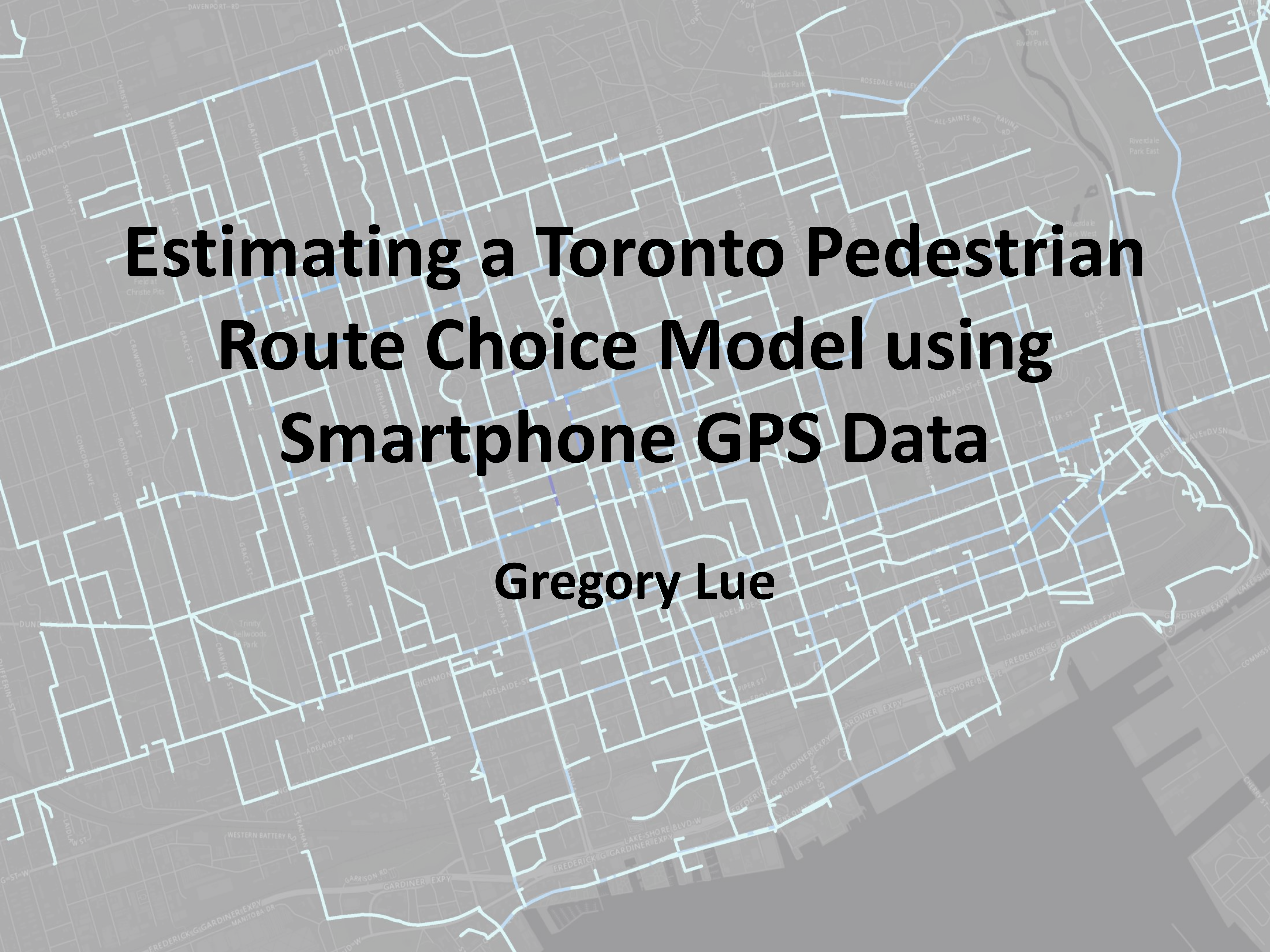
Estimating a Toronto Pedestrian Route Choice Model using Smartphone GPS Data



Gregory Lue

Presentation Outline

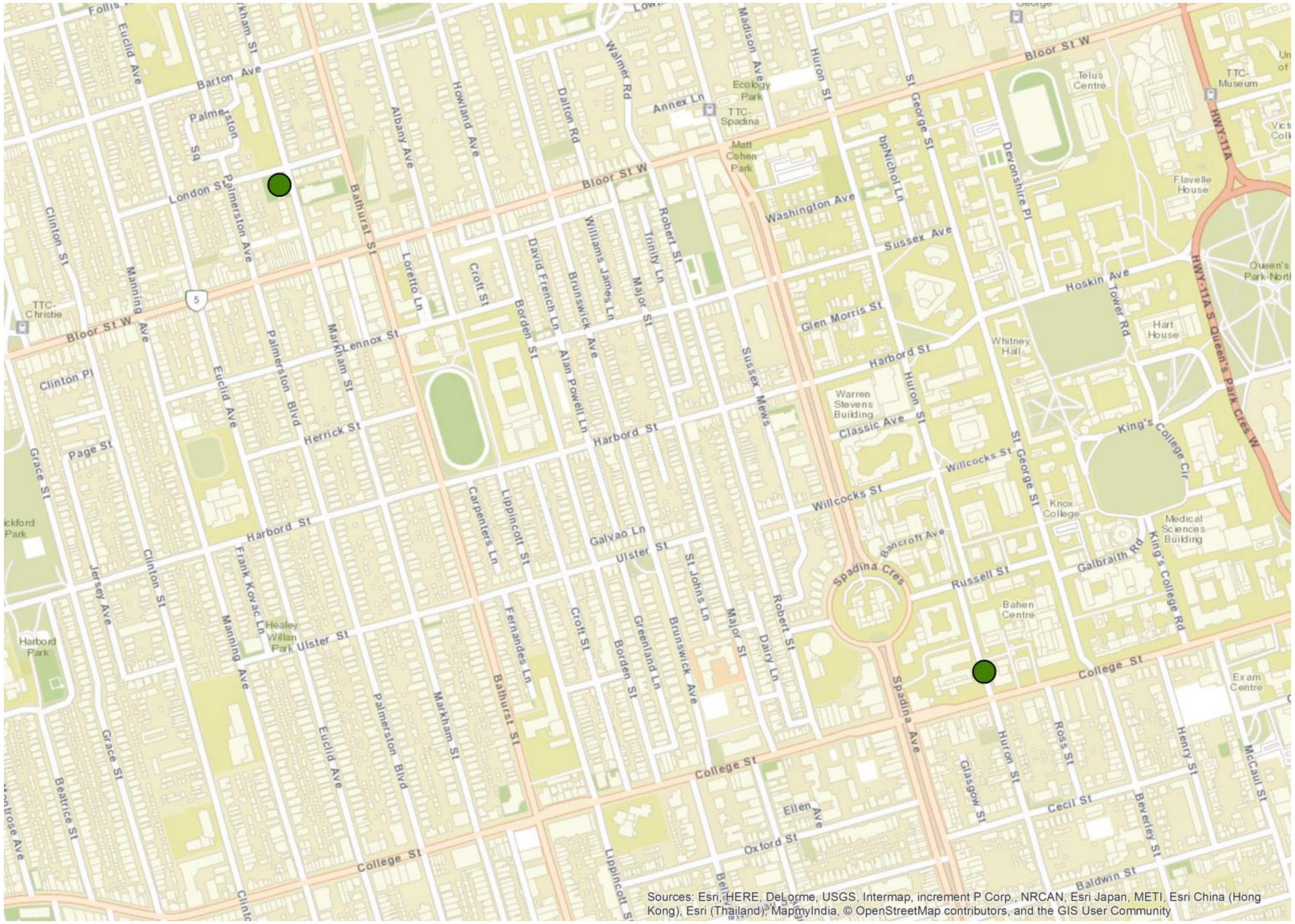
- Introduction
- Background
- Data
- Smartphone Data
- Alternative Route Generation
- Choice Model
- Toronto Case Study
- Results
- Route Generation Analysis
- Conclusions

1. Introduction

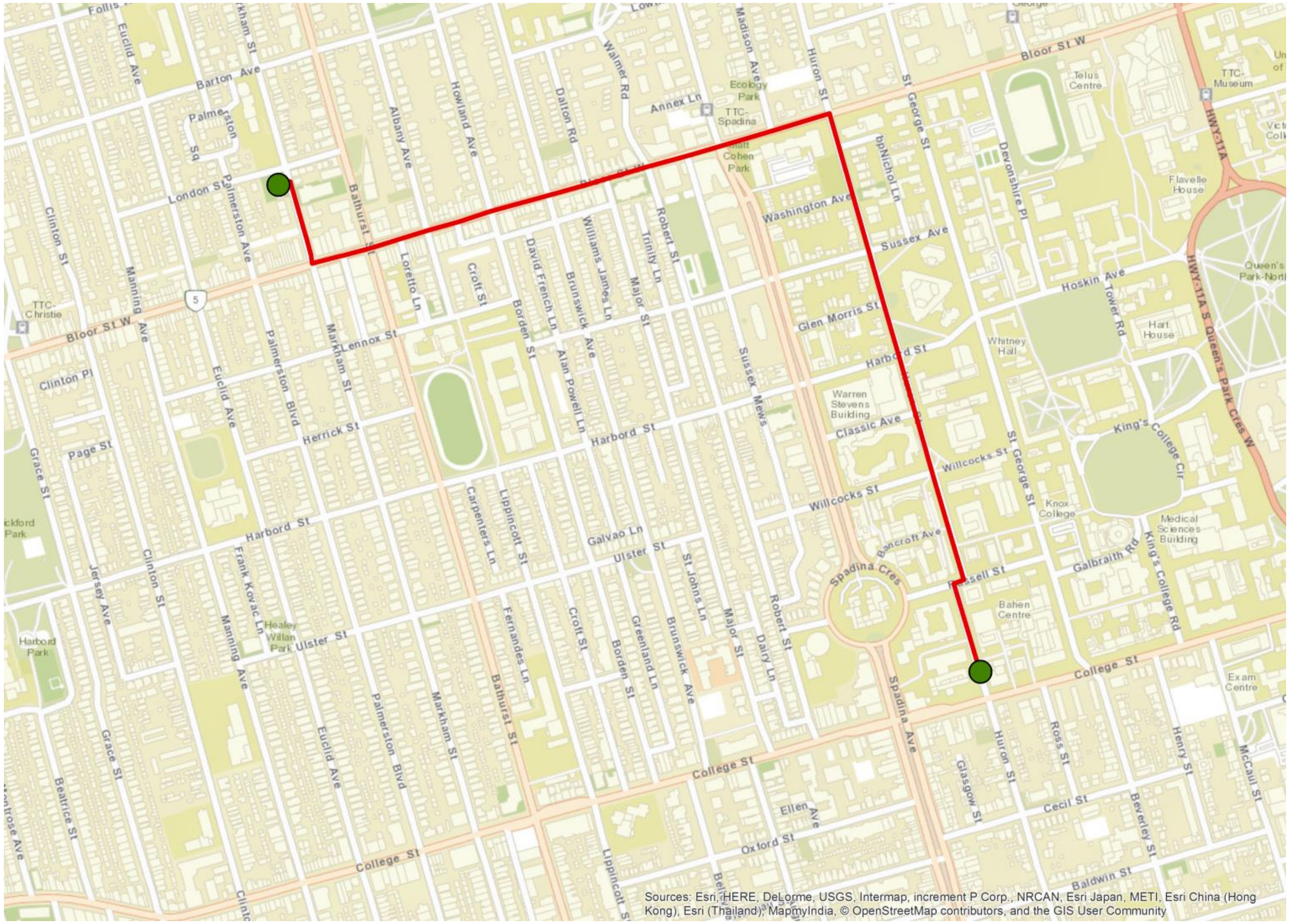
Study Motivations

- Travel demand models overlook walking trip routes
- City planning supports building walkable streets but measures are often qualitative
- Smartphone GPS surveys are becoming more common for data collection

