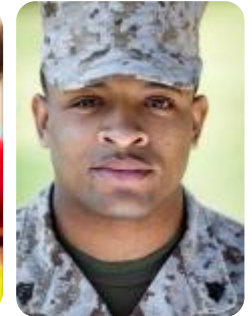


DRAFT
5/31/2016



The Value of Self-Reported Frequently Visited Addresses in GPS Assisted Travel Surveys

Timothy Michalowski
Dara Seidl

June 28, 2016

Agenda



- **Travel Surveys**
 - Location Data Collection Methods
- **Research Focus**
 - Self-reported addresses
 - vs.
 - Recorded GPS locations
- **Los Angeles County Simulation**
 - GIS Analysis of ~2 million parcels
- **Findings**



Travel Surveys



Travel Surveys



- **Metropolitan Planning Organizations (MPOs)**
Regional Transportation Planning Organizations (RTPOs)
 - Regional planning organizations
 - Urbanized areas 50,000+ population (409 MPOs in USA)
 - Manage regional funding for transportation projects
- **Exact travel location data is increasingly essential**
 - Shift to GPS offers ~3 meter accuracy
 - New activity based travel models
 - Continuous need for higher precision

Travel Surveys

Self-Report Methods



Travel: How did you get to Location 1?

1. What type(s) of transportation did you use to go to Location 11?

1 st		2 nd (if needed)		3 rd (if needed)	
1	Car, van, truck	4	Public Bus	7	Amtrak
2	Walk	5	Light Rail (Hiawatha)	8	Bicycle
3	School Bus	6	Commuter Rail (Northstar)	9	Motorcycle/Moped
				10	Taxi/Shuttle
				11	Dial-A-Ride
				12	Other (specify) _____

2. If you used a bus/train for this trip, did you use a pass? Yes No --> How much did you pay? _____

3. If you used car/van/truck or motorcycle/moped for this trip . . .

A. Were you the . . .? Driver Passenger

B. Including yourself, how many people were in the vehicle? 1 2 3 4+

Including yourself, how many are household members? 1 2 3 4+

Which household members were with you?
_____, _____, _____, _____

C. Was this vehicle from your household? Yes No

D. Did you pay a toll? Yes No

E. How much, in total, did you personally pay for parking? Nothing
\$ ____ . ____ Was the rate...? Hourly Daily Monthly Other

PAPER
TRAVEL DIARY

WEB
TRAVEL DIARY

PHONE CATI
TRAVEL DIARY

Travel information for every household member, every trip:

Start Location	Start Time	Travel Mode	Activities at Locations
End Location	End Time	Travel Segments	Behavioral Data

Travel locations as text addresses → Geocoded to coordinates

Travel Surveys

GPS Methods



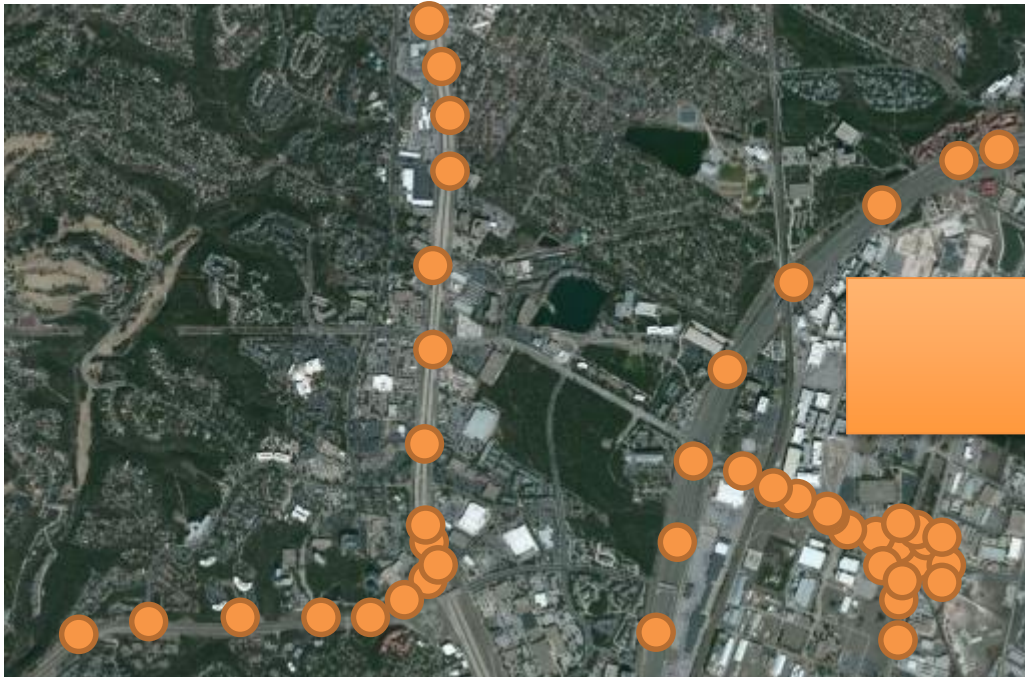
- GPS Loggers or GPS Smartphone Apps
- Passively record location via GPS every 1-30 seconds
- ~3 meter Geographic accuracy
- Participants instructed to carry GPS logger / GPS phone app everywhere
- Travel Survey 1 to 3 days



