Building and Calibrating Activity-Based Models: Stories from the Trenches

Erik Sabina, Regional Modeling Manager
Suzanne Childress, Senior Modeler

Project Funding

- STP Metro federal and RTD match
- Consultants - $1M
- In-house – somewhere between $800k - $1M

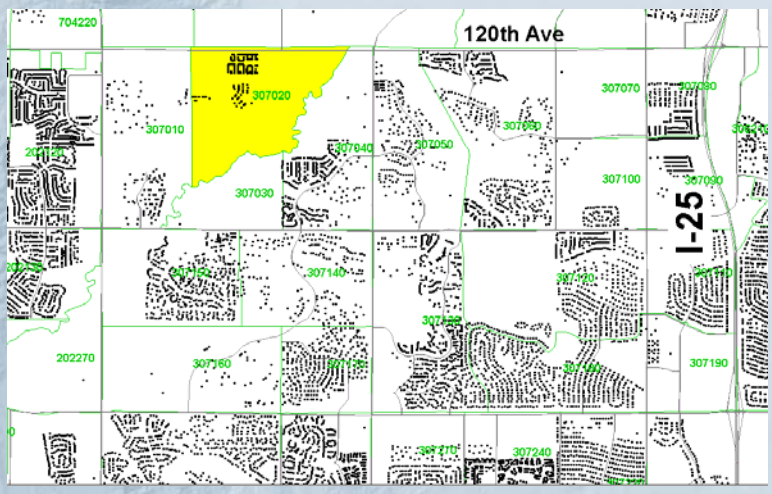
Stakeholder Buy-Off

- Vision Phase
- Technical and policy panels
- Periodic presentations to DRCOG Board, etc.
- The good news and bad news: if you sell it, they'll want it!
- Grand plans for "products in a stream"
  - Didn't really follow through
  - DRCOG bogged down for a long time building the "plumbing" and db.
    - Other responsibilities got in the way for staff.
  - Do this if you possibly can! Don't bail like we did!

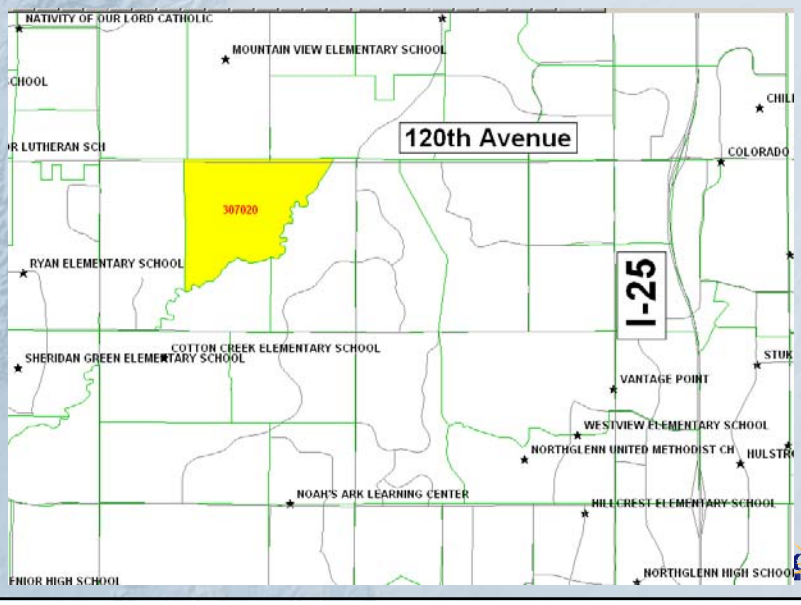
Basics of our design

- TransCAD for networks, skims, assignment and matrix repository
- C# for "plumbing" and logit solving components
  - Threaded and distributable
- MicroSoft SQL-Server for all non-matrix data
- Theoretical structure very similar to SACOG:
  - High level of geographic disaggregation
    - Especially useful for bike/ped
  - Very little explicit modeling of intra-household interactions
  - "Day travel pattern" model.

