Integrated Models and Data Needs for Large-Scale Simulations: Accessibility, Time-Space Prisms, Choice Sets, and Related Issues

Konstadinos (Kostas) Goulias
Outline

• “Typical” emerging schema
• “Typical” emerging data collection
• Behavior in Space-Time
• Modeling the environment & the agent
• Integration needs
• Computational issues
Southern California Region

- Simulator of Activities, Greenhouse gas Emissions, Networks, and Travel (SimAGENT) in Southern California
- Six Counties (population, percent change 2000-2010)
  - Imperial (174,528, D%=22.6)
  - Ventura (823,318, D%=9.3)
  - San Bernardino (2,035,210, D%=19.1)
  - Riverside (2,189,641, D%=41.7)
  - Orange (3,010,232, D%=5.8)
  - Los Angeles (9,818,605, D%=3.1)
Typical Schema SimAGENT

Baseline Year (t=1)
- Synthetic Population
- Accessibility by Time-of-Day
- Long Term Choices
- Car Ownership and Type
- Activity and Travel Scheduling
- Routes & Assignment
- Energy Consumption & Emissions

Agent and Environment Evolution
- Population Evolution
- Urban Landscape Evolution
- Infrastructural Changes
- Scenario Databases
- Information Fusion
- Accessibility Computation

One Year Later (t=t+1)
- Synthetic Population
- Accessibility by Time-of-Day
- Long Term Choices
- Car Ownership and Type
- Activity and Travel Scheduling
- Routes & Assignment
- Energy Consumption & Emissions
Data Needs

• Core Behavior and Household Characteristics
  – Other aspects – policy dependent (cars and costs, long term choices and lifestyle, attitudes)

• Other agents (firms, institutions, plans, and so forth)

• Landscape/Environment/Context
  – Activity locations
  – Homes/Jobs/Schools
  – Availability over time
Figure 4 The Data Collection Overall Scheme

- Core Survey (Household, Person, and Base Diary)
- In-Depth Mode Survey with Active Living Questions
- Expenditures and Budgeting Survey
- Residence, Workplace, and School Location
- Long Distance Component
- Toll Willingness to Pay
- GPS & GPS OBD (verification, special days, emissions)
- Panel of Households and Persons & 2-day Diary
- One Week Activity Diary
- In-Depth Car Ownership and Use
- Location Choice and Activity Satisfaction
- Retrospective and Prospective Location Choices
The Many Ds in California Land Use & Transp. Policy
(change in each indicator)

• Density = (Population + Employment)/square mile
• Diversity =
• Design = a*street network density + b*sidewalk completeness + c*route directness
• Destinations = from an origin zone i Sum over j of (attractions*impedance)
• Distance to transit

Local = depends on application; street network density = length of street in miles/area of neighborhood in square miles, sidewalk completeness = total sidewalk centerline distance/total street centerline distance, route directness = average airline distance to center/average roadway distance to center

INDEX 4D uses a=0.0195, b=1.18, and c=3.63
Goals

• Develop the base background data and information to develop the Ds
• Create indices at a fine spatial level (display activity “deserts”)
• Add the temporal (time of day) dimension to the indices
• Create a method that can be used in forecasting/scenario building
• Focus on Destinations intended as opportunities and not as a gravity model or other functions that require behavioral assumptions
Objectives

- Map an entire region (even a state) to show access to activity opportunities
- Use 15 different types of “industries” (retail, education, health, etc)
- Account for different speeds on network by time of day
- Account for different opening-closing hours by county and industry type
- Use largely available data with any needed rectifications using secondary sources of information
- Develop a multi-scale process (TAZ/tract, blockgroups, blocks, land parcels)
- Compute these for different years (2000-2003, 2008, 2010, 2025, 2035) –future years are for land use scenarios
- Test these indicators in behavioral model specifications
- Use these indicators to study spatial equity for sociodemographic groups
Ingredients

• A geocoded database that enumerates all opportunities classified by the type of activity – first order of approximation can be number of business establishments classified based on NAICS-SIC

• Employment by type of industry to use as verification but also as a direct indicator of attractiveness

• Other data to fuse, merge, verify, validate built environment estimates

• A detailed network with all models – start with highways and add other travel modes

• Estimates of travel speeds on each link

• An algorithm to build the indicators
Data Used (in Los Angeles+)