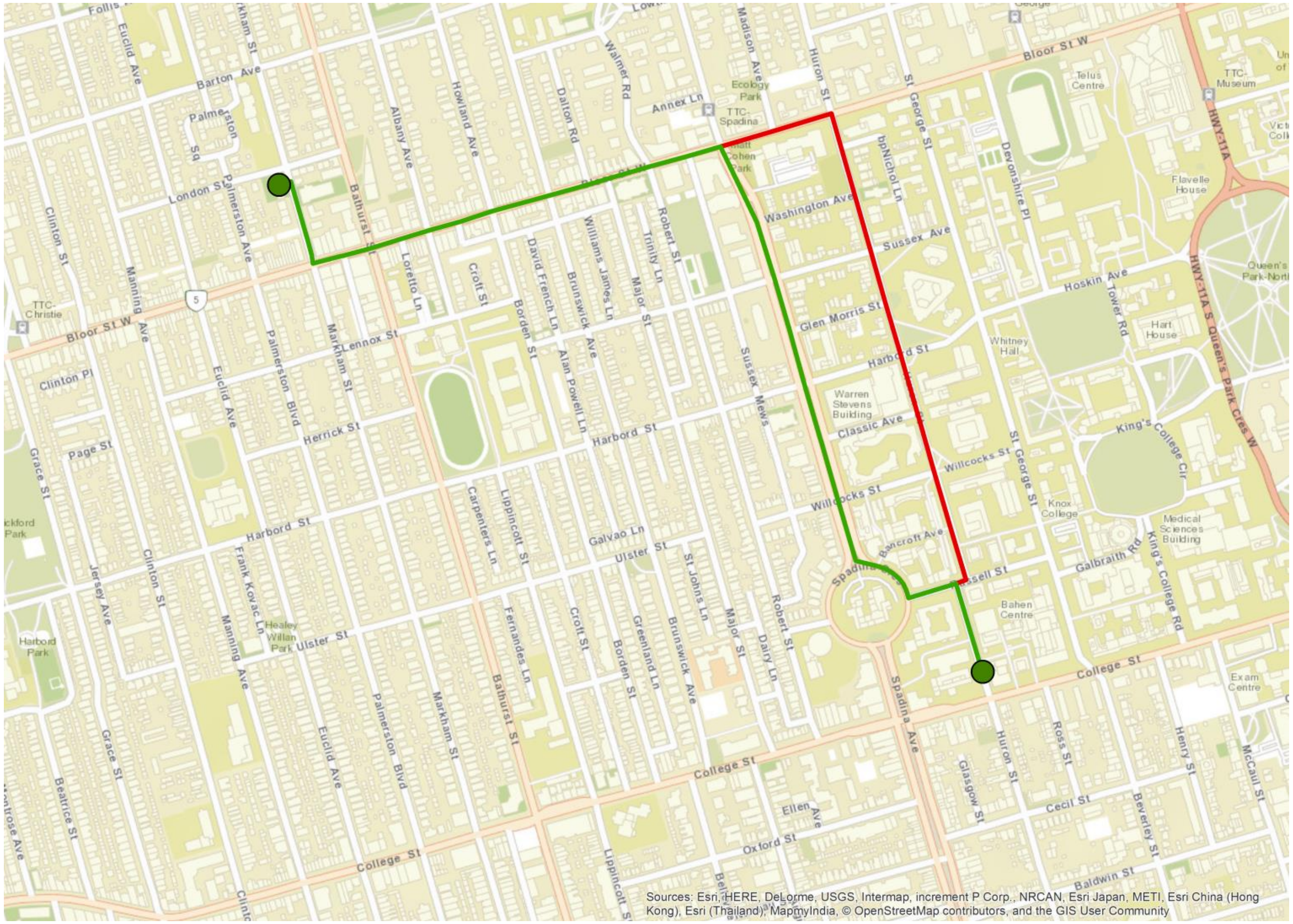
Route Choice



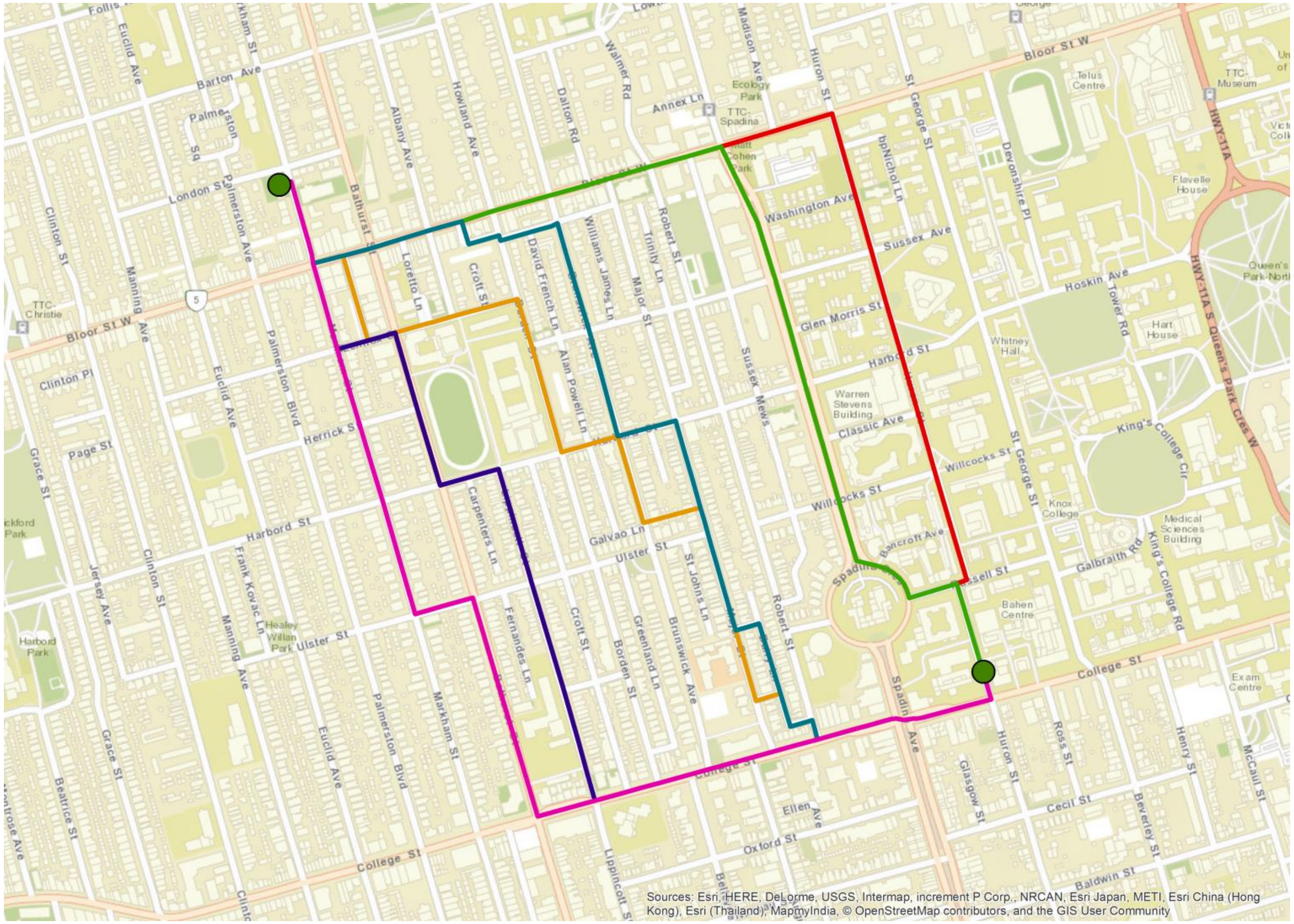
Route Choice



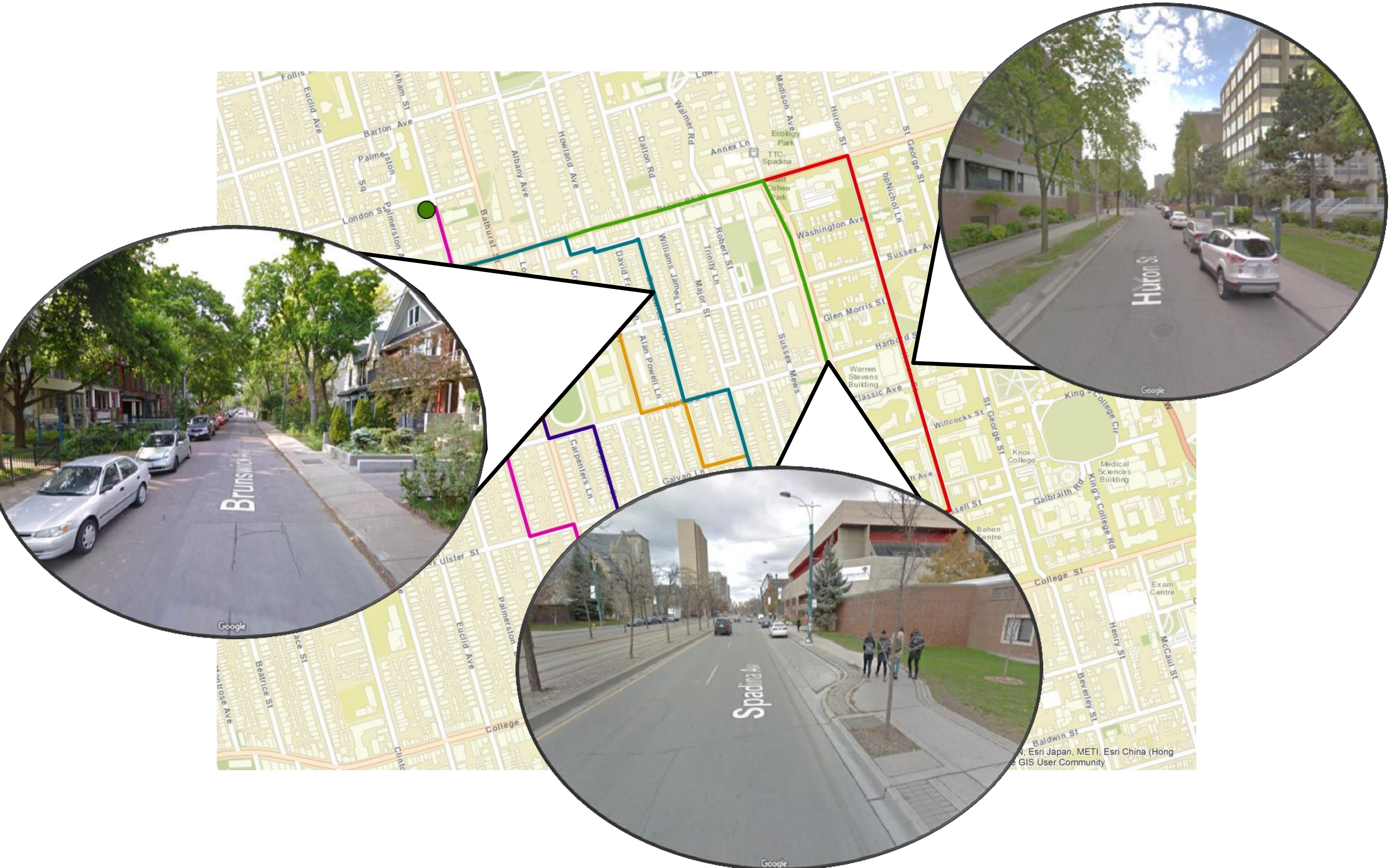
Route Choice



Route Choice



Route Choice



2. Background

Built Environment

- Built environment – Buildings, transportation systems, open space, and land-use that support communities and impact human health (City of Toronto, 2015)
- Various measures:
  - Perceived measures
  - Observed measures
  - Geographic measures

Built Environment and Pedestrian Travel

- Effects of built environment on walking rates
- Effects of built environment on walking routes
  - Very few studies
  - Mainly qualitative

Built Environment and Pedestrian Travel

- Guo (2009)
  - One more intersection per 100m increased utility by 0.3 min, increasing sidewalks by 6ft increases utility by 0.5 min, and people willing to walk 2.9 min to avoid hilly topography
- Dill and Broach (2015)
  - turns equivalent to +50m, upslopes of 10% are twice as costly, unsignalized arterial path perceived as +70m, busy roads 14% longer, commercial neighborhoods 28% shorter

3. Data

Street Network Data

- Toronto Open Data
  - Street Network
  - Sidewalk Conditions
  - Signalized Intersection Locations
  - Land Use
- Elevation
- Walk Score

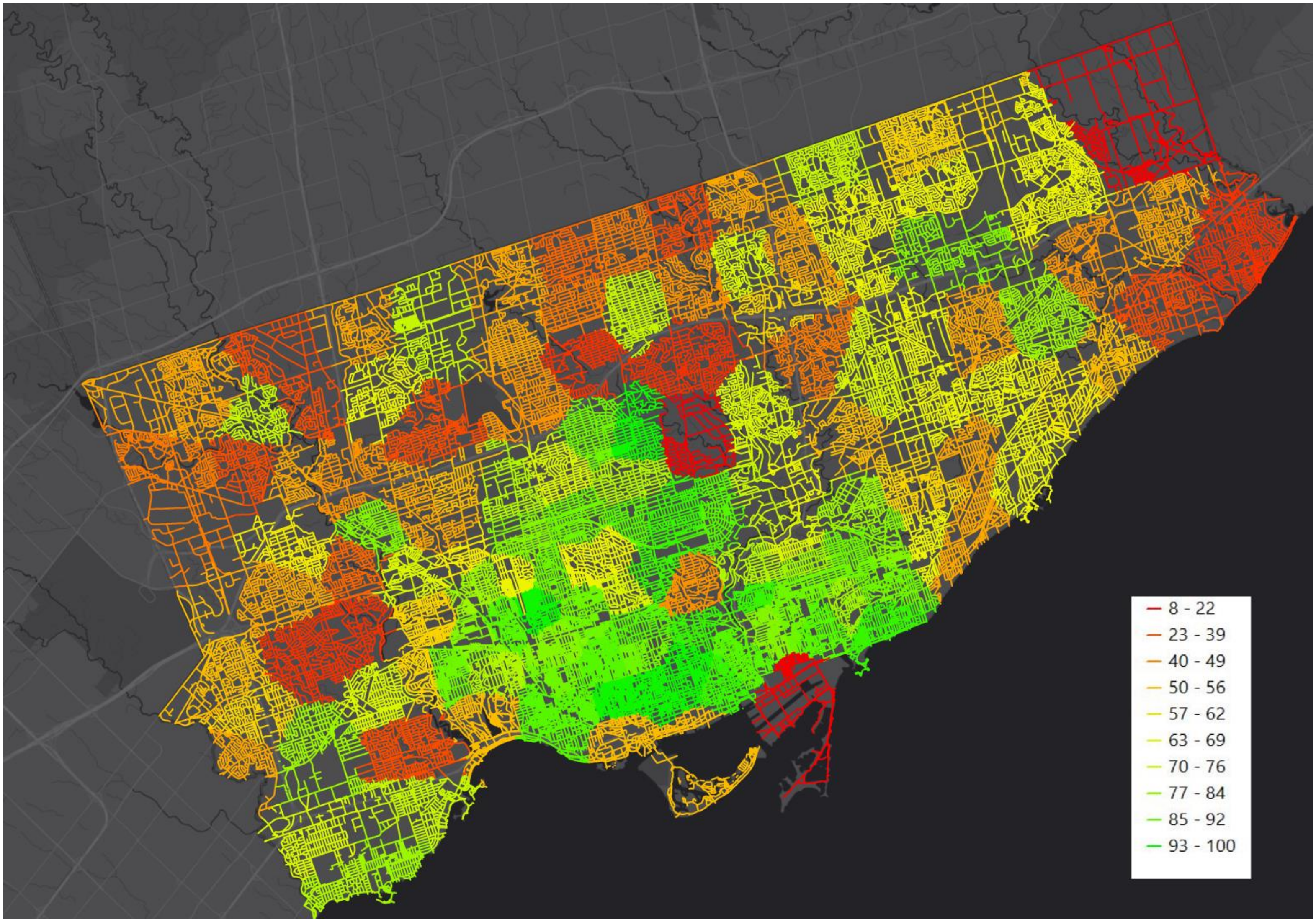
Walk Score

- Considers proximity to amenities, walking infrastructure, population density, block length, intersection density