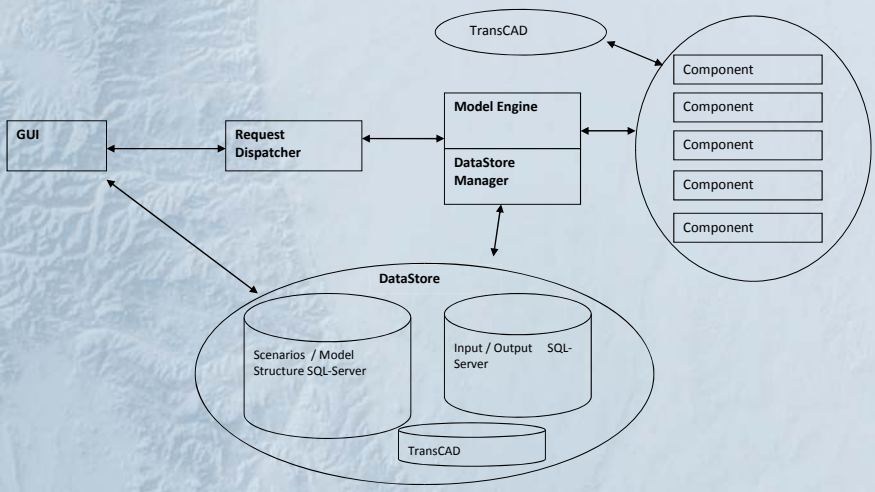
Point data



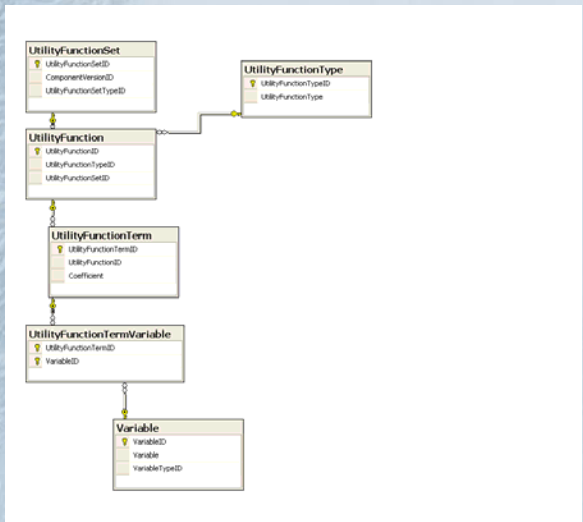
More point data



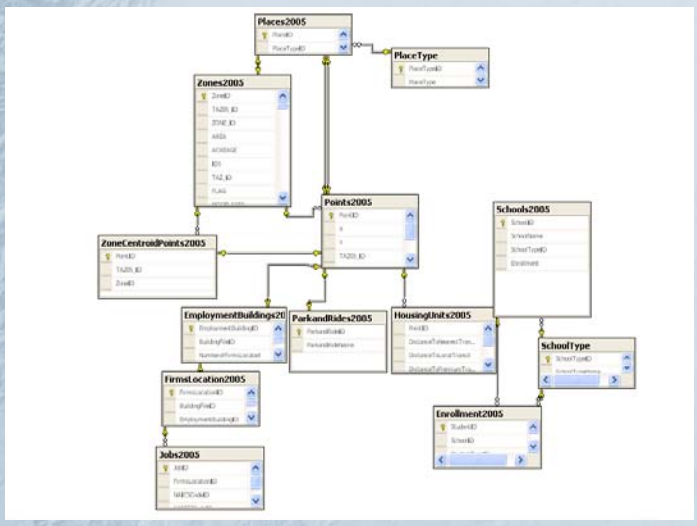
The Big Picture



Some database design

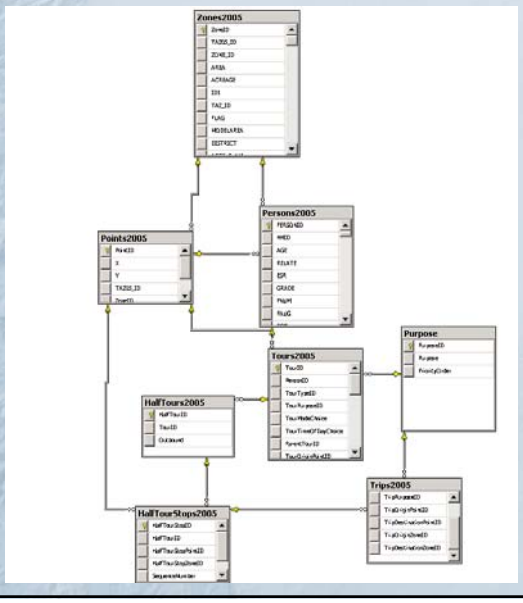


Some more database design



DRCOG

Yet more database design



DRCOG

Model Components

- **1. Population Synthesizer**
  - Network Skims
  - Aggregate Mode/Destination Choice Logsum Generator
  - Mode Choice Logsum Generator
- **2. Regular Workplace Location Choice**
- **3. Regular School Location Choice**
- **4. Auto Availability**
  - Intermediate Stop Logsum Generator
- **5. Daily Activity Pattern Choice**

Exact Number of Tours Choice
Work Tour Destination Type Choice Model
Work-Based Subtour Generation Choice

6. Tour Primary Destination Choice

7. Tour Main Mode Choice

8. Tour Time of Day Choice

- Intermediate Stop Generation Choice
- Intermediate Stop Location Choice

9. Trip Mode Choice

10. Trip Departure Time Choice

11. Assignment

DRCOG

A few details and thoughts

- Logsums
- Simulations
- Possible inconsistencies you can get
- Intra-hh interaction models may reduce or eliminate these
- Our attitude: model at level n so it is accurate at level n+1

DRCOG

PopSyn

- Borrowed from Atlanta
- Designed by John Bowman
- Coded by Wu Sun (now at SANDAG)
- Controls households at small geography
- Can make it control at regional level
- No person control

PopSyn output processor

- We build a set of points as model input
  - Based on our zone-based land use forecast
- Assign dwelling unit point to each household
- Assign job point to each job
- Calculate distance to transit from each household

Regular workplace location choice

- Nested logit
- Disaggregate logsums
- Size variables
- Very large choice set
- One of the slow ones

Regular school location choice

- Actually four models (different school levels)
- Also disaggregate logsums and large choice sets
- And so also rather slow
- But not as slow, because of fewer variables in the utility function
- A bit of intra-hh interaction here

Aggregate logsum generator