- CTPP vintage 2000 (considered as a baseline)
- Dunn & Bradstreet post processed to provide summaries by block group (via a joint project with LANL & UCOP)
- US Census Block, block groups, TAZs, and Tracts
- SCAG network (highway and transit)
- Teleatlas California 2000
- InfoUSA (later vintage) just for comparison
- SCAG Travel Survey Post-Census Commute Arrival to Work and Departure from Work

- All this is usually available for MPOs
Conceptually we Build Buffers

Off Peak Network-based Access to an opportunity area (potential path area)

Peak Period Network-based Access to an opportunity area
Tracking density of opportunity by time of day – Just based on the information about arrival and departure from work of retail employees
Time of day profile of available opportunity

- Arrival and departure time of workers in travel survey (for each county and each industry type)
- Retail workers at work in LA county
2am
3am
5am
7am
10am
12 midnight
Min and Max in AM Peak

Los Angeles County - Finance and Insurance

Percent of Workers Available

Percent used to compute Min

Percent used to compute Max
Riverside opportunity map
Max in 7PM to 6AM (Night Time) period of retail within 10 minutes for each block
Superimposing Transit Routes
And the Land Parcels!
Compare retail accessibility between two different urban forms/street design.
Accessibility and Policy

- Land-use strategies aiming at changing business location -> impact the opportunities available and the accessibility of zones for activity purposes.

- In modeling we need to represent the opportunities available as a measure of attraction.

- But we may not have all this information available -> create a framework, test its feasibility, and then derive indicators at different levels – depending on data availability.

- Enhance modeling framework by computing different accessibility measures and accommodating these in the re-estimation.
Integration at Multiple Levels

• Land Use-Activity Opportunities
  – Spatial shifts/change in economic activity → spatial shifts in activity opportunities
  – Spatial shifts/change in employment → spatial shifts in work, residence, and school location choices
  – Spatial shifts in economic opportunity (jobs) → immigration/out-migration
  – Demographic evolution → location/relocation patterns, labor market changes, activity location preferences change and so on and on....
  – WE NEED A TIME SEQUENCED SPATIO-SOCIAL EVOLUTIONARY SYSTEM!
  – Verify all this with “clean” longitudinal data --> we are working with some noisy data
Feasibility of Using Time-Space Prism to Represent Available Opportunities and Choice Sets for Destination Choice Models

In the previous slides we reviewed accessibility indicators based on:

- **Spatial units** = 203,621 blocks
- **Highway network and travel time**
  - almost every local road of the network with an estimate of travel time between blocks
- **Temporal units**
  - Estimated speeds based on 4 time periods (AMpeak, PMpeak, Midday, and Night time) for travel time
  - 24 time periods for opportunity availability (based on opening and closing time)
  - Activity schedules by minute from an ABM (not reviewed here but in other seminars)
Components considered

- **Available opportunities by time of day**
  - 24 time periods and 15 industry types
  - a) Agriculture, forestry, fishing and hunting and mining; b) Construction; c) Manufacturing; d) Wholesale trade; e) Retail trade; f) Transportation and warehousing and utilities; g) Information; h) Finance, insurance, real estate and rental and leasing; i) Professional, scientific, management, administrative, and waste management services; j) Educational; k) Health; l) Arts, entertainment, recreation, accommodation and food services; m) Other services (except public administration); o) Public administration; p) Armed forces

- **Travel time by time of day**
  - 4 time periods
  - AM peak (6am-9am), PM peak (3pm-7pm), off-peak (9am-3pm), and night time (7pm-6am)

- **Individual differences in activity scheduling**
Data Used

- **Household Travel Survey for SCAG** (collected from spring 2001 to spring 2002) containing 16,939 households and 40,376 individuals.
- **Detailed roadway network including time dependent travel speed/travel time**
- **Employment data of 15 industry types** that combine the US Census Transportation Planning Package (CTPP) of year 2001 and Dun & Bradstreet (D&B)
- **Geographic information of the census blocks and the block groups** in SCAG region (available from the Census website, [http://www2.census.gov/cgi-bin/shapefiles2009/state-files?state=06](http://www2.census.gov/cgi-bin/shapefiles2009/state-files?state=06)) (for more details see Chen et al, 2011)
- **InfoUSA** that is a database enumerating all the business establishments in the region
- **Land parcel database** to locate and enumerate major shopping centers
Time-space prism