Travel Surveys GPS Methods



1 second of travel = 1 GPS point
1 person, 1 day of travel = ~10,000 GPS points
1,000 persons = ~1 million GPS points

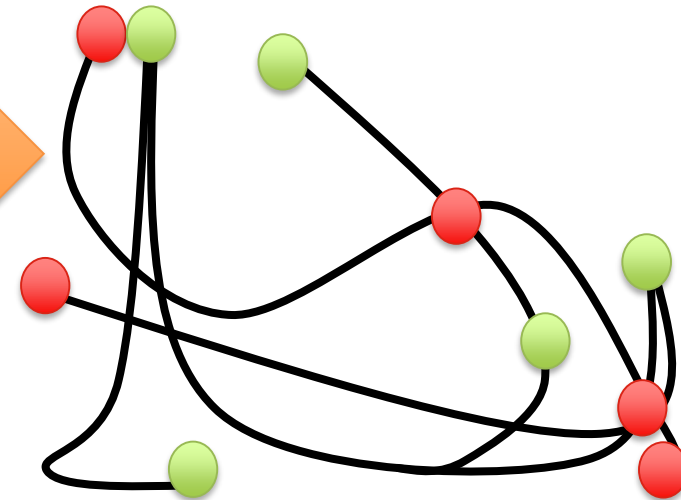


~1 million data points at 3 meter accuracy

GIS Data Processing

ORIGINS → **DESTINATIONS**
POINTS for each trip

ROUTE
LINES for each trip



Research Focus



Research Focus



- In standard recruit survey = **“Frequently visited locations”**
- Where is your... Home? Work? School? Shopping? Other Frequented Locations?
- Locations collected as Text Addresses
Addresses → Data Entry → Geocoded to Lat/Lon Coordinates
- **How do self-reported “frequently visited addresses” compare to GPS travel survey points?**



Research Focus



Annie Provided Frequently

Visited Addresses:

Home = A

Work = B

School = C

Shopping = D

Other = E



Completed 1 day Travel Survey
with GPS data =

Trip 1 = B

Trip 2 = C

Trip 3 = D

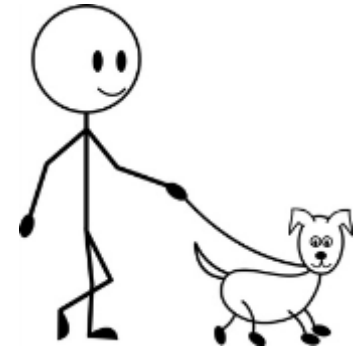
Trip 4 = E

Research Focus



Bruce Provided Frequently Visited Addresses:

Home = A
Work = B
School = C
Shopping = D
Other = E



Completed 1 day Travel Survey with GPS data =

Trip 1 = O

Trip 2 = E

Trip 3 = J

Trip 4 = Q

Research Focus



Annie Provided Frequently Visited Addresses:

Home = A
Work = B
School = C
Shopping = D
Other = E



Completed 1 day Level Survey with GPS data =

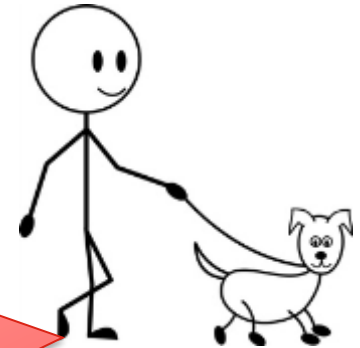
Trip 1 = B
Trip 2 = C
Trip 3 = D
Trip 4 = E



*Useful? or
Redundant
Survey Data?*

Bruce Provided Frequently Visited Addresses:

Home = A
Work = B
School = C
Shopping = D
Other = E



Completed 1 day Level Survey with GPS data =

Trip 1 = O
Trip 2 = E
Trip 3 = J
Trip 4 = Q



*Useful? or
Irrelevant
Survey Data?*

Research Question



How do self-reported “frequently visited addresses” compare to GPS travel points?

To answer:

- Need a large geographic study area...
- Diverse housing / commercial / Institutional Land Uses...
- GPS travel points / Survey data...
- Yet preserve participant locational privacy...