- Mode-destination choice logsum
- Represents aggregate accessibility of a place to all other places
- Simplified due to some variables not yet being known
- Also simplified to reduce computational burden

Daily Activity Pattern

- Probably the most complex of the models
- Simultaneously predicts 14 variables
- Person/hh variables
- Variables crossing person/hh characteristics with amount of activity in the day
- Variables concerning combinations of different activities during the day

Exact number of tours

- DAP says "none or some" tours or stops for each purpose
- This one calculates how many for the "some" cases

Work Tour Destination Choice

- For each work tour, does it go to the regular workplace
- Or somewhere else?

Work-based subtour generation

- By convention, all tours are assumed to start at home…
- Except tours that leave work and return there again
- Most common of these is "eat meal"

Tour time of day simulation

- Catch-22 when doing these models sequentially…
- Destination choice needs skims (varying by time of day)
- But time of day model runs AFTER destination choice
- So we run simulation at this point
- And the real tour time of day model becomes a conditional model

Tour primary destination choice

- Several models by purpose (I forget how many!)
- Similar in some ways to Regular Work Location Choice
- And slow!
- And very demanding of computer memory.

Tour main mode choice

- Paired with trip mode choice (below)
- Includes non-motorized modes…
- Which is a big reason we built a point-based model
- Primary mode for the tour

Tour time of day

- Limits on the eligibility of times that can be chosen…
- That conditional model comment again
- "Shift variables" (example of part time worker crossed with duration shift)

Intermediate Stop Generation

- Purpose constants
- Person variables
- Tour time of day variables
- "how busy is the day" variables

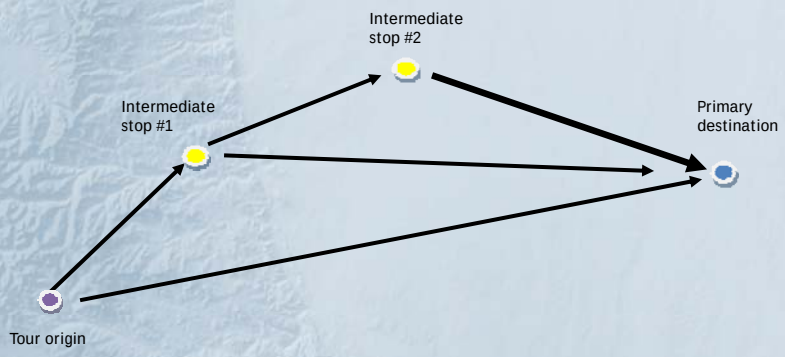
Trip time of day simulation

- Situation analogous to tour time of day
- Skims for intermediate stop destination need time of day to be known