Hägerstrand (1970)
Activity scheduling
Computation of Potential Path Area

Home (block i) to Work (block j) via block k
<table>
<thead>
<tr>
<th>Origin</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>1.3</td>
<td>3.7</td>
<td>10.9</td>
<td>13.0</td>
<td>20.1</td>
<td>19.8</td>
<td>19.8</td>
<td>22.4</td>
<td>22.6</td>
<td>19.6</td>
<td>21.1</td>
</tr>
<tr>
<td>2</td>
<td>1.3</td>
<td>0.0</td>
<td>2.4</td>
<td>9.6</td>
<td>11.6</td>
<td>21.4</td>
<td>21.1</td>
<td>21.1</td>
<td>23.7</td>
<td>23.9</td>
<td>20.9</td>
<td>22.4</td>
</tr>
<tr>
<td>3</td>
<td>3.7</td>
<td>2.4</td>
<td>0.0</td>
<td>7.3</td>
<td>9.3</td>
<td>23.8</td>
<td>23.5</td>
<td>23.5</td>
<td>26.1</td>
<td>26.3</td>
<td>23.2</td>
<td>24.8</td>
</tr>
<tr>
<td>4</td>
<td>10.9</td>
<td>9.6</td>
<td>7.3</td>
<td>0.0</td>
<td>2.0</td>
<td>31.0</td>
<td>30.7</td>
<td>30.8</td>
<td>33.4</td>
<td>33.5</td>
<td>30.5</td>
<td>32.0</td>
</tr>
<tr>
<td>5</td>
<td>13.0</td>
<td>11.6</td>
<td>9.3</td>
<td>2.0</td>
<td>0.0</td>
<td>33.0</td>
<td>32.7</td>
<td>32.8</td>
<td>35.4</td>
<td>35.5</td>
<td>32.5</td>
<td>34.0</td>
</tr>
<tr>
<td>6</td>
<td>19.1</td>
<td>20.4</td>
<td>22.7</td>
<td>30.0</td>
<td>32.0</td>
<td>0.0</td>
<td>8.7</td>
<td>8.8</td>
<td>11.4</td>
<td>11.5</td>
<td>8.5</td>
<td>10.0</td>
</tr>
<tr>
<td>7</td>
<td>22.4</td>
<td>23.7</td>
<td>26.1</td>
<td>33.3</td>
<td>35.4</td>
<td>3.3</td>
<td>0.0</td>
<td>3.1</td>
<td>2.6</td>
<td>2.8</td>
<td>11.7</td>
<td>4.3</td>
</tr>
<tr>
<td>8</td>
<td>19.6</td>
<td>20.9</td>
<td>23.3</td>
<td>30.5</td>
<td>32.6</td>
<td>0.5</td>
<td>9.3</td>
<td>0.0</td>
<td>11.9</td>
<td>12.1</td>
<td>9.0</td>
<td>10.6</td>
</tr>
<tr>
<td>9</td>
<td>19.8</td>
<td>21.1</td>
<td>23.4</td>
<td>30.7</td>
<td>32.7</td>
<td>0.7</td>
<td>0.5</td>
<td>0.4</td>
<td>0.0</td>
<td>0.2</td>
<td>9.1</td>
<td>1.7</td>
</tr>
<tr>
<td>10</td>
<td>19.6</td>
<td>20.9</td>
<td>23.3</td>
<td>30.5</td>
<td>32.5</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
<td>2.9</td>
<td>0.0</td>
<td>8.9</td>
<td>1.5</td>
</tr>
<tr>
<td>11</td>
<td>19.8</td>
<td>21.1</td>
<td>23.5</td>
<td>30.8</td>
<td>32.8</td>
<td>0.8</td>
<td>0.4</td>
<td>0.5</td>
<td>3.1</td>
<td>3.3</td>
<td>0.0</td>
<td>1.7</td>
</tr>
<tr>
<td>12</td>
<td>21.0</td>
<td>22.3</td>
<td>24.7</td>
<td>32.0</td>
<td>34.0</td>
<td>2.0</td>
<td>1.7</td>
<td>1.6</td>
<td>4.4</td>
<td>4.6</td>
<td>10.4</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Computation of Potential Path Area

Home (block \(i\))

Work (block \(j\))

Block \(k\)

\[ \text{Travel time from } i \text{ to any } k \]
\[ \text{Travel time from any } k \text{ to } j \]
\[ \text{Travel time from } i \text{ to } j \text{ through any } k \]

