEHD account for roughly 85% of BEA
- If they are independent, together they would account for 98% of BEA
- Research in Ohio suggests they are close to independent
Rural Employment Allocation Process

Using InfoGroup and LEHD together
• First at the TAZ level, identify any cases where InfoGroup (IG) and LEHD differ by > 200 employees for any industry
  – Determine whether to use IG, LEHD or split the difference based on IG and LEHD coverage of that industry in that county or manual investigation for very large discrepancies
  – Replace IG and/or LEHD estimate with chosen value to create ‘cleaned’ versions
• Second, choose how to use the two datasets together and scale them to BEA
  – Using ‘cleaned’ IG and LEHD calculate the ratio of their sum to the BEA total for each county and industry
    If < 1.5, scale maximum of ‘cleaned’ IG, LEHD to BEA
    If between 1.5 and 2.5, scale average of ‘cleaned’ IG, LEHD to BEA
    If > 2.5, possibly scale minimum of ‘cleaned’ IG, LEHD to BEA
Socioeconomic Forecasts
County Control Totals

Sources of Forecasts
- UT’s Center for Business and Economic Research (CBER) – population only
- Woods & Poole
- MPO forecasts
- Historic growth rates and trends

Recommended Control Totals
- If 2 sources (rural employment)
  - If W&P within 10% of historic, use it, otherwise average with historic
- If 3 sources
  - use middle estimate
- If 4 sources (urban population)
  - If MPO forecast is not highest or lowest, use it
  - If MPO forecast is highest, use second highest
  - If MPO forecast is lowest, use second lowest

Anderson County Population Projections

- Historic
- CBER
- W&P
- Lowest Post ’70
- Avg. ’70-’10
- Highest Post ’70
Allocation to TAZ

**MPO areas**
- Use MPO growth allocations
- Absolute growth may not match exactly if control totals differ, but same pattern will be assumed

**Non-MPO areas**
- Population allocation where growth occurred between 2000 and 2010
  - Except imposing a floor of 30% of 2010 population in declining areas
- Employment allocation an average of two allocation processes:
  - Scaling 2010 employment to 2040 control total
  - Allocate future growth where growth has occurred between 2002 and 2012 (according to LEHD)
Employment Density

PERSONS PER SQUARE MILE

0 - 5  5 - 25  25 - 50  50 - 100  100 - 500  500+

2010

2040
ATRI Truck GPS Data
ATRI Truck GPS Data

What’s ATRI?

**American Transportation Research Institute (ATRI)**

- non-profit funded by the trucking industry
- Receives over 4 Billion GPS truck positions annually from member organizations
- Cannot disclose the individual raw truck traces, but can provided processed data products which avoid disclosure
- Basis of FHWA’s Freight Performance Measures Webtool
- Used for major corridor studies, I-95, I-70
- Incorporated in Indiana & Iowa’s statewide models
- Will be primary basis of truck model in v2 Tennessee model
Indiana Experience

Data
• Eight week sample
• 16 million records
• 305,000 trucks
• 2 million truck trips

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>2006 Model</th>
<th>2010 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>6,689</td>
<td>5,898</td>
</tr>
<tr>
<td>Avg. Count</td>
<td>1,379</td>
<td>1,264</td>
</tr>
<tr>
<td>RMSE</td>
<td>69.3%</td>
<td>60.6%</td>
</tr>
<tr>
<td>Avg Error</td>
<td>5.4%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>MAPE</td>
<td>74%</td>
<td>42%</td>
</tr>
</tbody>
</table>

Method
• Used existing commodity-flow based model to pivot off of expanded ATRI data
Data Processing

What constitutes a stop?

Anonymized GPS records converted to ODs

- Criteria based on speed and time
- Duration of a stop necessary to avoid counting traffic stops as destinations

<table>
<thead>
<tr>
<th>from TAZ</th>
<th>to TAZ</th>
<th>distance</th>
<th>time</th>
<th>elapsed time</th>
<th>speed</th>
<th>status1</th>
<th>status2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>101032</td>
<td>66.0</td>
<td>57.7</td>
<td>57.7</td>
<td>68.6</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>101032</td>
<td>101033</td>
<td>16.3</td>
<td>14.3</td>
<td>72.0</td>
<td>68.6</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>101033</td>
<td>101015</td>
<td>26.8</td>
<td>27.9</td>
<td>99.9</td>
<td>57.5</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>101015</td>
<td>101015</td>
<td>0.0</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
<td>stopped</td>
<td>stopped</td>
</tr>
<tr>
<td>101015</td>
<td>101015</td>
<td>0.2</td>
<td>2.7</td>
<td>7.7</td>
<td>5.2</td>
<td>stopped</td>
<td>stopped</td>
</tr>
<tr>
<td>101015</td>
<td>101015</td>
<td>0.3</td>
<td>9.8</td>
<td>17.5</td>
<td>2.0</td>
<td>stopped</td>
<td>stopped</td>
</tr>
<tr>
<td>101015</td>
<td>101015</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>28.2</td>
<td>moving</td>
<td>stopped?</td>
</tr>
<tr>
<td>101015</td>
<td>2035</td>
<td>37.1</td>
<td>60.0</td>
<td>60.3</td>
<td>37.1</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>2035</td>
<td>18099</td>
<td>67.8</td>
<td>65.4</td>
<td>125.7</td>
<td>62.2</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>18099</td>
<td>27006</td>
<td>5.9</td>
<td>5.4</td>
<td>131.1</td>
<td>65.3</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>27006</td>
<td>18023</td>
<td>10.0</td>
<td>15.9</td>
<td>147.0</td>
<td>37.8</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>18023</td>
<td>18023</td>
<td>0.0</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
<td>stopped</td>
<td>stopped</td>
</tr>
</tbody>
</table>
Data Cleaning