Intermediate Stop Destination

- Detour generalized time variables
- Zonal development characteristics

Half Tour intermediate stop location example



Trip mode choice

- A lot of the usual
- Plus conditioning on the tour main mode choice

Intermediate Stop Departure Time Choice

- Constants and shift variables
- Trip purpose variables (this is one model)
- Person characteristic variables
- "how busy is the day" variables (size of remaining available windows)

Assignment

- Still using static equilibrium for highway (10 periods)
- Some merging with remaining trip-based elements
- Four transit periods
- OD transit assignment for drive to transit is really hard!

Speed balancing

- Several possibilities
- At present, based on "segments"
- Will be exploring various options.

Big efforts at the start

- Modelers and software people teaching each other
- Estimation data TOOK A LONG TIME!
  - Side trip to the land of transit path-building
    - Validating model versus survey data
    - Trying to solve the problem of transit paths with multiple transit modes
- Side trip to the land of UrbanSIM/Opus
- Building the "plumbing"
  - Lots of discussions about programming philosophy (and despite how it sounds, these are terribly important!)
- Building a prototype logit solving component

Team assembly and issues

- Use my programmer theorem from Chicago
  - Have a system, and stick to it!
- Lot of discussion over who makes the detailed design calls
- Use Sharepoint or similar!
- Multiple estimators can cause some issues
  - Nothing different from any other large project, multi-person team
- Consultant staff was pretty stable over a long period
- More partnerships can be very cool
  - But hard to reconcile too many competing desires

Humorous Mathematical Interlude

- Sabina's First Law of Software Development

$$\forall X, Y \in Developers \; \forall c, d \in Strategies$$
$$(X \neq Y \wedge approach(X, c) \wedge approach(Y, d)) \rightarrow c \neq d)$$

- Proof by inspection (if you've ever tried to do it)

In-house stresses and struggles

- We had a lot of really smart, motivated people
- And such people have strong opinions!
- Biggest challenge for the PM: coordinating those strong wills
- Bleeding edge on the software was a challenge:
  - Judgments were heavy on opinion, fact-starved!
- We sold it well, which then caused mgt pressure to finish!

In-house versus contractor

- Plumbing in-house:
  - We accidentally had a really strong IT developer
  - If you don't, outsource this!
- I don't know how you can learn your sw except by developing some of it!
  - Outsource basic logit solver, then you specialize it for some components?
- Definitely estimate some models
  - Use the coach model

Our favorite mistakes

- Use the same code to create skims that you'll use to run the model when it's done.
- Trim your ambitions regarding partnerships.
- Get estimation files right away (before the estimators forget where the final version is!)
- Start thinking about (and preparing!) estimation data really early.
  - Some of it depends on the model design
  - But lots doesn't (you can start now on this part!)
- But do think about data you don't have now, and how you might get it.
- Start to think now about level of geographic detail.

Key steps in software

- Learning to call TransCAD from C#
- TransCAD matrix read/write dll
- Getting a basic version of the "plumbing" working
- Various performance "wins"
- Especially figuring out basic threading (huge)

A closer look at some results

- All come from db queries stored in the db
- Location choice models are not doubly constrained
- We're in the early days of trying to explain these results!

Staff turnover

- Lost senior modeler just before the consultant started.
- Lost another senior modeler just one year into consultant project.
- Lost another senior modeler in the middle of calibration.
- Socio-econ group turned over twice during the project.

Current issues

- 64 bit version
- Hardware selection
- Documentation and document control

Thing we STILL can't do well

- Transportation facility price (need distributed vot.)
- Location/migration of households by type
  - Aging households
  - Gentrification
  - Lifecycles of neighborhoods
  - EJ

Contact information

- Erik Sabina
  - esabina@drcog.org
- Suzanne Childress
  - schildress@drcog.org