Blocks that are accessible within the time budget: PPA
Potential path areas for different time windows (AMpeak)
Available opportunities within PPA for different activity types (8am to 9am)

- **Purchasing activity/Retail Banking/Finance**
  - Accessible businesses: 40,593
  - Open: 32,377
  - Employees: 143,452
  - Accessible businesses: 26,768
  - Open: 21,334
  - Employees: 90,697
  - Accessible businesses: 8,528
  - Open: 7,078
  - Employees: 97,416
  - Accessible businesses: 13,110
  - Open: 10,881
  - Employees: 142,299
Different purchasing activity types

Grocery stores
- Accessible: 288
- Open: 230

Regional shopping centers
- Accessible: 35
- Open: 28

Accessible: 189
Open: 151

Accessible: 25
Open: 20

PPA for 75 min time budget
PPA for 75 min time budget with 20min minimum activity duration
PPA by time of day

- AM peak
- Midday
- PM peak
- Night time
Available opportunities by time of day

AM peak

Midday

PM peak

Night time

- PPA for 75 min time budget
- PPA for 75 min time budget with 20min minimum activity duration
If have flexibility in arrival and departure time. Which value of accessibility matters to behavior (home, work, route or all)?

Density of workers present at work

Density of workers
- Low
- High

PPA for 75 min time budget
PPA for 75 min time budget with 20min minimum activity duration
Nature of Place
Changes with Time of Day and
With Whom we Interact
Beginning of activity
- 22:00-23:59
- 12:00-13:59
- 00:00-1:59

Solo Activities

With Family

With Friends
Some of the findings of this section

- Spatial distribution of alternatives has impacts on behavior
- Temporal dimensions of behavior also have impacts on
  - the spatial extent of choice sets and
  - the nature of alternatives that are distributed over space
- Flexibility (when, where, and how) makes some alternatives more attractive and some other less attractive
- These have to be accounted for in simulation models to achieve behavioral realism and to enhance model’s ability to predict the impact of policies related to time (i.e., staggered arrival at work, flexible work hours)
- Further development
  - Intra-household interaction and life cycle stages (Yoon and Goulias, 2010) and
  - Subjective perception of space (mental maps) and attitudes
Integration Needs

- We need estimates of travel time on all links of the entire network – traffic assignment should give us all this information by some time interval and we do not know the right interval to use.
- Travel times need to be multi-modal (walk + bus) of any combination analyzed for policy – not a trivial task but feasible (see appendix)
- Availability of opportunities in this example is point-based (individual establishment) → model of open-closed hours of establishments?
- Knowledge (spatial cognition) and preferences/attitudes (sense of place) are not included and we know are important in delimiting spatial regions of considered opportunities → integrate with cognitive filters?
Computation

• Initial versions take days to run on 40 core servers
• Second iteration of newer versions take hours to run
• BUT: slow improvement in handling input-output from different components-still need human intervention to check databases and verify interfaces work properly
• ONE core application designed to integrate modules is needed now!
Acknowledgement

GeoTrans lab.
University of California Transportation Center
United States Department of Transportation Eisenhower Fellowship program
University of California Office of the President UC Lab Fees Program
University of California Office of the President Multicampus Research Program Initiative
Sustainable Transportation
Southern California Association of Government

Thank You to The Contributors:

Dr. Yali Chen, Dr. Seo Youn Yoon, Dr. Ting Lei,
Kathleen Deutsch, and Srinath Ravulaparthy
Base map of opportunities

FROM BLOCK GROUPS TO BLOCKS & ADDING NETWORK DETAIL
SCAG network from a four-step application (year 2003)
Detailed network (adding local...
Block Level Centroid Connectors
Allocation from Block groups to Blocks

- Function of land area, population, amount of: freeways, arterials, collectors, ramps (negative and positive influence depends on industry type).