Solution: Simulate a GPS travel survey in LA County!

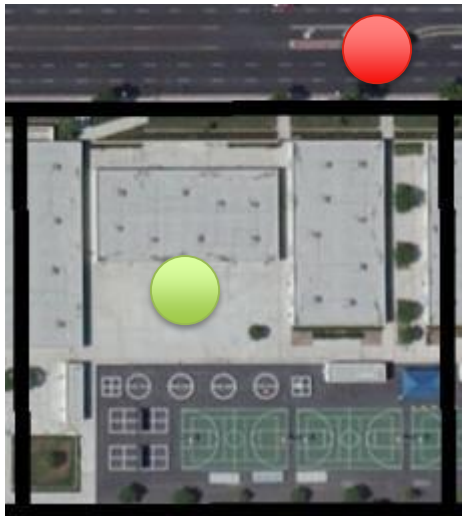
Los Angeles County Simulation



Los Angeles County Simulation



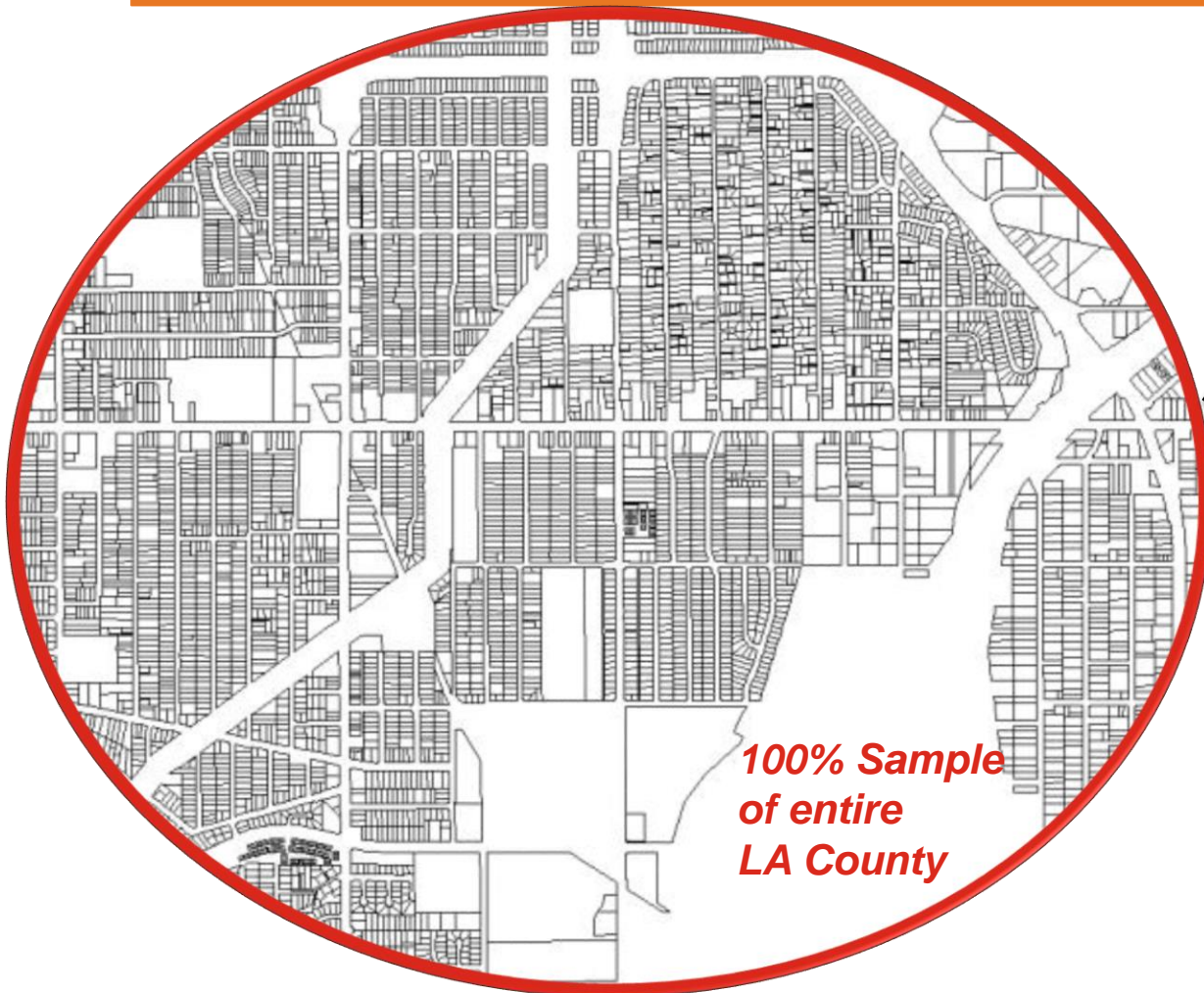
Data needed = Self-Report Addresses & GPS data