Walk Score	Description
90-100	Walker's Paradise - Daily errands do not require a car
70-89	Very Walkable - Most errands can be accomplished on foot
50-69	Somewhat Walkable - Some errands can be accomplished on foot
25-49	Car-Dependent - Most errands require a car
0-24	Car-Dependent - Almost all errands require a car

(Walk Score, 2016)

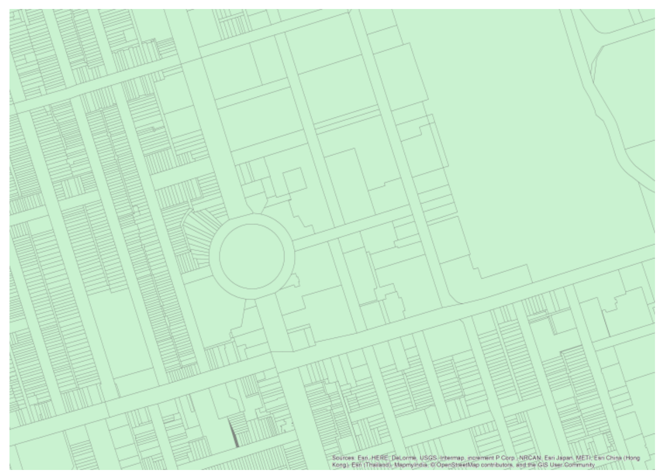
Walk Score



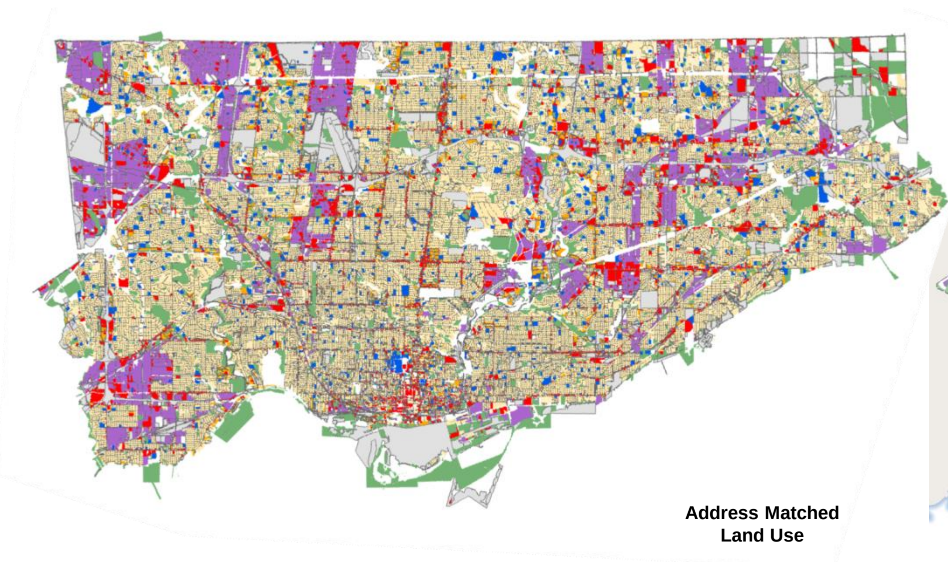
Land Use

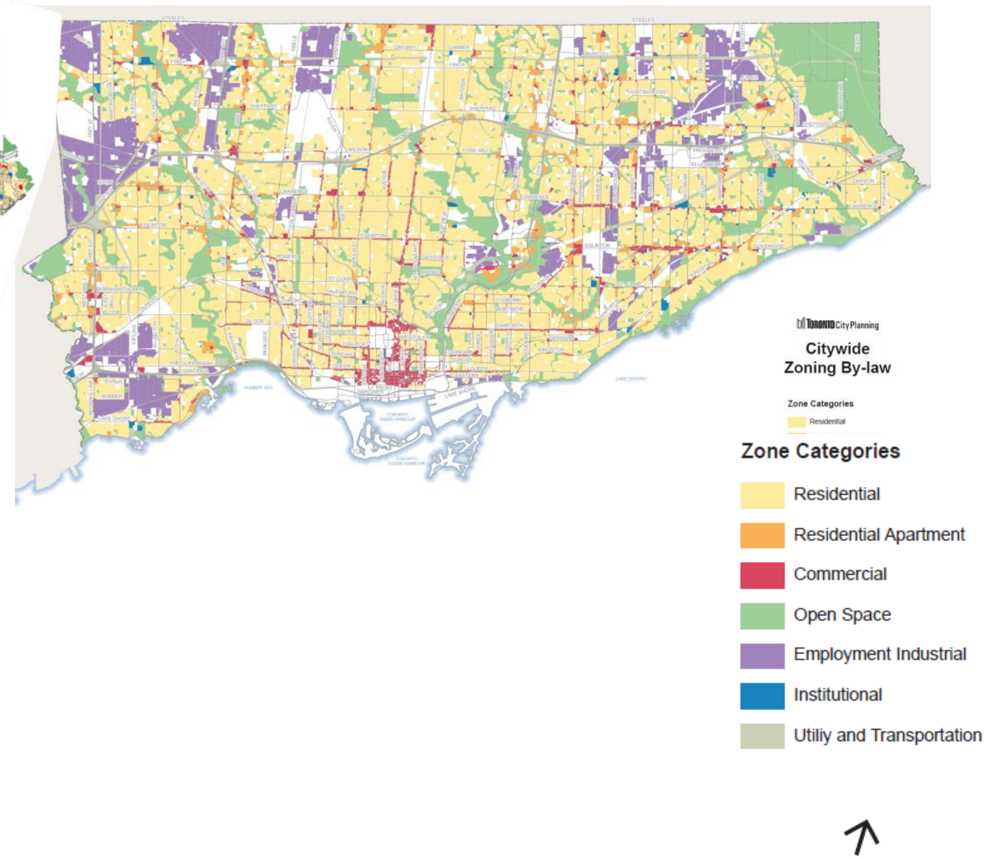
- Address point with land use
- Land parcel

Need to merge these files and convert into a "land use frontage" measure

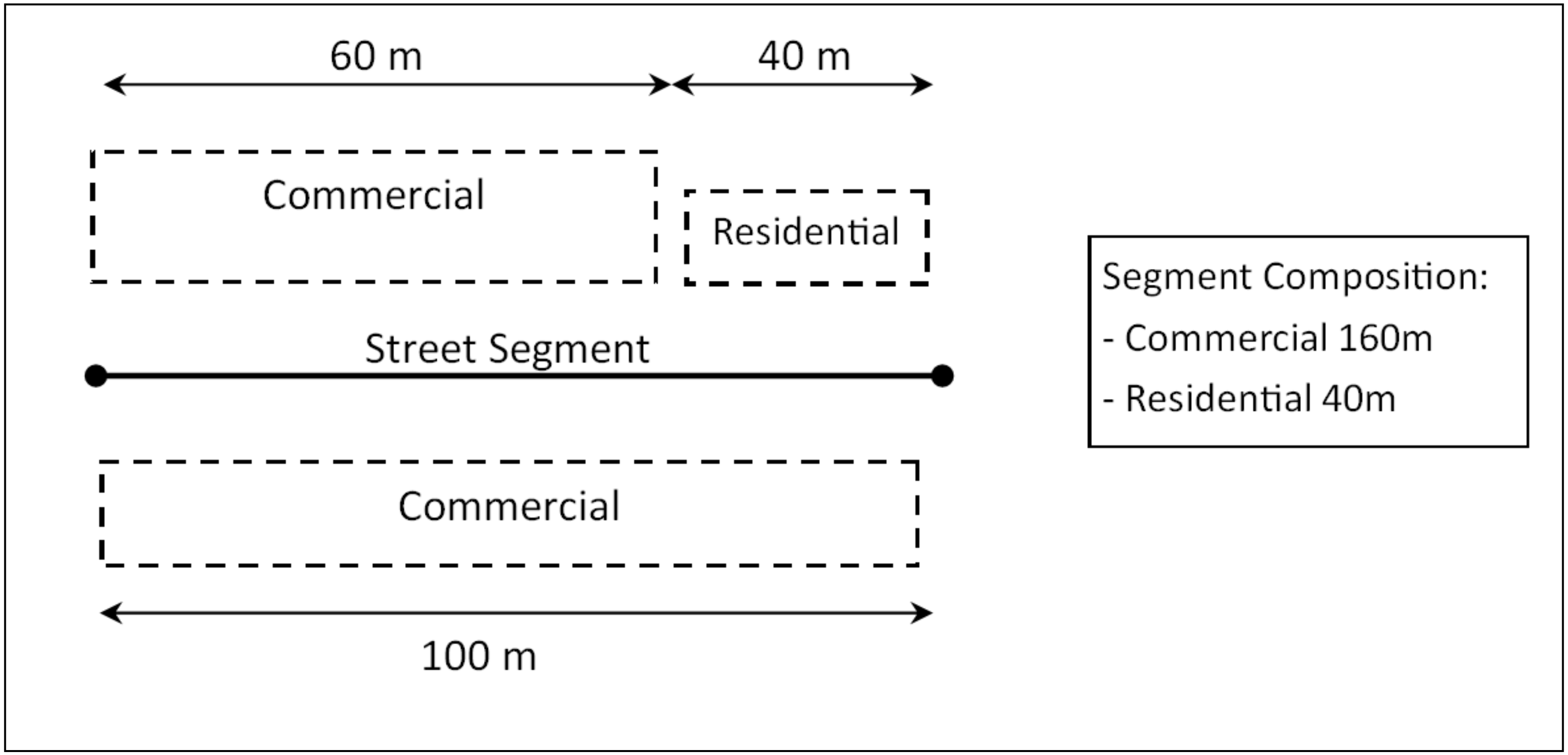


Land Use Comparison





Land Use


60 m
40 m
Commercial
Residential
Street Segment
Commercial
100 m
Segment Composition:
- Commercial 160m
- Residential 40m

4. Smartphone Data

Smartphone Data

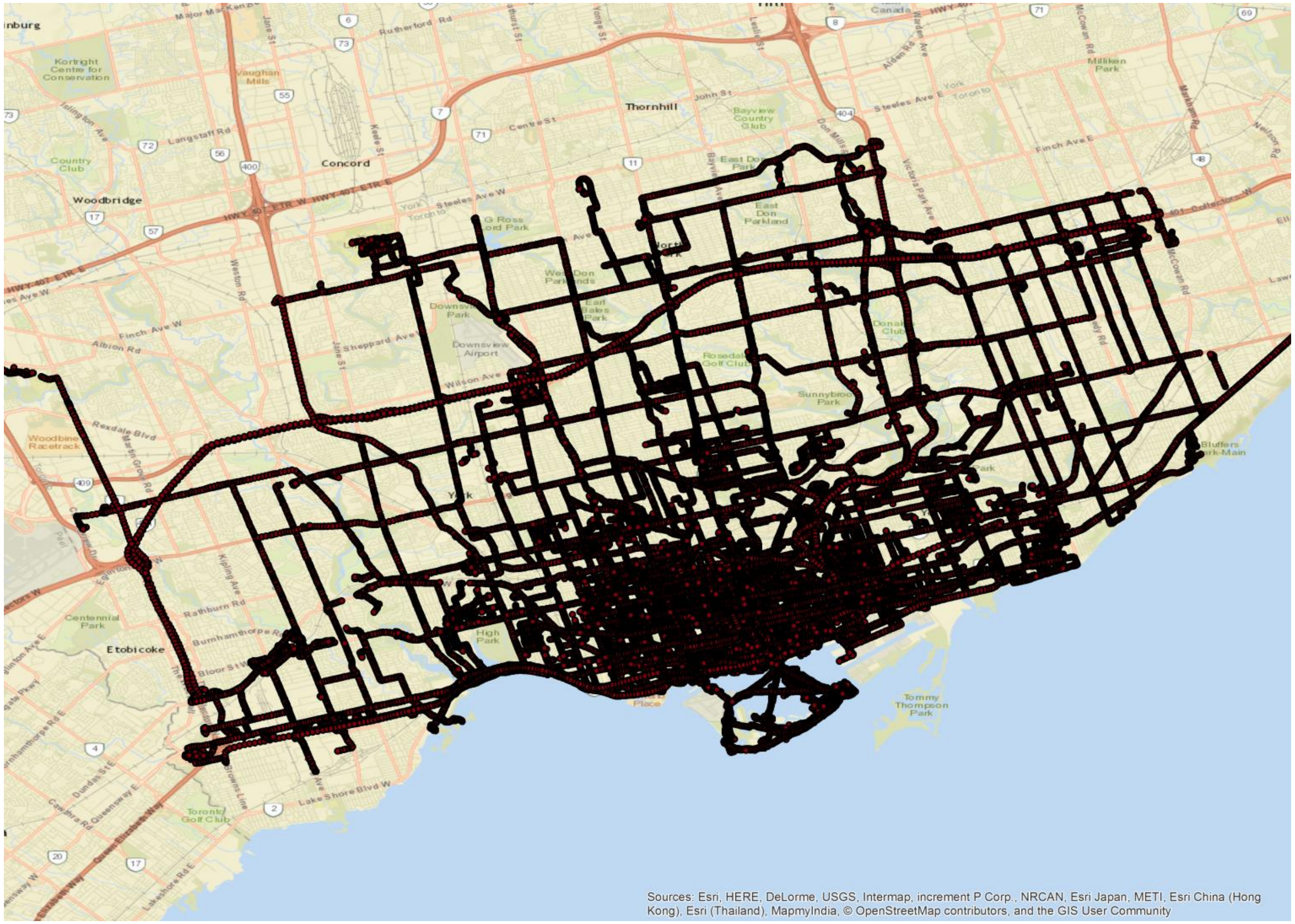
- Collected during the Waterfront Project in 2014
- 4 week survey period starting in November
- Passive GPS location
  - Records location after 50m of travel distance from previous point