- Tested many different regression models and decided on Poisson regression (count data regression models)

- Kept the total equal to the block group for each block group.
Using these regression models (one for each industry type), the block group employment for each industry type is distributed into each block within a block group as shown below:

a. Calculate the estimated block employment ($\hat{E}_i$) for block $i$ by applying the block characteristics and the estimated regression model.

b. Calculate the percentage of employment ($P_i$) for each block $i$ within each block group $j$.

$$P_i = \frac{\hat{E}_i}{\sum_{i \in B_j} \hat{E}_i},$$

where $B_j$ is the set of all blocks within block group $j$.

c. Calculate the block employment ($E_i$) by multiplying the percentage with the block group employment

$$E_i = P_i \times \text{(employment in block group } j \text{ that block } i \text{ belongs to)}.$$

The same process is repeated for $j = 10,631$ block groups and 15 industry types.
Agriculture Density

Block group level (observed)  Block level (predicted from model)

Blocks in the Ocean – special treatment
Retail Density

Block group level (observed)

Block level (predicted allocation model)
Retail Density

Block group level (observed)  Block level (predicted from model)
We split this in two separate sets of indicators
Education & Health Density

Block group level (observed)  Block level (predicted from model)
Check if reasonable
Issues Found and Resolved

- Block groups with employees but no businesses
- Block groups with businesses but no employees
- Disagreements between D&B number of persons by industry and CTPP
- Differences in NAICS-SIC classifications
- CTPP reports education and health as one industry (split in two)

AVAILABILITY OF OPPORTUNITIES BY TIME OF DAY
Time of Day Components

• Travel time on the network -> use travel speed from calibrated four step model that has four wide time periods (AM Peak - 6AM to 9AM, Midday - 9AM to 3PM, PM Peak - 3PM to 7PM, Night - 7PM to 6AM)

• Opening-closing times for businesses -> use arrival and departure time of workers in post Census survey (16k hhs)

• To account for the “wide” periods of the network times we use MIN and MAX for the number of employees by industry that can be reached within a temporal buffer of 10, 20, and 50 minutes

• The buffers 10 vs 50 resemble local vs regional accessibility

• Shortest path computed using aprox. 200,000 by 200,000 origin destination matrix using TRANSCAD.
Employee Classification

1. Wholesale Trade
2. Retail Trade
3. Administrative and Support Services
4. Educational Services
5. Healthcare and Social Assistance
6. Arts, Entertainment and Recreation
7. Accommodations and Food Services
8. Public Administration
9. Agriculture/Forestry/Fishing
10. Construction
11. Manufacturing – Durable and Nondurable goods
12. Transportation and Warehousing
13. Information
14. Finance or Insurance
15. Professional, Scientific and Technical Services
Using a survey of workers for each respondent $q$ that works, we have the arrival time at work (say $t_{qa}$) and departure time from work (say $t_{qa}$).

Define a dummy variable $\delta_{qgh}$ that takes a value of one if individual $q$ works in industry type $g$ and county $h$, and zero otherwise.

Consider a specific hour of the day $k$, and let $T_s$ be the start time of this hour $k$, and let $T_e$ be the end time of this hour $k$.

Define another dummy variable $\alpha_{qk}$ that takes a value 1 if $t_{qz} \geq T_s$ and $t_{qe} \leq T_e$, and zero otherwise.

Intuitively, $\alpha_{qk}$ equals one if the employee $q$ is at the work place at the hour $k$ and zero otherwise.

From the travel survey, we also have a weight $W_q$ for person $q$, which expands the individual so that the expanded sample across all surveyed individuals is representative of the population.

An estimate of the number of reachable employees in industry type $g$ at county $h$ at time $k$ ($A_{ghk}$) is as follows:

$$A_{ghk} = \sum_{q} \delta_{qgh} \alpha_{qk} W_q$$
The total number of reachable workers in industry $g$ at county $h$ at time $k$ can be expressed as a function of the percentage of reachable workers in industry $g$ and county $h$:

$$P_{ghk} = \frac{A_{qhk}}{\sum_q \delta_{qgh} W_q}.$$ 

The next slide shows examples of the reachable percent of workers in Los Angeles County by time of day.
Identification of Activity Opportunity Clusters

Based on: Ravulaparthy S., P. Dalal, Y. Chen, and K. G. Goulias (2011)
Step 2: G* Analysis

• Spatial clusters analysis
  – Spatial dependence
  – Spatial heterogeneity

• Getis-Ord $G^*(d)$ statistic is a local measure of spatial dependence
  – Measures spatial association between any given point in space to all other neighboring points
  – Defined by a distance of $d$
  – Neighborhood blocks sharing a boundary

• Getis-Ord $G^*$ statistic calculates the z-score
  – Statistical significance of clustering

• z-score significance
  – Higher or lower values indicates high intensity of clustering
  – Near zero indicates no clustering in study area