Self-reported address of school
1650 Echo Park Ave, Los Angeles, CA 90026
Geocoded = 34.082729, -118.254231

Coordinates of trip point to school
Recorded from GPS
34.08273, -118.2542

Los Angeles County Simulation



**100% Sample
of entire
LA County**



Los Angeles County
GIS Data Portal – Open Source
~2.2 million parcels layer

~200,000 parcels removed
No Address / Duplicate Addresses

**~2 million parcels
w/ unique addresses**

Los Angeles County Simulation

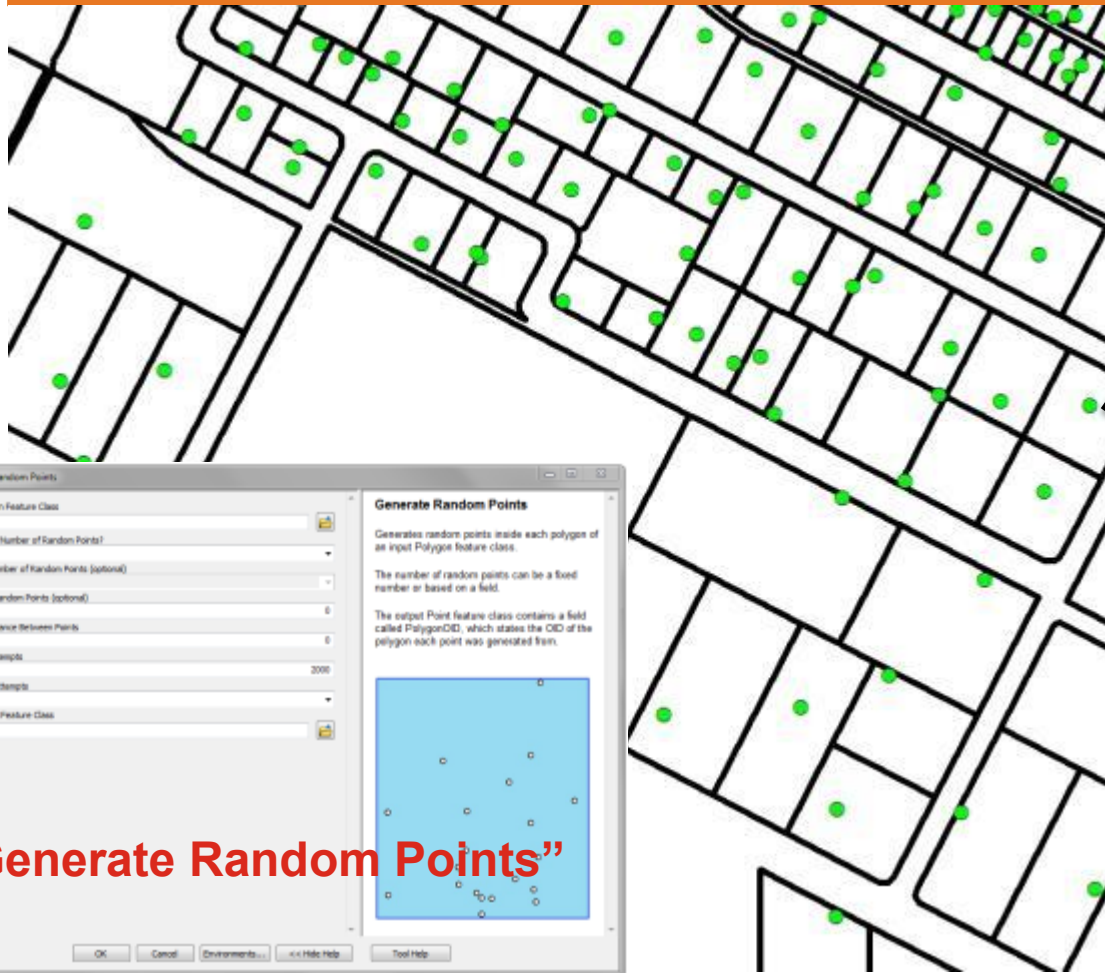


In travel surveys, GPS points vary in location within parcels

Parcel vary in in size & shape