Smartphone Data

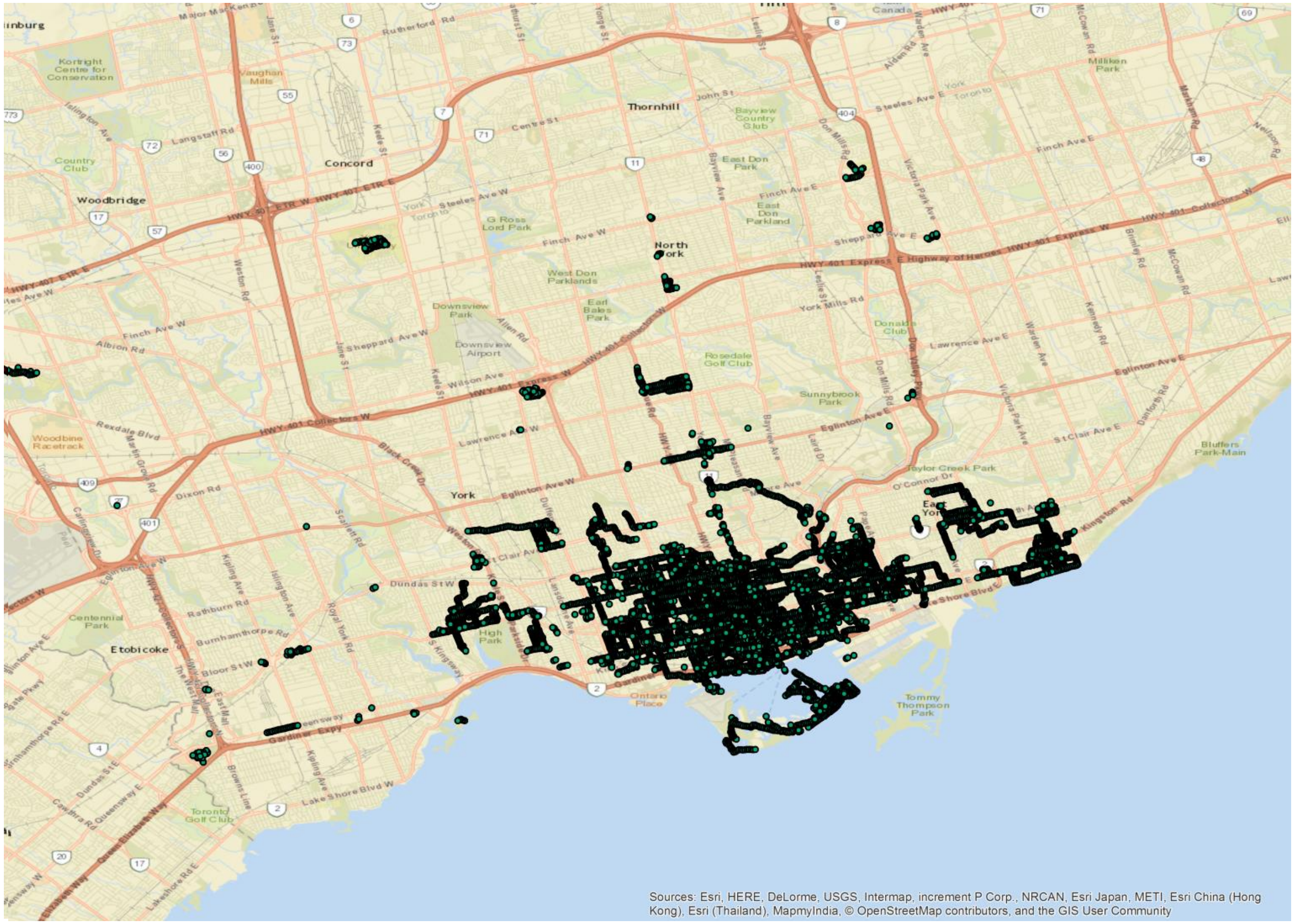
- Post Survey Data Processing
  - Trip ends determined based on 3 minute dwell time
  - Travel modes were inferred based on speed profiles (87% success rate for mode detection)
  - Trip purpose was not collected

*Outlined in paper by Harding, Zhang, & Miller (2015)

Data Cleaning

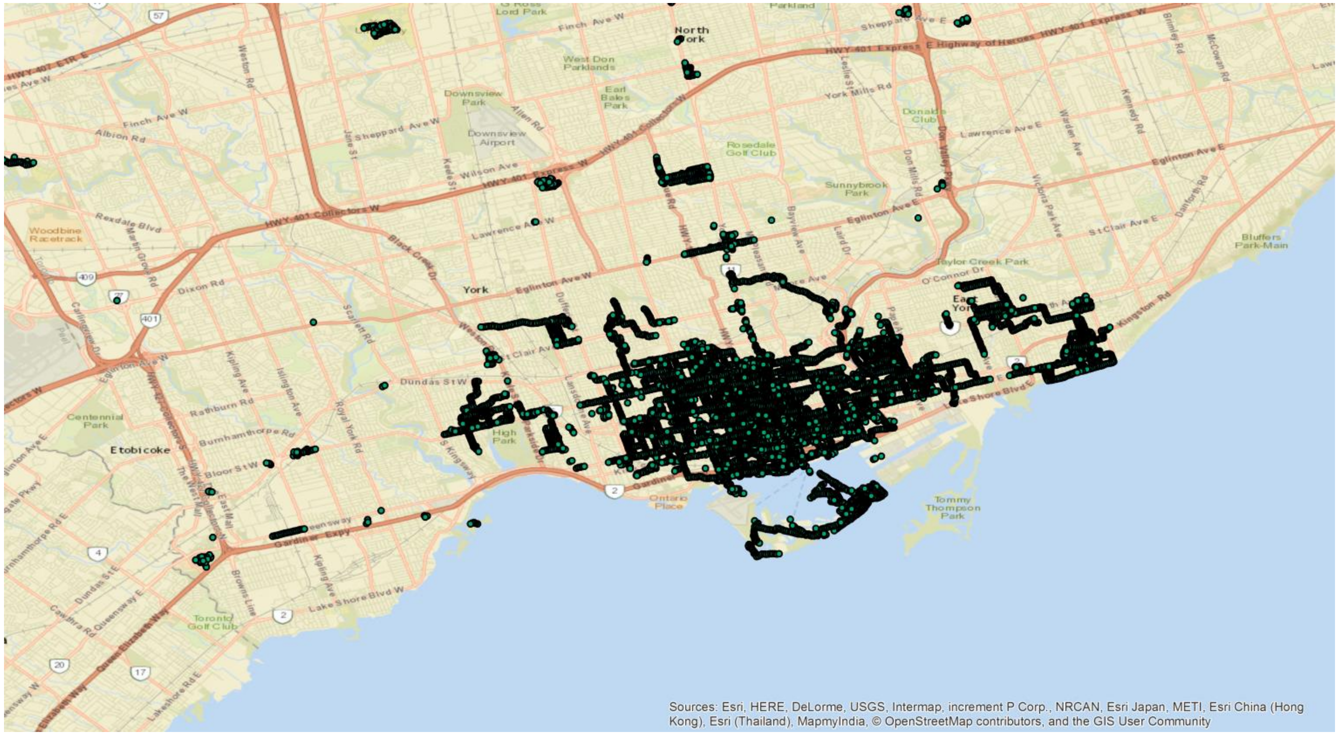


Data Cleaning



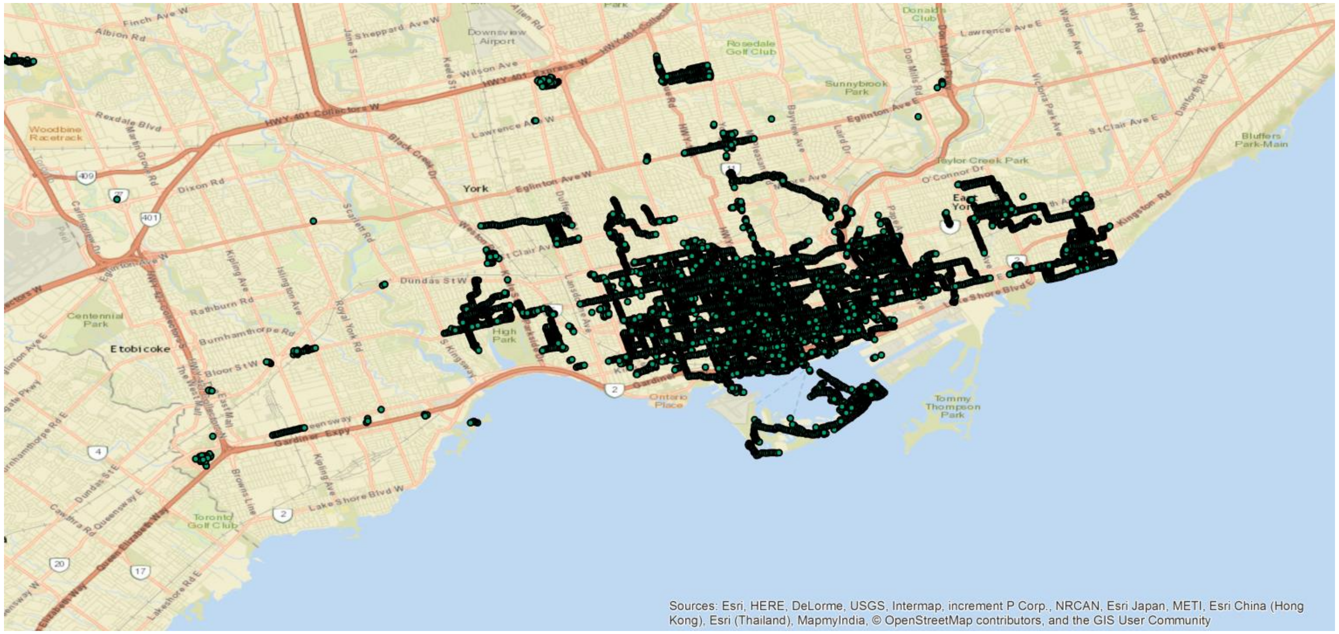
Data Cleaning

- 3193 walking trips across 103 individuals



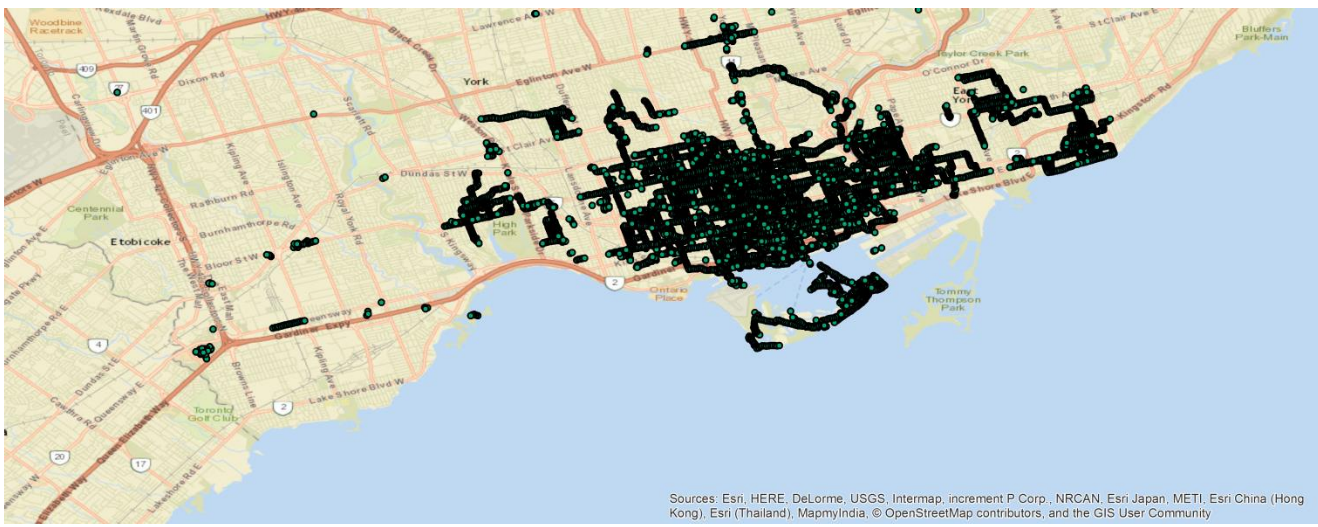
Data Cleaning

- 3193 walking trips across 103 individuals
- Remove trips with large gaps (200m)