Jacquez (2008), Getis and Ord (1992) and www.esri.com
Step 2: G* Analysis

- Data input for Getis-Ord G* statistic
  - AM peak (6AM to 9AM) block level accessibility measures within 10 minutes for finance, retail trade, arts/entertainment, healthcare & education
  - Average land values (price per unit area)
- Output
  - Computed z-scores for the six spatial indicators at block level
  - High opportunities and high land values are positive z-scores
  - Low opportunities and lower land values are negative z-scores
- G* clusters
  - Red indicates high value and clusters
  - Blue indicates low value and clusters
  - Yellow indicates lack of clustering
G* Clusters

Arts/Entertainment

Education
G* Clusters

Finance

Healthcare
G* Clusters

Retail

Land Value
Space-time Prism of Travel by Public Transit

Data Source: LACMTA master trip database
Ting Lei, Jan 2012
Maximum number of reachable retail employees for a 20-minute buffer by transit by time of day in Central Los Angeles.

Maximum number of reachable education employees for a 20-minute buffer by transit by time of day in Central Los Angeles.

Maximum number of reachable retail employees for a 20-minute buffer by time of day in Central Los Angeles by Private Car

Maximum number of reachable education employees for a 20-minute buffer by time of day in Central Los Angeles by Private Car
• Moving to finer resolutions (more TAZs or even points) increases ability to represent built environment but creates the need to maintain databases over time
• There are some interesting issues about spatial allocation – for now we used count data regression models
• Assign synthetic households to blocks but also assign parcels/housing units to synthetic households
• Start the path toward a demographic microsimulator for at least 25 years and identify methods for fine spatial resolution
• Interface with land use model can offer added information but the space resolution will always be an issue to consider
• Interface with the more recent “sustainability tool(s)” that use GIS maps of parcels open a new gateway to possibilities
Base map of opportunities

FROM BLOCK GROUPS TO BLOCKS & ADDING NETWORK DETAIL
SCAG network from a four-step application (year 2003)
Detailed network (adding local roadways)
Block Level Centroid Connectors
Allocation from Block groups to Blocks

- Function of land area, population, amount of: freeways, arterials, collectors, ramps (negative and positive influence depends on industry type).

- Tested many different regression models and decided on Poisson regression (count data regression models)

- Kept the total equal to the block group for each block group.
Using these regression models (one for each industry type), the block group employment for each industry type is distributed into each block within a block group as shown below:

a. Calculate the estimated block employment ($\hat{E}_i$) for block $i$ by applying the block characteristics and the estimated regression model

b. Calculate the percentage of employment ($P_i$) for each block $i$ within each block group $j$.

$$P_i = \frac{\hat{E}_i}{\sum_{i \in B_j} \hat{E}_i},$$

where $B_j$ is the set of all blocks within block group $j$.

c. Calculate the block employment ($E_i$) by multiplying the percentage with the block group employment

$$E_i = P_i \times (\text{employment in block group } j \text{ that block } i \text{ belongs to}).$$

The same process is repeated for $j = 10,631$ block groups and 15 industry types.
Agriculture Density

Block group level (observed)

Block level (predicted from model)

Blocks in the Ocean – special treatment
Retail Density

Block group level (observed)

Block level (predicted allocation model)
Retail Density

Block group level (observed)

Block level (predicted from model)
We split this in two separate sets of indicators
Education & Health Density

Block group level (observed)

Block level (predicted from model)
Check if reasonable
Issues Found and Resolved

- Block groups with employees but no businesses
- Block groups with businesses but no employees
- Disagreements between D&B number of persons by industry and CTPP
- Differences in NAICS-SIC classifications
- CTPP reports education and health as one industry (split in two)

AVAILABILITY OF OPPORTUNITIES BY TIME OF DAY
Time of Day Components

- Travel time on the network -> use travel speed from calibrated four step model that has four wide time periods (AM Peak - 6AM to 9AM, Midday - 9AM to 3PM, PM Peak - 3PM to 7PM, Night - 7PM to 6AM)
- Opening-closing times for businesses -> use arrival and departure time of workers in post Census survey (16k hhs)
- To account for the “wide” periods of the network times we use MIN and MAX for the number of employees by industry that can be reached within a temporal buffer of 10, 20, and 50 minutes
- The buffers 10 vs 50 resemble local vs regional accessibility
- Shortest path computed using aprox. 200,000 by 200,000 origin destination matrix using TRANSCAD.
Employee Classification