Buildings / Visted Locations within parcels vary

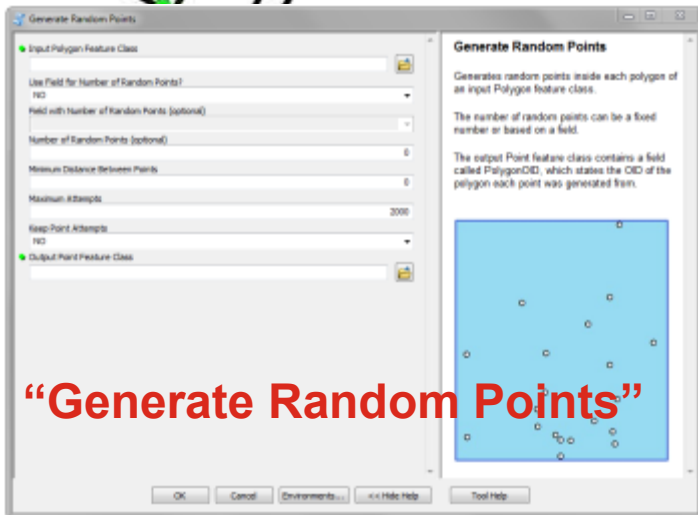
GPS Trip Origins 
GPS Trip Destinations 

Los Angeles County Simulation



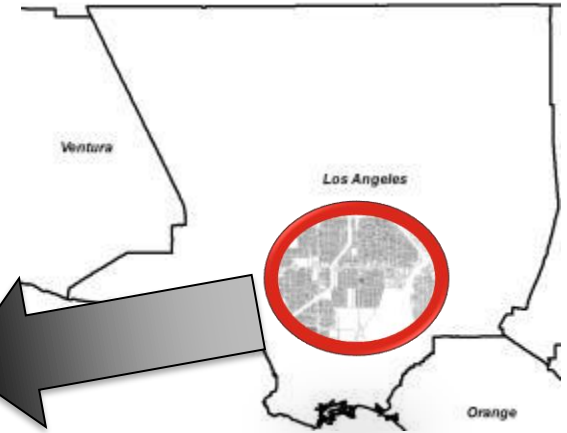
**1 random point per parcel
= Simulated GPS Data Points**

**~2 million random points in
LA County Parcels**



“Generate Random Points”

Los Angeles County Simulation



Auto Batch Geocoding:
We built an **ESRI GIS Locator** from
Open Source LA County Network
~94% geocoded (~1.9m)
~6% ungeocoded (~100k)

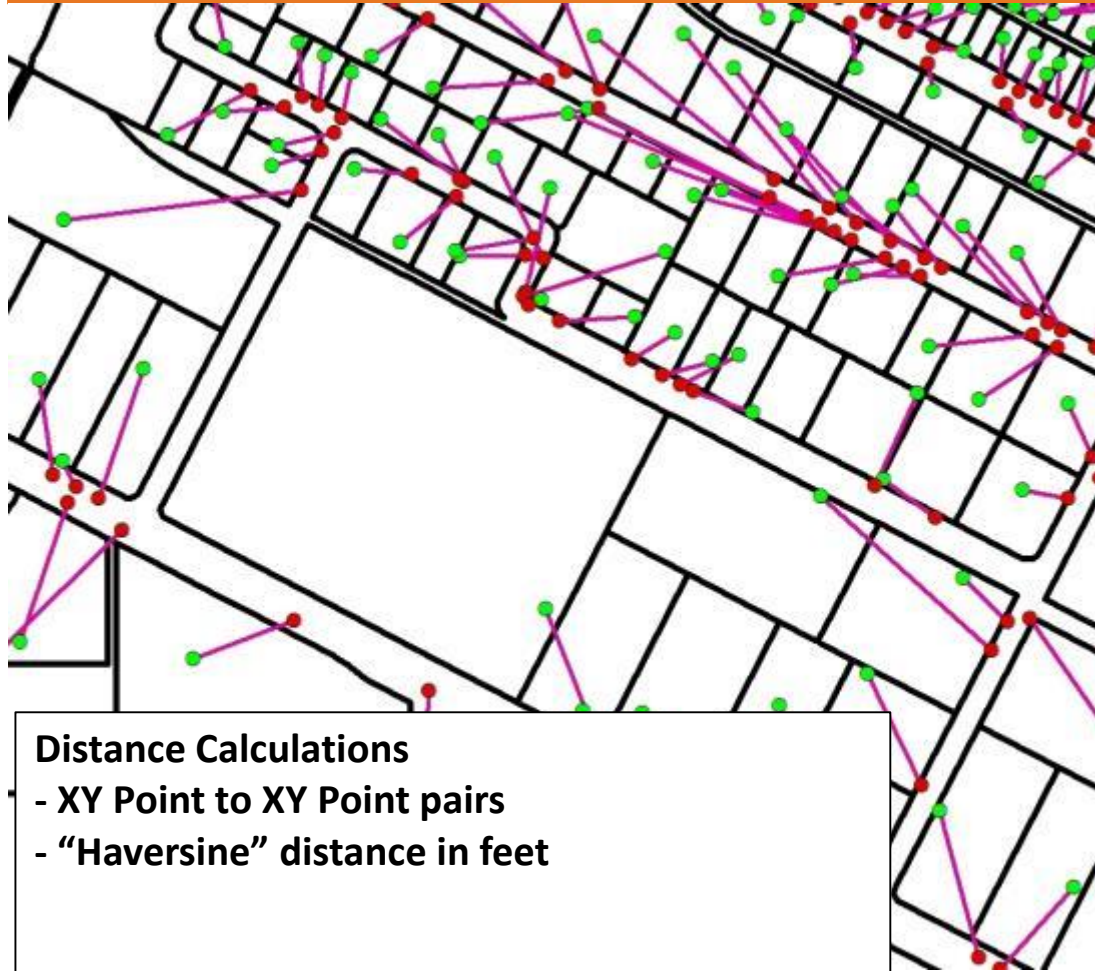
*Ungeocoded =
Property Address Data Errors
"0 PYLON SIGN SITE CERRITOS, CA"*