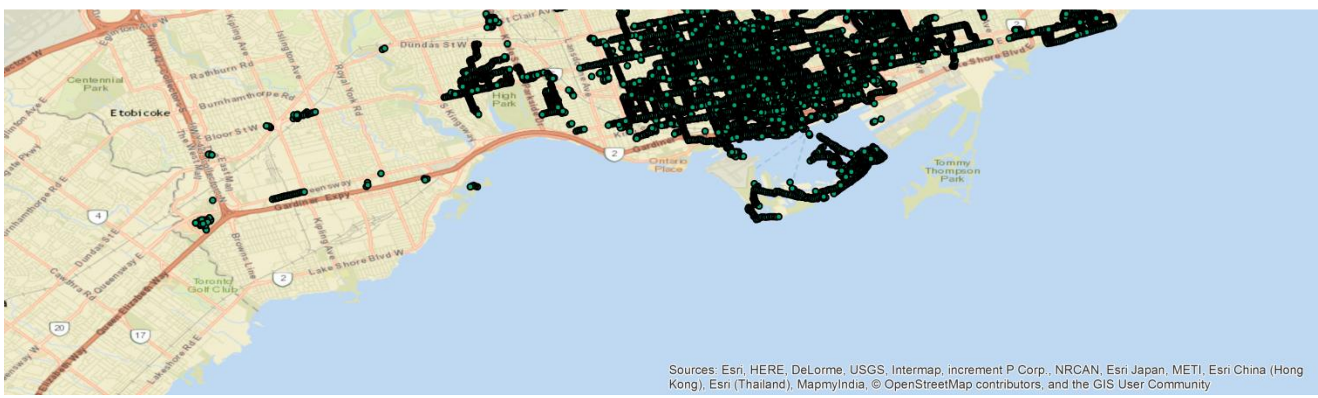
Data Cleaning

- 3193 walking trips across 103 individuals
- Remove trips with large gaps (200m)
- Remove trips with 3 or less points

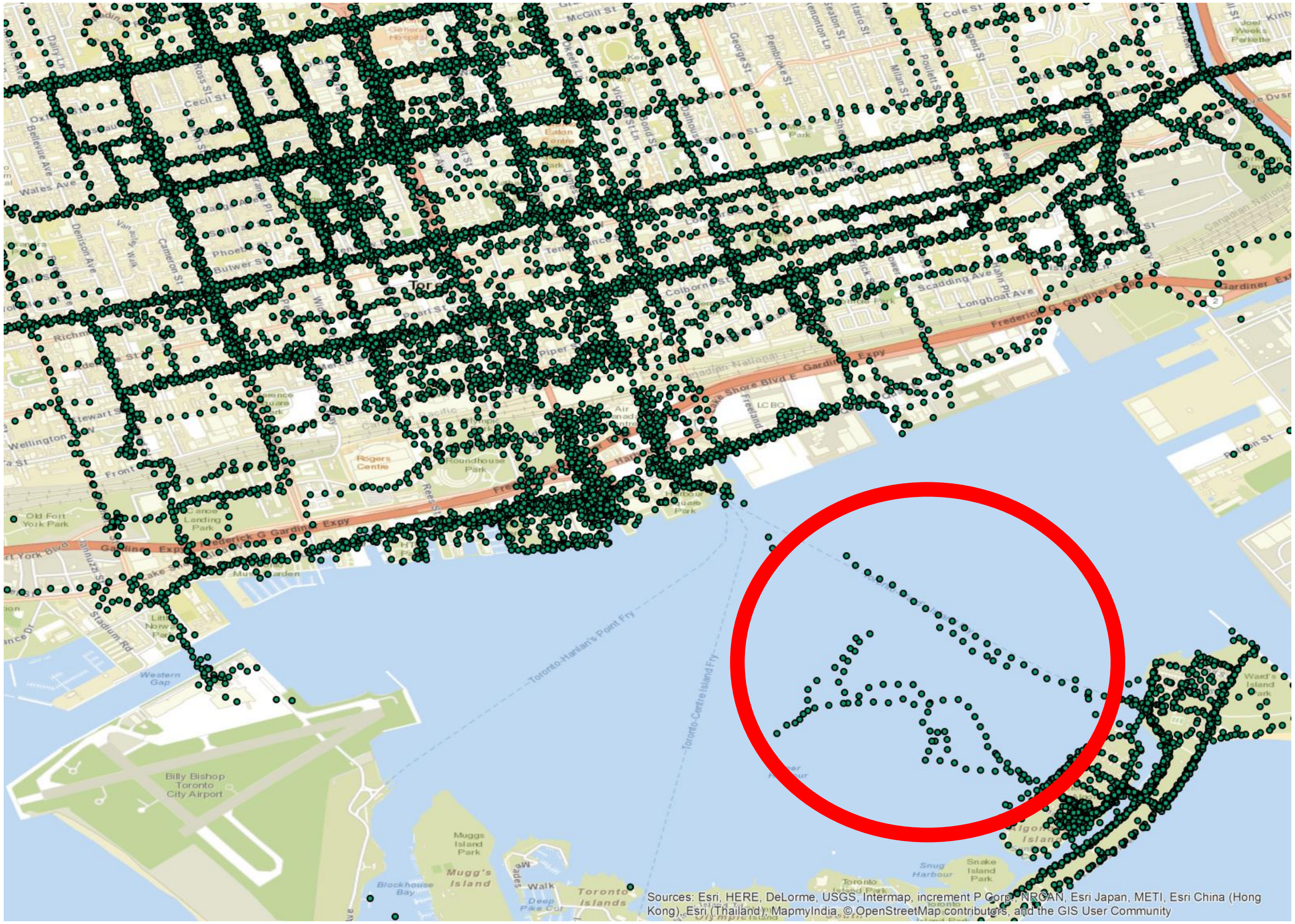


Data Cleaning

- 3193 walking trips across 103 individuals
- Remove trips with large gaps (200m)
- Remove trips with 3 or less points
- Remove mislabelled walk trips

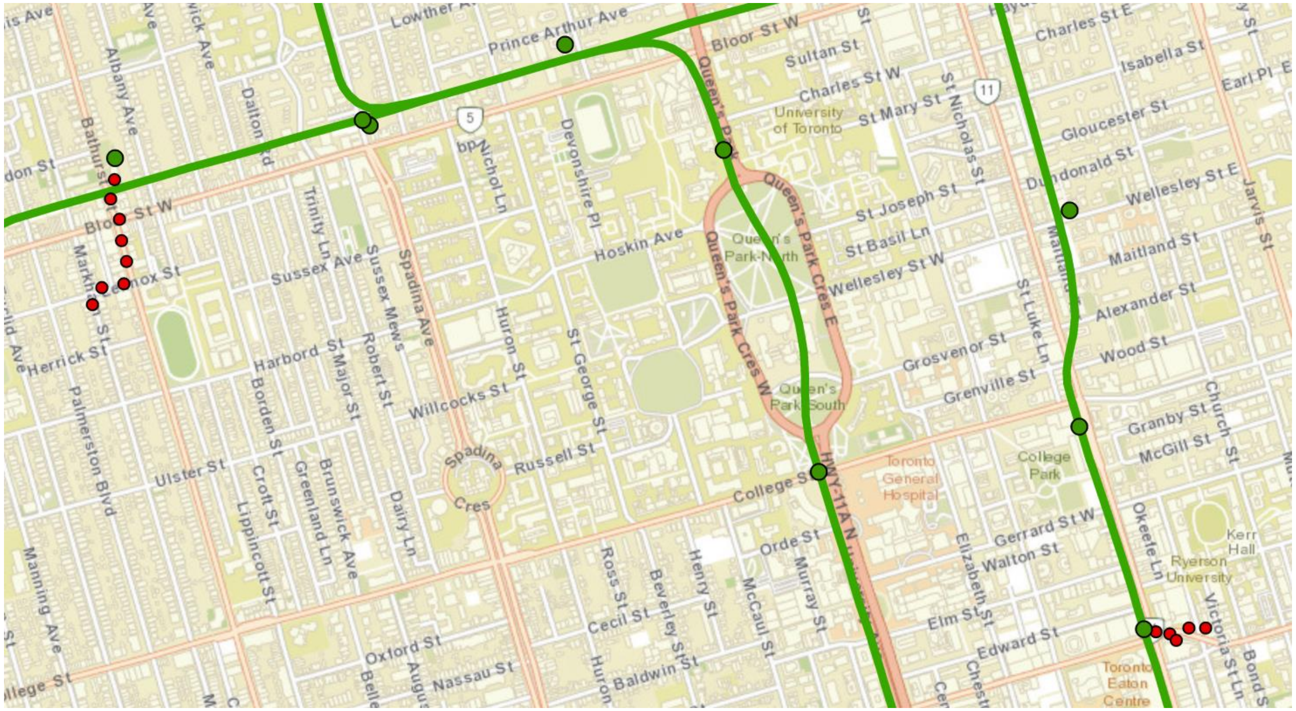


Data Cleaning



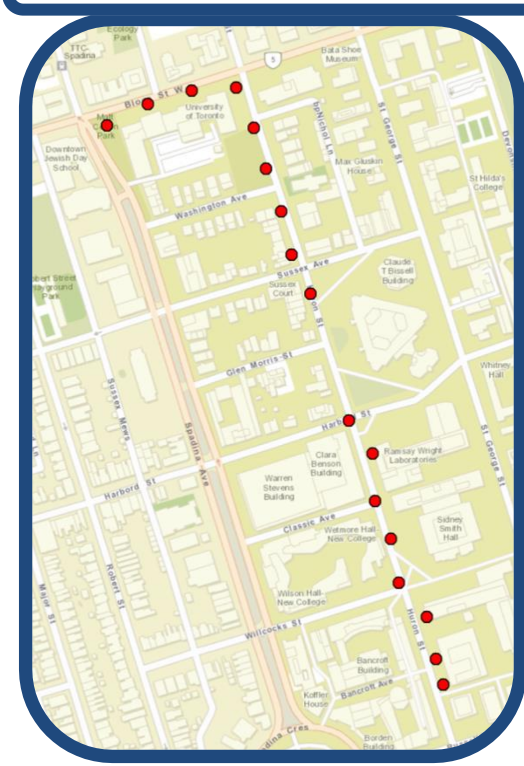
Large Gap Trips

- Check gaps if they coincide with subway stations
- Break trip into two walking trips