1. Wholesale Trade
2. Retail Trade
3. Administrative and Support Services
4. Educational Services
5. Healthcare and Social Assistance
6. Arts, Entertainment and Recreation
7. Accommodations and Food Services
8. Public Administration
9. Agriculture/Forestry/Fishing
10. Construction
11. Manufacturing – Durable and Nondurable goods
12. Transportation and Warehousing
13. Information
14. Finance or Insurance
15. Professional, Scientific and Technical Services
Using a survey of workers for each respondent $q$ that works, we have the arrival time at work (say $t_{qa}$) and departure time from work (say $t_{qa}$).

Define a dummy variable $\delta_{qgh}$ that takes a value of one if individual $q$ works in industry type $g$ and county $h$, and zero otherwise.

Consider a specific hour of the day $k$, and let $T_s$ be the start time of this hour $k$, and let $T_e$ be the end time of this hour $k$.

Define another dummy variable $\alpha_{qk}$ that takes a value 1 if $t_{qa} \geq T_s$ and $t_{qa} \leq T_e$, and zero otherwise.

Intuitively, $\alpha_{qk}$ equals one if the employee $q$ is at the work place at the hour $k$ and zero otherwise.

From the travel survey, we also have a weight $W_q$ for person $q$, which expands the individual so that the expanded sample across all surveyed individuals is representative of the population.

An estimate of the number of reachable employees in industry type $g$ at county $h$ at time $k$ ($A_{ghk}$) is as follows:

$$A_{ghk} = \sum_{q} \delta_{qgh} \alpha_{qk} W_q$$
The total number of reachable workers in industry $g$ at county $h$ at time $k$ can be expressed as a function of the percentage of reachable workers in industry $g$ and county $h$:

$$P_{ghk} = \frac{A_{qhk}}{\sum_q \delta_{qgh} W_q}.$$ 

The next slide shows examples of the reachable percent of workers in Los Angeles County by time of day.
Identification of Activity Opportunity Clusters

Step 2: G* Analysis

- Spatial clusters analysis
  - Spatial dependence
  - Spatial heterogeneity
- Getis-Ord $G^*(d)$ statistic is a local measure of spatial dependence
  - Measures spatial association between any given point in space to all other neighboring points
  - Defined by a distance of $d$
  - Neighborhood blocks sharing a boundary
- Getis-Ord $G^*$ statistic calculates the z-score
  - Statistical significance of clustering
- z-score significance
  - Higher or lower values indicates high intensity of clustering
  - Near zero indicates no clustering in study area

Jacquez (2008), Getis and Ord (1992) and www.esri.com
Step 2: G* Analysis

- Data input for Getis-Ord G* statistic
  - AM peak (6AM to 9AM) block level accessibility measures within 10 minutes for finance, retail trade, arts/entertainment, healthcare & education
  - Average land values (price per unit area)

- Output
  - Computed z-scores for the six spatial indicators at block level
  - High opportunities and high land values are positive z-scores
  - Low opportunities and lower land values are negative z-scores

- G* clusters
  - Red indicates high value and clusters
  - Blue indicates low value and clusters
  - Yellow indicates lack of clustering
G* Clusters

Arts/Entertainment

Education
G* Clusters

Finance

Healthcare
G* Clusters

Retail

Land Value
Maximum number of reachable retail employees for a 20-minute buffer by transit by time of day in Central Los Angeles.

Maximum number of reachable education employees for a 20-minute buffer by transit by time of day in Central Los Angeles.

Maximum number of reachable retail employees for a 20-minute buffer by time of day in Central Los Angeles by Private Car

Maximum number of reachable education employees for a 20-minute buffer by time of day in Central Los Angeles by Private Car
• Moving to finer resolutions (more TAZs or even points) increases ability to represent built environment but creates the need to maintain databases over time
• There are some interesting issues about spatial allocation – for now we used count data regression models
• Assign synthetic households to blocks but also assign parcels/housing units to synthetic households
• Start the path toward a demographic microsimulator for at least 25 years and identify methods for fine spatial resolution
• Interface with land use model can offer added information but the space resolution will always be an issue to consider
• Interface with the more recent “sustainability tool(s)” that use GIS maps of parcels open a new gateway to possibilities
Thank You & Questions?