PropertyLocation
12363 185TH ST CERRITOS CA 90703
12355 HEDDA DR CERRITOS CA 90703
12351 HEDDA DR CERRITOS CA 90703
12357 LA JARA LN CERRITOS CA 90703
12238 186TH ST ARTESIA CA 90701
12249 185TH ST ARTESIA CA 90701
18617 CORTNER AVE CERRITOS CA 90703
18825 CORTNER AVE CERRITOS CA 90703

**1 geocoded address per parcel
= Simulated Self-Reported
Addresses**

**~2 million addresses
from LA County parcel
"property location"
geocoded**

Los Angeles County Simulation



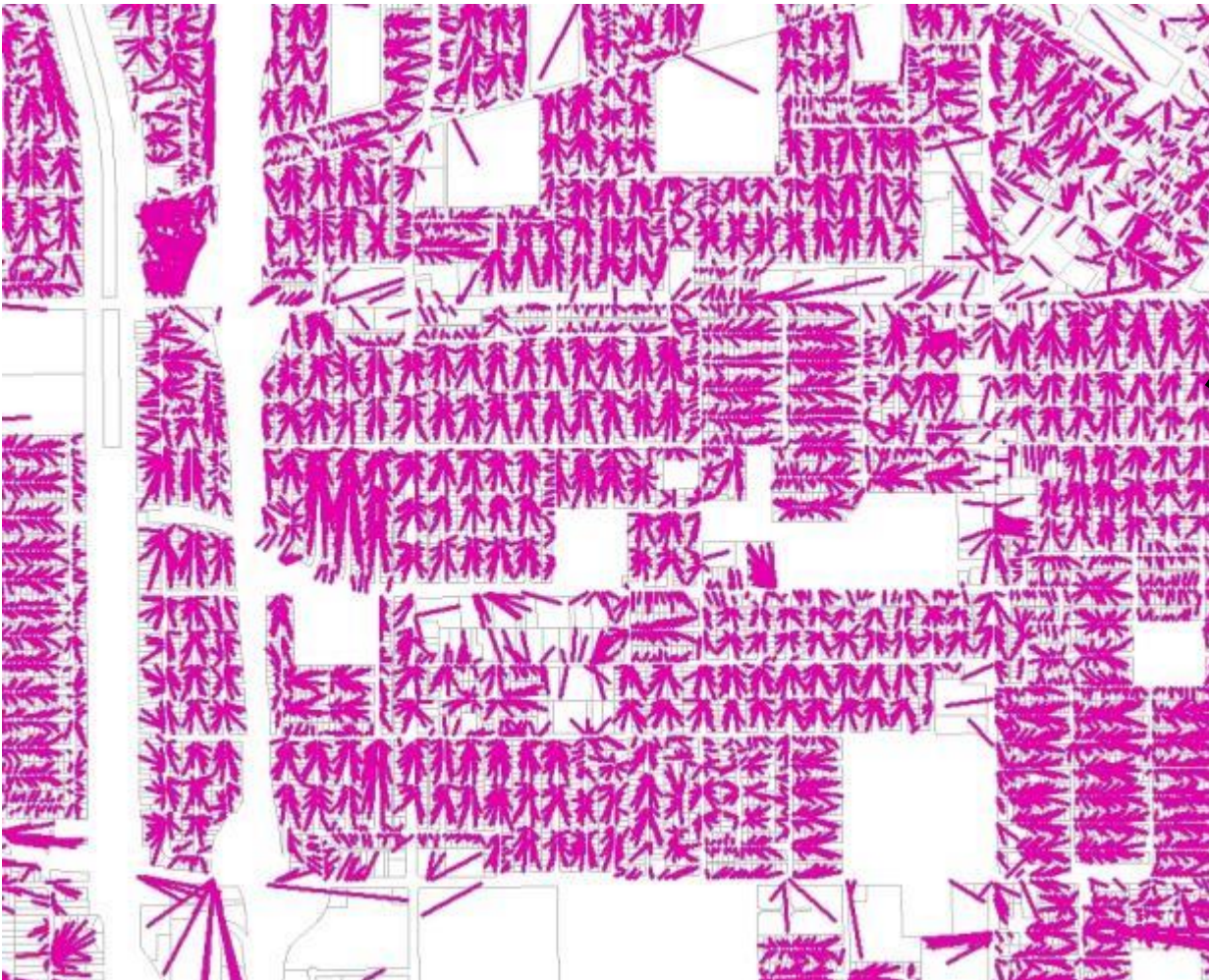
GPS point vs. Self-Report Addresses

~2 million random parcel points vs. ~2 million addresses geocoded

Distance Calculations

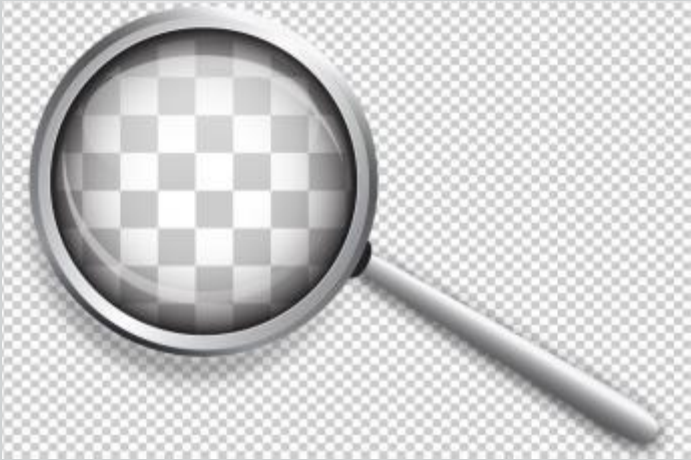
- XY Point to XY Point pairs
- "Haversine" distance in feet

Los Angeles County Simulation



**GPS point vs. Self-Report
Addresses**

Findings



Findings - Overall



- **170.8 ft Mean Distance**
126.7 ft Median Distance
- Self reported geocoded addresses vs. GPS point on parcel
- Variation by land use:

General Use Type	Median Distance	Mean Distance	Number of Parcels
Dry Farm	705.7	8100.0	108
Irrigated Farm	448.5	2172.0	189
Miscellaneous	237.5	705.5	3,892
Industrial	188.8	323.7	35,968
Recreational	184.1	581.8	1,455
Institutional	166.0	273.5	8,340
Commercial	141.4	270.6	78,616
Unavailable	134.6	190.7	25
Residential	124.9	160.4	1,740,344
Overall	126.7	170.8	1,868,937

Findings - Residential



- **160.4 ft Mean Distance**
124.9 ft Median Distance
- Self reported geocoded addresses vs. GPS point on parcel

Specific Use Type	Median Distance	Mean Distance	Number of Parcels
Rooming/Boarding House	139.3	231.9	424
Homes for Aged and Others	163.4	224.7	920
Five or More Units or Apartments (Any Combination)	141.6	183.3	65,361
Three Units (Any Combination)	129.4	169.9	35,381
Double, Duplex, or Two Units	125.7	169.7	103,723
Four Units (Any Combination)	128.1	169.5	32,174
Single Family Residence	123.9	157.9	1,501,443
Overall	124.9	160.4	1,740,344