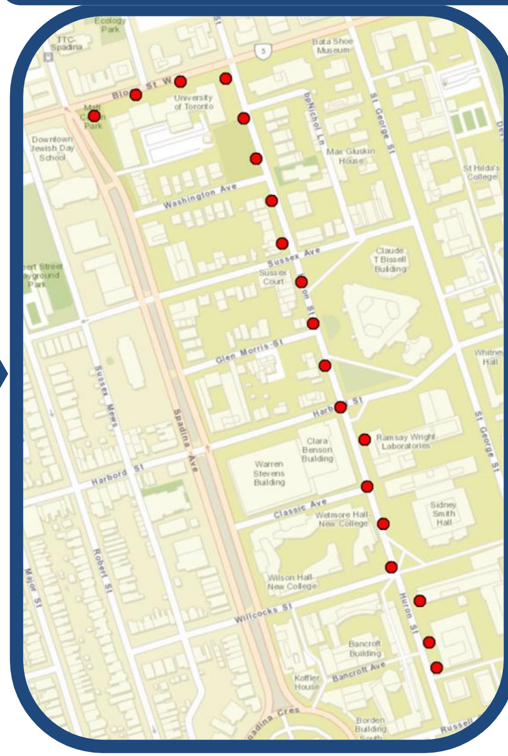


Walk Trip Solving Process

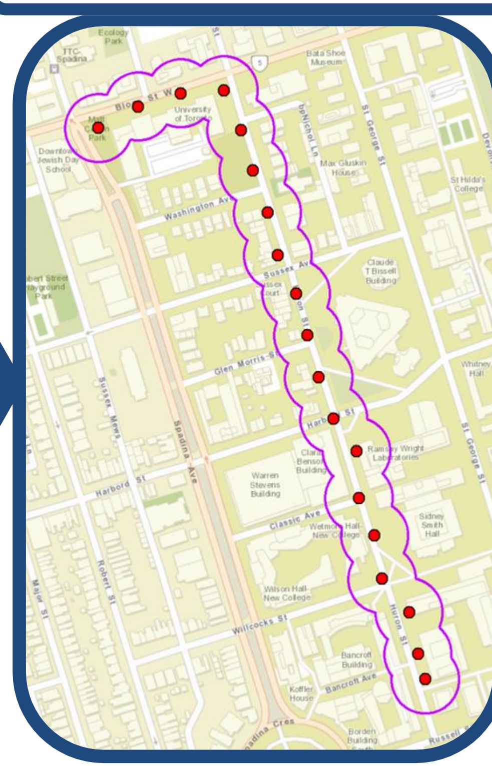
1. Import GPS Points



2. Fill Gaps

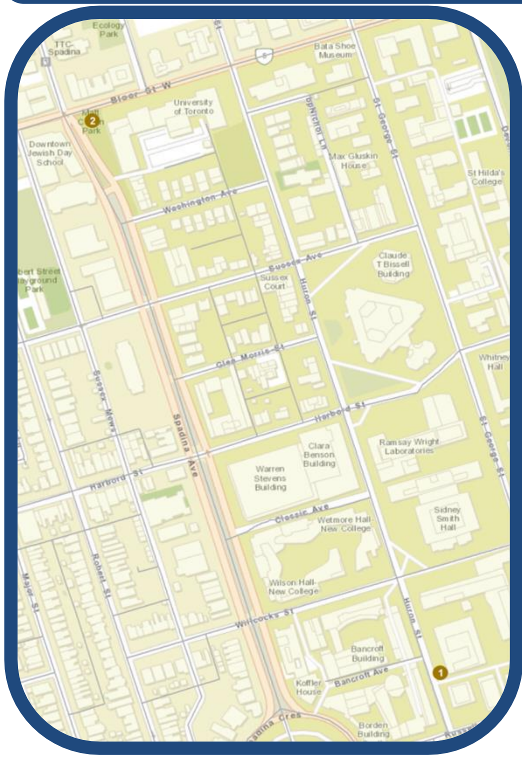


3. Create buffer area

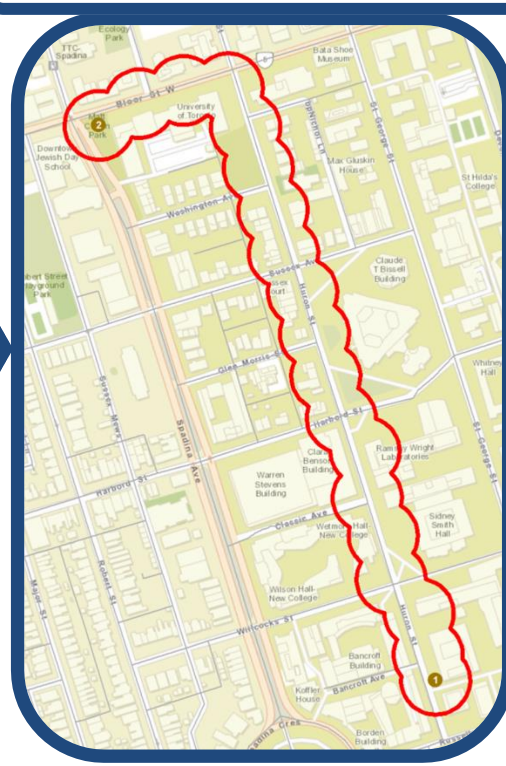


Walk Trip Solving Process

4. Add Origin/Destination



5. Add Buffer Restriction



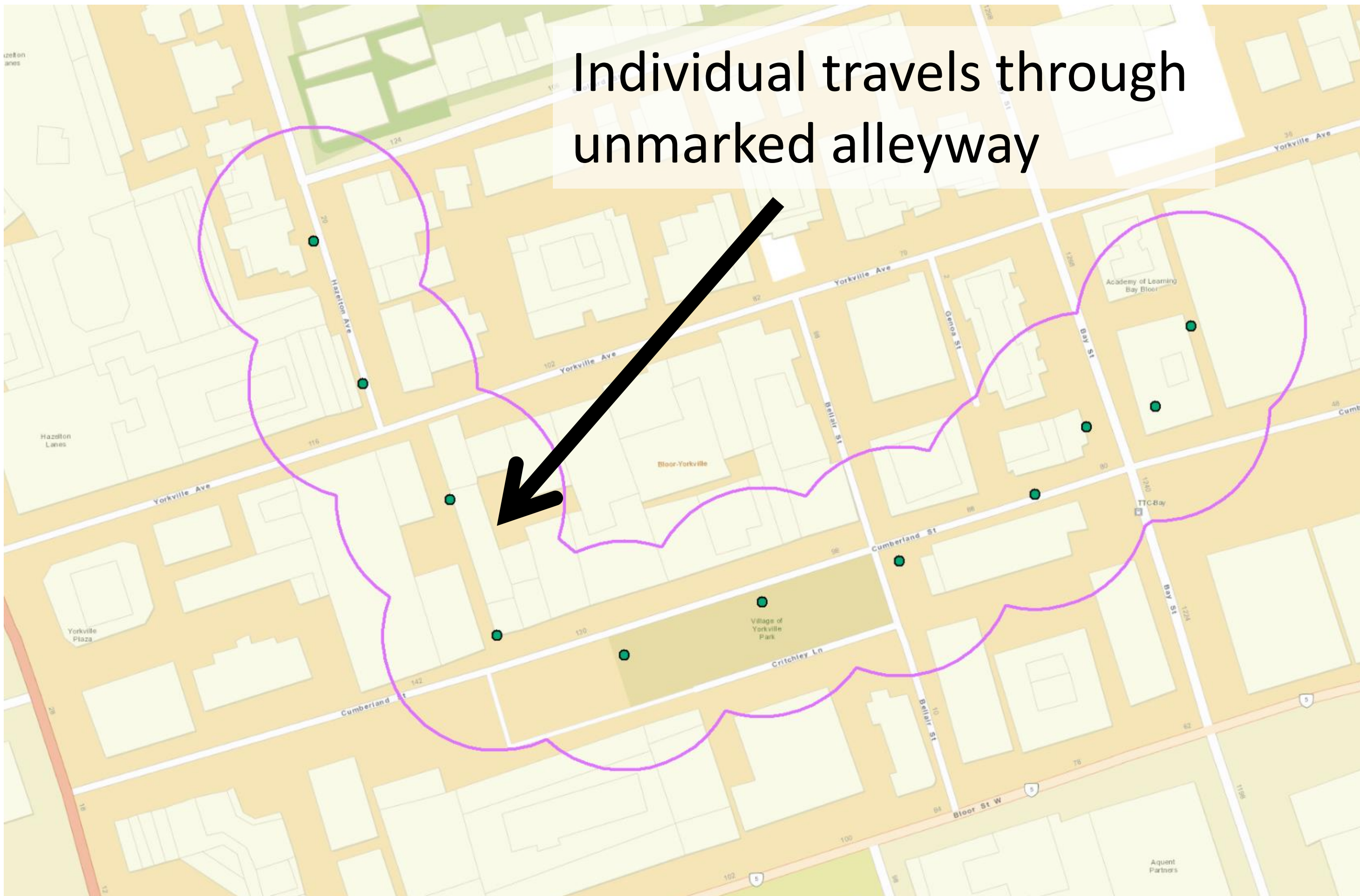
6. Solve Route

(Dalumpines & Scott, 2011)

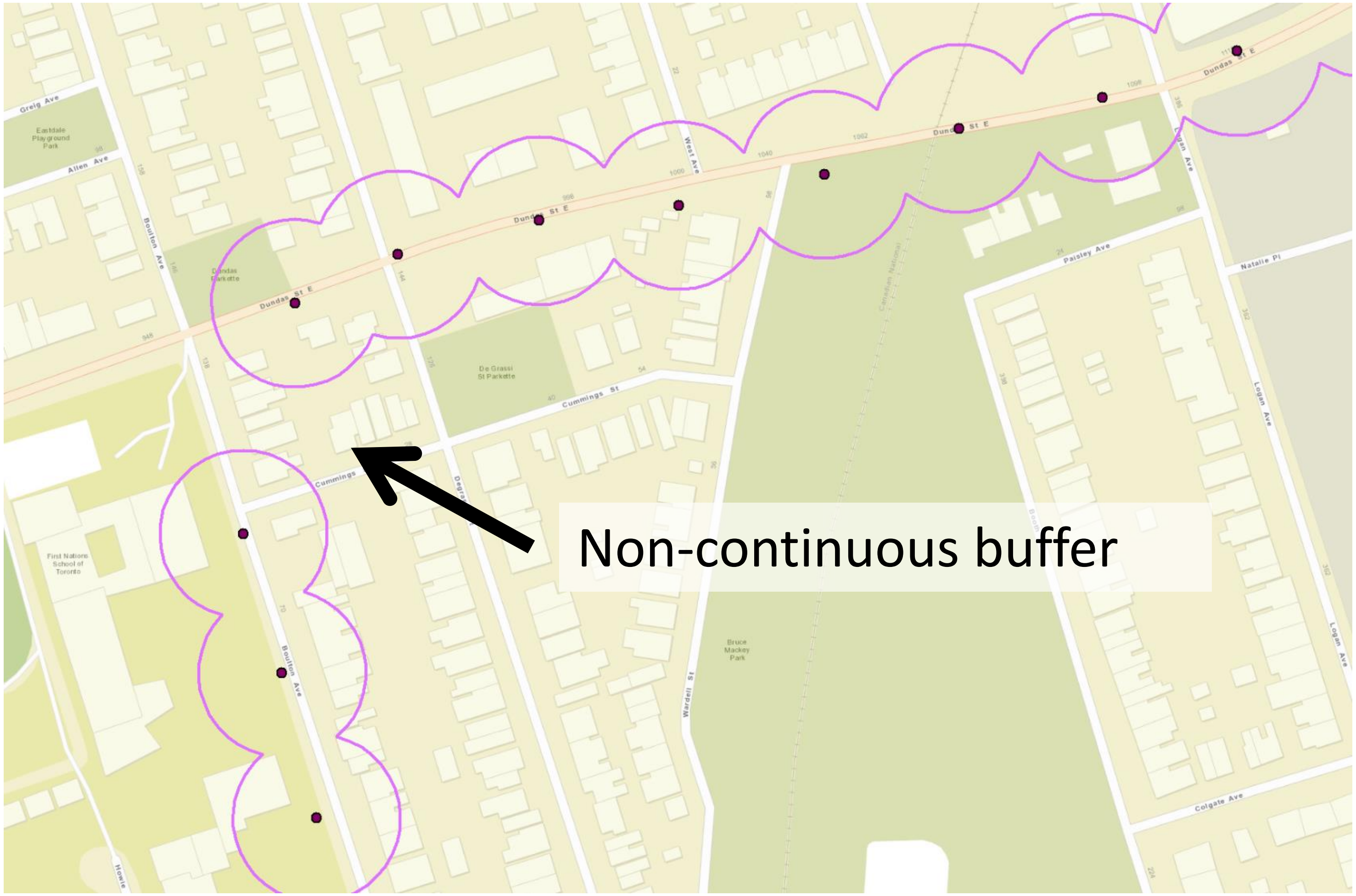
Map-Matching Issues

- Pedestrian trips can go through buildings or open spaces
- Alternate routes may exist within buffer area
- Large gaps may make buffer area not continuous
- Filling GPS points in straight line may cut corners

Walk Trip Issues



Walk Trip Issues



4. Alternative Route Generation

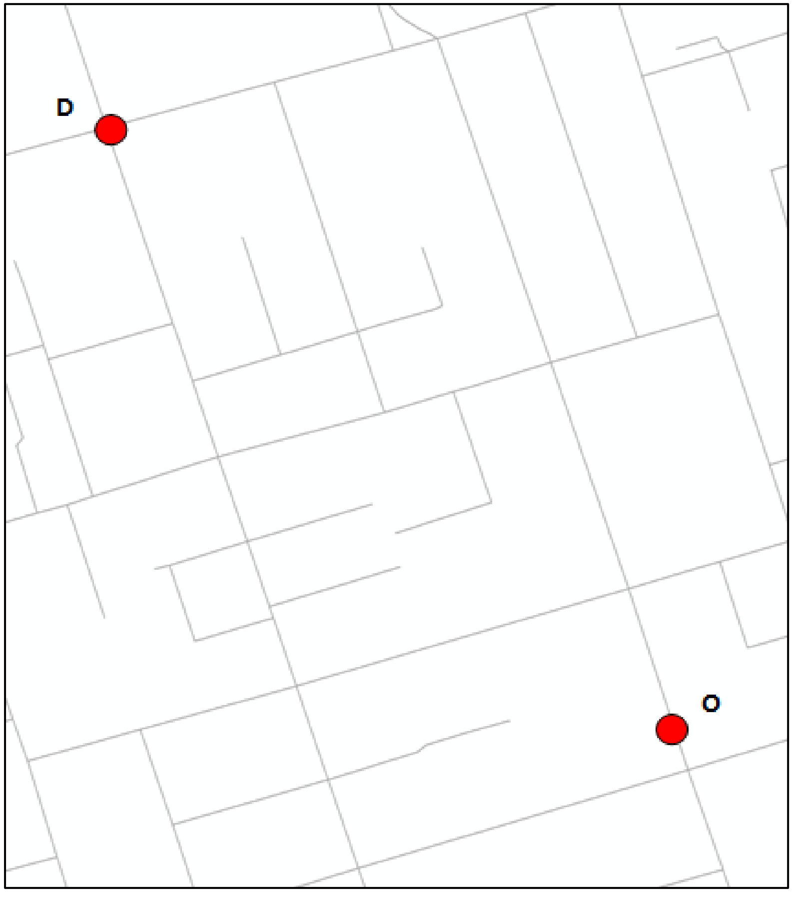
Stochastic Route Generation

- Biased random walk algorithm
- Builds the route link by link, making its way to the destination
- At each node it assesses the next links to take
- Probabilities of each branching link are determined
- Monte Carlo simulation decides which link is chosen

(Freijinger, 2007)