Need to clean/filter data for several reasons

- GPS blips
  - GPS location jumps from one place to another in a way that it could not be travel
- Start/End time
  - Trip fragments / partial trips in progress at beginning/end of sample periods
- Internal circuity
  - Correct for undetected stops
- Missing location data
- Long intrazonal trips (undetected stops)
Data Expansion

It’s a big sample, but it’s still a sample.

Simple Scaling
- Single uniform expansion factor
  - sample truck VMT to HPMS truck VMT

Preliminary Weighting
- Varying weights by trip length
- Weights developed from another study

Final Weighting
- Varying weights by
  - Region
  - Trip length
- Weights developed by analyzing results of ODME
Tennessee ATRI Data

ATRI data includes ~11% of the multi-unit trucks on the road for 56 days

- Observations over 8 weeks in 2013
- Over 234,000 individual trucks
- Over 6.5 million truck trips (5.7 million after cleaning)
- DOT/RSG estimate 14 million daily multi-unit truck VMT in TN
- ATRI data (cleaned) contained 84.1 million truck VMT in TN over 56 days for a daily average of 1.5 million
- This represents 10.7% of the estimated total.
  - To produce the simple scaled OD table, this meant factoring down the raw ATRI data by 0.1672 to represent a daily number.
Household Survey Data
Combining NHTS & MPO Surveys

Datasets
- NHTS Add-On for Tennessee
  - Oversampled rural areas
- MPO surveys
  - Complete/complement NHTS
  - Used Middle TN & Knoxville

Re-weighting and combining
- Controls
  - Region
  - Household size by vehicles
  - Person age
- Iterative Proportional Fitting

ACS vs Unweighted

ACS vs Weighted
Phase 2 of the Statewide Model Update
Phase 2: Model Development

**Advanced Trip-based Passenger Model**
- Advanced trip generation
- Destination choice models
- Peak hour models

**Truck / Freight Model**
- Still being scoped

**Validation**

**Post-processing**
- Traffic statistics
- Post-processing
Advanced Trip Generation

Non-linear Regression Models

- Allow multiple explanatory variables
  - Effect of area type /accessibility
  - Effect of seniors / children
- Capture both rational non-linearities
  - Diminishing returns to scale
  - Interaction effects

Poisson Distributed Household Variables

- Reduces aggregation as in cross-class
- But don’t require stratification curves, etc.

Source: Wikimedia Commons
Destination Choice Models

Account for More Factors
- Number of Attractions
- Travel Time / Impedance
- Effect of Residence Location on Willingness to Travel
- Psychological Boundaries
  - River Crossings
  - Ridgeline Crossings
  - Major Highway Crossings
  - State / County Line Crossings
- Walkability of Destination
- Mixture of Land Uses at Destination
- Convenience for Trip-Chaining
- Spatial Auto-correlation Effects
Peak Hour Models

Nested Logit Models

• Must account for duration of long trips
• Upper nest determines at least some portion of the trip occurs in the AM peak hour, PM peak hour or both
• Lower nest determines how much of the trip occurs in the peak hour
• Will consider differences in peak hour factors related to
  - area type / accessibility of origin and destination
  - trip length
  - region
Truck Model

Model Structure
• Three step, like traditional passenger models
• Segmentation
  – by vehicle type
    * light commercial vehicles
    * single unit trucks
    * multi-unit trucks
  – by trip type
    * commercial passenger trip
    * service delivery trip
    * freight delivery trip

Developed based on ATRI data, pivot off of ATRI data
Validation

Demand Validation
• Generation Rates
• Trip Lengths
• JTW Patterns

Assignment Validation
• Will produce similar statistics as for MPO models
• Different criteria for statewide models

%RMSE for Statewide Models from NCHRP 08-36-91

<table>
<thead>
<tr>
<th>Volume Range</th>
<th>Average</th>
<th>Avg + 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5000</td>
<td>92.2</td>
</tr>
<tr>
<td>5000</td>
<td>10000</td>
<td>51.2</td>
</tr>
<tr>
<td>10000</td>
<td>20000</td>
<td>46.7</td>
</tr>
<tr>
<td>20000</td>
<td>30000</td>
<td>32.4</td>
</tr>
<tr>
<td>30000</td>
<td>40000</td>
<td>29.1</td>
</tr>
<tr>
<td>40000</td>
<td>50000</td>
<td>18.0</td>
</tr>
<tr>
<td>50000</td>
<td>60000</td>
<td>18.6</td>
</tr>
<tr>
<td>60000</td>
<td>+</td>
<td>22.2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>54.5</td>
</tr>
</tbody>
</table>
Vince Bernardin, PhD, RSG
Statewide Model Update Project Manager
Vince.Bernardin@RSGinc.com
812.200.2351