Findings - Commercial

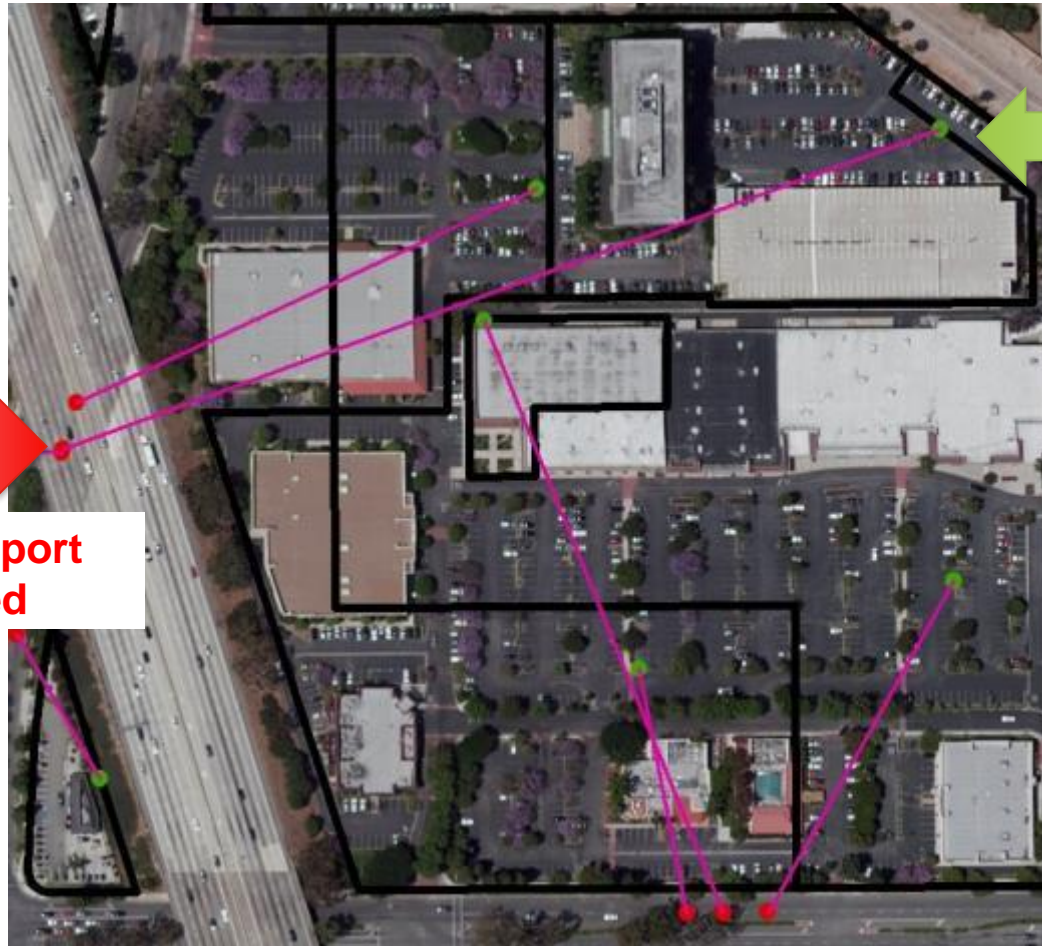


- **270.6 ft Mean Distance**
141.4 ft Median Distance
- Self reported geocoded addresses vs. GPS point on parcel

Specific Use Type	Median Distance	Mean Distance	Number of Parcels
Department Store	348.6	753.3	288
Shopping Center (Neighborhood, Community)	251.6	735.7	2,640
Hotel and Motel	156.0	522.2	1,803
Shopping Center (Regional)	336.4	464.0	520
Theater	173.9	344.1	139
Athletic and Amusement Facility	194.5	302.9	281
Bowling Alley	269.2	287.4	40
Service Station	147.8	268.6	1,786
Supermarket	193.4	267.3	615
Parking Lot (Commercial Use Property)	143.2	264.2	7,208
Auto, Recreation Equipment, Construction Equipment Sales and Service	143.3	252.3	7,598
Office Building	149.1	243.7	11,099
Bank, Savings and Loan	162.7	220.8	871
Professional Building	139.9	208.6	3,643
Non-Auto Service and Repair Shop, Paint Shop, or Laundry	127.1	206.5	903
Overall	141.4	270.6	78,616



Findings - Commercial



**Simulated GPS point:
Parking lot space**

**Commercial / Industrial
Parcels represented
greater margin of
distance error**

**Simulated Self-Report
Address Geocoded**

Findings



- Use of self-report addresses more justifiable for residential parcels than for commercial parcels
- Single family households have lowest error (124 feet median distance)
- Department stores, shopping centers have highest error (300-700 feet distance)
- “True” participant commercial/industrial/institutional locations better served by GPS data than geocoding

Greater Significance



- Travel Surveys require great detail → high respondent burden, lower recruit rates
- Need Return on Investment = high location precision, accuracy from participants
- Collection of self-report addresses adds time, costs, potential for error, and participant burden:
 - Home Address? Always collected, **lowest error**
 - Work Address? Include for household profile? But **increased error**
 - School Address? Include for household profile? But **increased error**
 - Other / Shopping = Collection of addresses represents additional participant burden with **increased errors, no clear use**



Thank you

Timothy Michalowski, Senior Statistical GIS Analyst
Dara Seidl, Statistical GIS Analyst

t.michalowski@srbi.com

d.seidl@srbi.com

Twitter - @abtsrbi