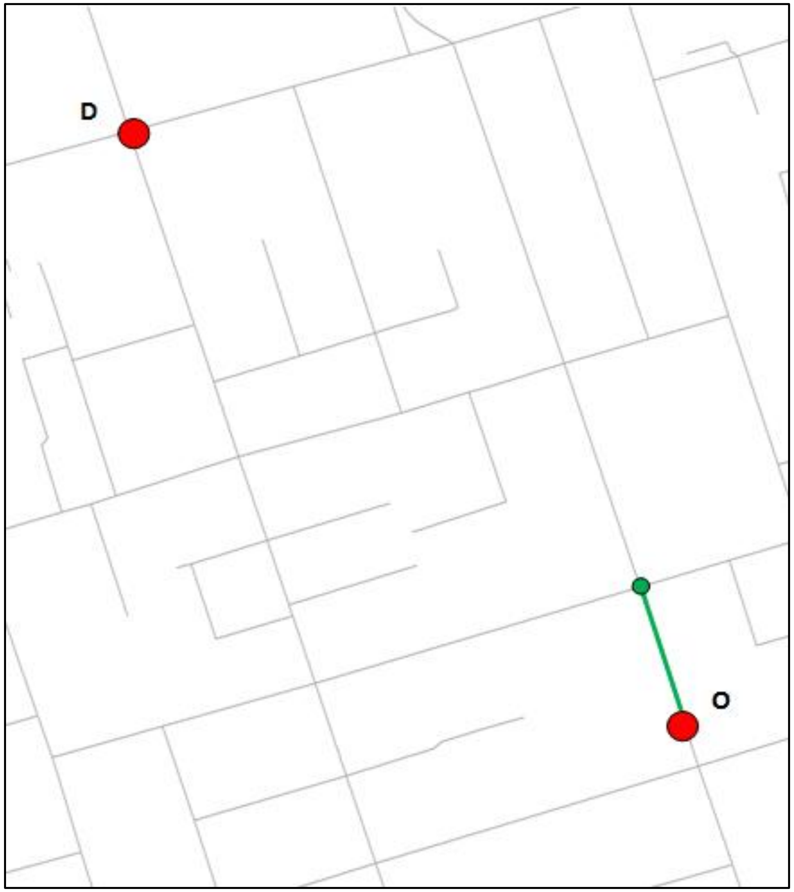
Route Generation Process

1. Import origin and destination



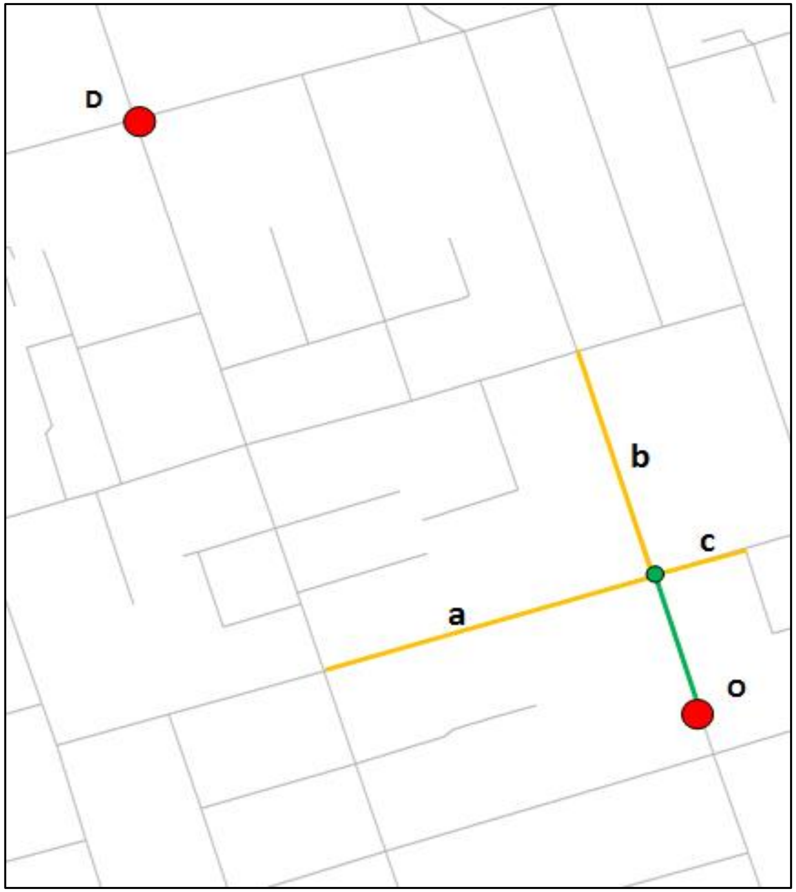
Route Generation Process

2. Determine origin street segment



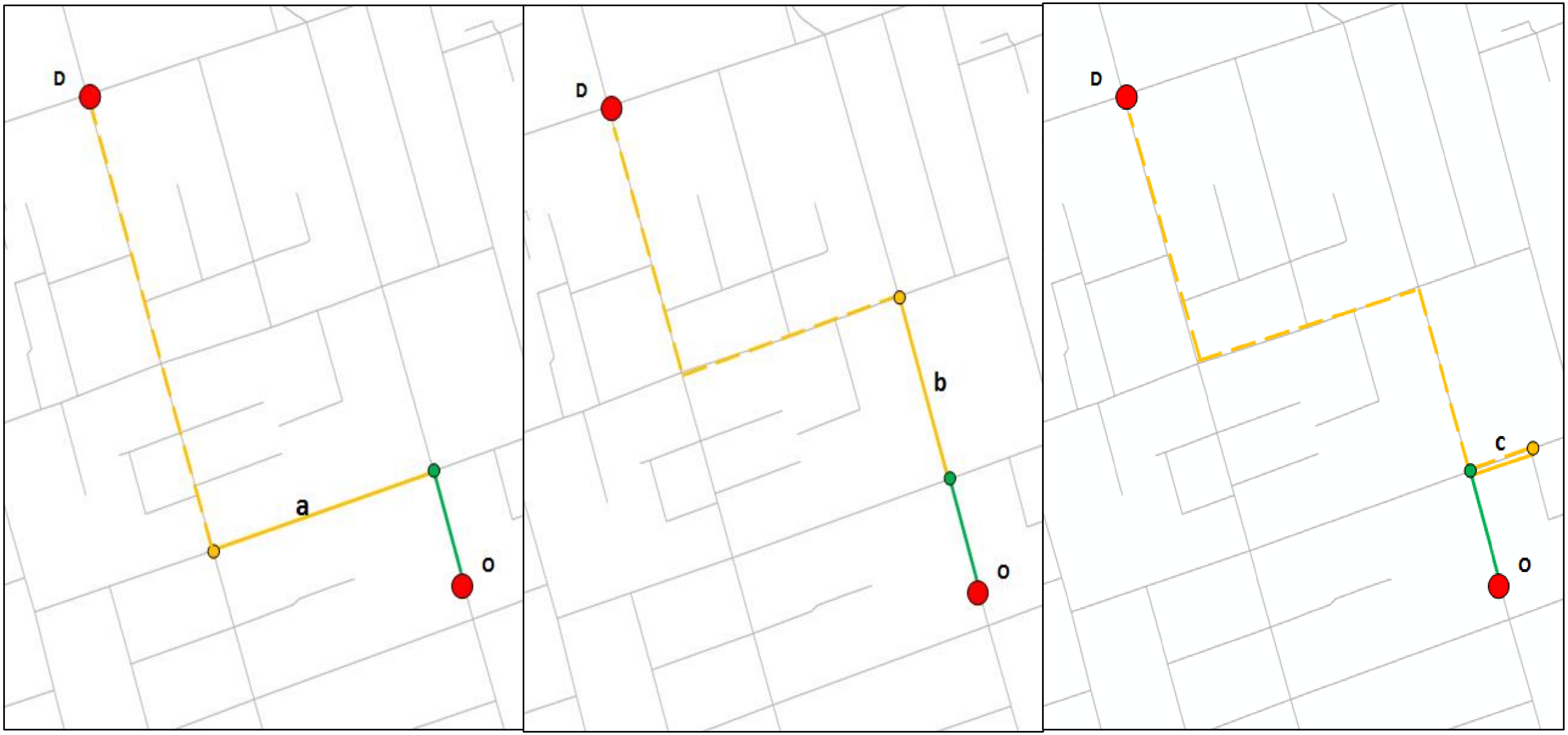
Route Generation Process

3. Find the street segments connected to the source node



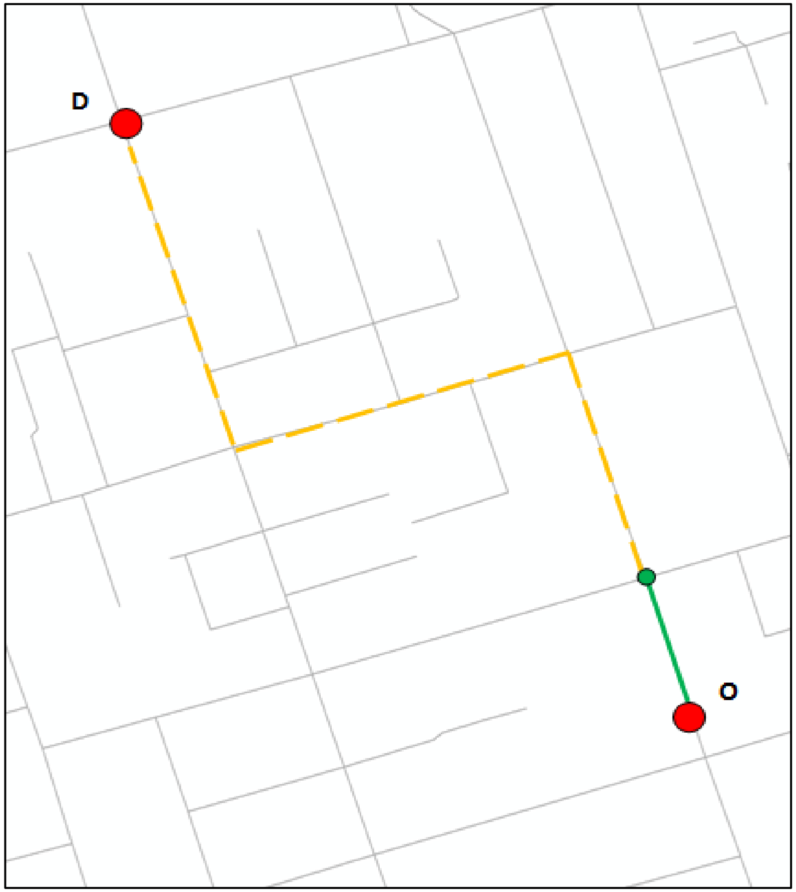
Route Generation Process

4. Determine the cost for each street segment



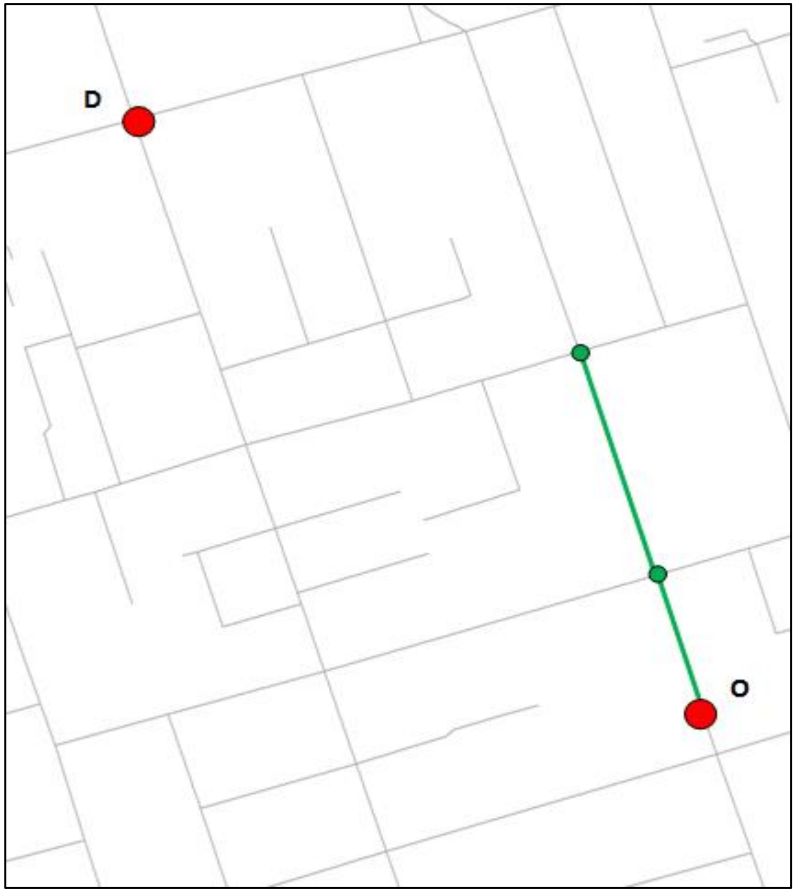
Route Generation Process

5. Determine the cost from the source node to the destination



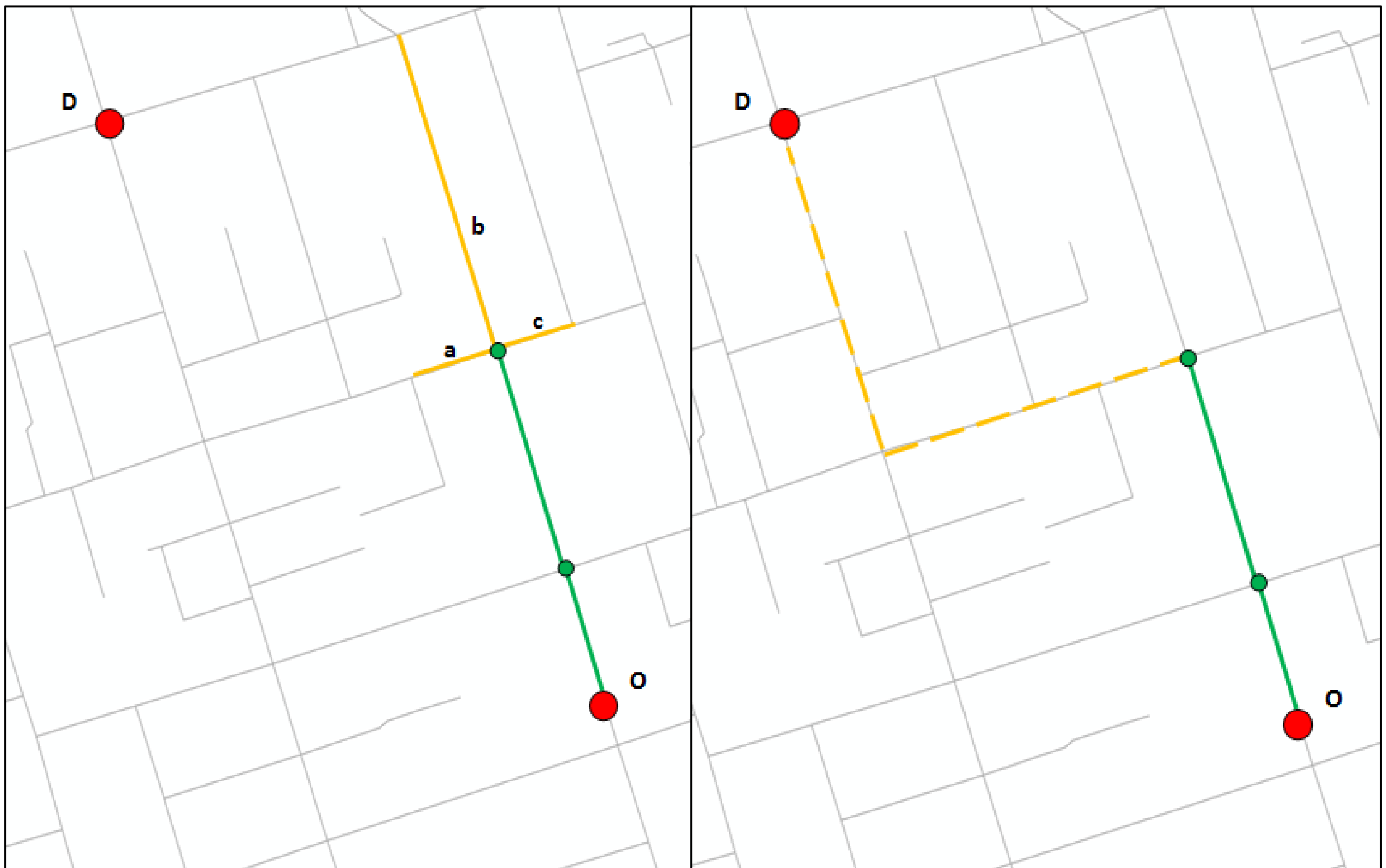
Route Generation Process

6. Calculated probabilities and use Monte Carlo simulation to select next segment



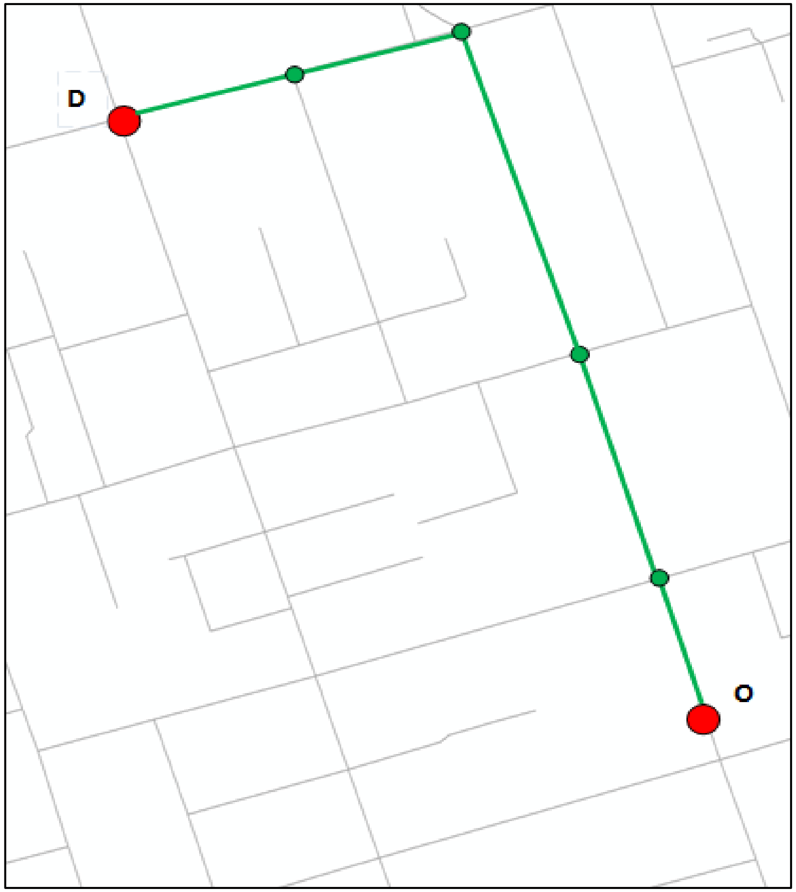
Route Generation Process

7. Repeat process for newly selected segment and source node



Route Generation Process

8. Once destination segment is reached, stop process and generate route



Route Generation Rules

$$P(i) = \frac{1 - \left(1 - \left(\frac{SP(v,D)}{cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}{\sum_{i \in M} 1 - \left(1 - \left(\frac{SP(v,D)}{cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}$$

Where:

Probability of choosing link i out of possible outgoing links (M)

Source node v and sink node w

SP(v,D) is the shortest path/least cost path from source node v to destination D

Cost(i) is the cost of link i

α and β are parameters that make the probability more sensitive to increase in cost.

(Freijinger, 2007)

Route Generation Rules

- No node is traversed twice. If a loop is detected, the route generation attempt fails.
- U-turns are not needed
- The generated path does not exceed two times the shortest path between O and D
- The route does not pass the destination link
- If a dead end is reached, the route generation attempt fails and the dead end segment is recorded so it is not considered again. After 10 attempts, the iteration is abandoned
- Travel on street segments that go in a direction away from the destination are heavily penalized (cost=9999m) unless they are on the shortest path from the source to the destination.

Route Generation Rules

- Additional Modifications
  - Turns equivalent to +50m
  - Travel on streets with complete sidewalks is 10% shorter

5. Choice Model

Path Size Logit Model

$$P(i|C_n) = \frac{e^{\mu(V_{in}+\ln(PS_{in}))+\ln\left(\frac{k_{in}}{q(i)}\right)}}{\sum_{j\in C_n} e^{\mu(V_{jn}+\ln(PS_{jn}))+\ln\left(\frac{k_{jn}}{q(j)}\right)}}$$

Where:

Cn is the choice set for user n (includes chosen route)

μ is the logit scale term

V_{in} is systematic utility for alternative i for user n

PS_{in} is the expanded path size factor for alternative i for user n

k_{in} is the number of times alternative i is randomly drawn. If chosen route, $k_{in}+1$

q(i) is the probability of choosing a route containing the street segments. It is calculated as the product of each link choice probability

(Freijinger, 2007; Frejinger, Bierlaire, and Ben-Akiva, 2009)

Path Size Logit Model

$$PS_{in} = \sum_{a \epsilon \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon C_n} \left(\frac{L_i}{L_j}\right)^{\phi} \delta_{aj}}$$

Where:

Γi is the set of links in path i

La is the length of link a

Li is the length of path i

Lj is the length of path j

δ_{aj} equals 1 if link a is on path j and 0 otherwise

ϕ is a parameter that controls the impact of route length in the correction factor

(Ramming, 2002)

6. Toronto Case Study

Route Characteristics

Observed walk trip characteristics	
Total Number of Trips	776
Number of Users	71
Average Number of Trips	9.6
Max Number of Trips per User	167
Trips by Females	28.0%
Mean Distance (m)	926.8
Travel on streets with complete sidewalks	88.8%
Travel on off-street paths	6.0%

Alternative route characteristics	
Mean Distance (m)	1000.6
Travel on streets with complete sidewalks	80.2%
Travel on off-street paths	4.2%
Average Number of Unique Alternatives	7.4

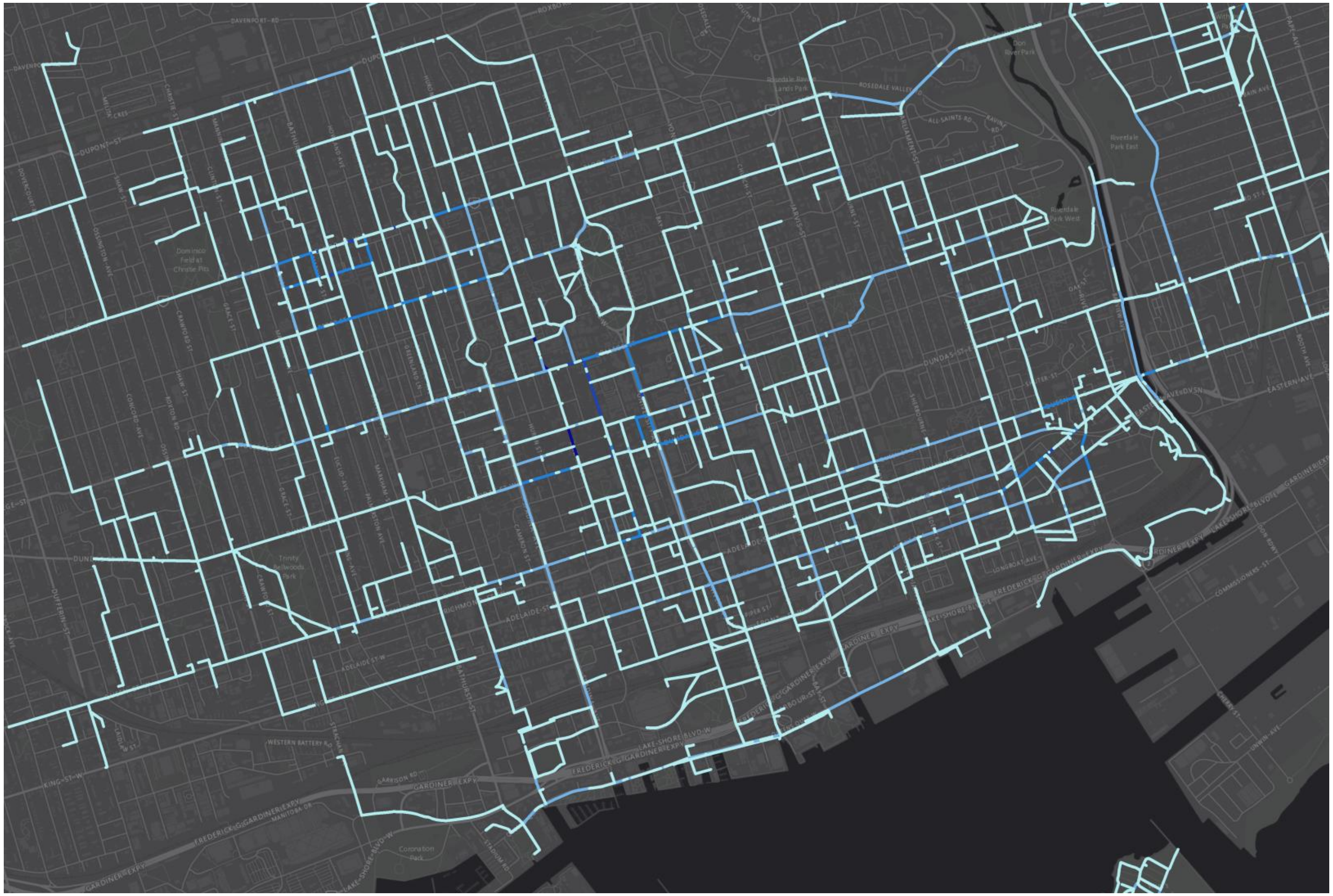
Route Variables

Name	***Description***
Length	Total route length
Turns	Total number of turns in route
Sidewalk both sides	Length of road (m) with sidewalk on both sides
Signalized Intersection	Number of signalized intersections in route
Minor arterial road	Length of route (m) on minor arterial road
Arterial Road	Length of route (m) on major or minor arterial road
Collector road	Length of route (m) on collector road
Land commercial	Length of route (m) with commercial land use frontage
Land office	Length of route (m) with office land use frontage
Land park	Length of route (m) with park land use frontage
Percent land park	Percent of route with park land use
PS	Path size correction factor
Sample correction	Probabilistic sampling correction factor
Additional variables tested	Pedestrian crossovers, steep slopes, major arterial road, local road, incomplete sidewalk, Walk Score, low residential land, high residential land, industrial land, institutional land

Socioeconomic Interaction Terms

- Gender
- Age
- Student Status
- Employment Status
- Income level
- Time of Day

Observed Trips



5. Results

General Model Results

The utility equation for route i is given be the following equation:

$$U_i = \beta_{Length} * Length_i + \beta_{Turns} * Number\ of\ Turns_i$$
$$+ \beta_{Sidewalk\ Both\ Sides} * Length\ Sidewalk\ Both\ Sides_i$$
$$+ \beta_{Siginalized\ Intersection}$$
$$* Number\ of\ Signalized\ Intersections_i + \beta_{PS} * ln(PS_i)$$
$$+ ln\left(\frac{k(i)}{q(i)}\right)$$

General Model Results

	Coefficient*
Length (m)	-0.02
Turns	-0.645
Length with sidewalk on both sides of the road	0.00665
Number of signalized intersections	0.669
ln(PS)	1.53
Log-likelihood (Null)	-1488.946
Log-likelihood (Model)	-785.99
Rho squared	0.472
N	776

* all coefficients significant at $p<0.05$

General Model Distance Trade-off

Attribute
Turn
Signalized Intersection
Sidewalk Both Sides

+33m

General Model Distance Trade-off

Attribute

Turn

Signalized Intersection

Sidewalk Both Sides

-36m

General Model Distance Trade-off

Attribute
Turn
Signalized Intersection
Sidewalk Both Sides

-33% distance

General Model Distance Trade-off

Attribute	**Distance Equivalent (m)**
Per additional..	
Turn	+32
Signalized Intersection	-34
Change in perceived distance along..	
Sidewalk both sides	-33%

Non-Significant Variables

- Land use
- Development density
- Steep slopes
- Walk Score

Interaction Model Results

The interaction term model's utility equation for route i is given by the following equation:

$$U_i = \beta_{Length} * Length_i + \beta_{Length\ Female} * Length_i * Gender_{Female} + \beta_{Turns} * Number\ of\ Turns_i$$
$$+ \beta_{Sidewalk\ Both\ Sides} * Length\ Sidewalk\ Both\ Sides_i + \beta_{Siginalized\ Intersection}$$
$$* Number\ of\ Signalized\ Intersections_i + \beta_{Minor\ Arterial\ Student}$$
$$* Length\ on\ Minor\ Arterial_i * Student + \beta_{Arterial\ Age\ 25} * Length\ on\ Arterial_i$$
$$* Age\ under\ 25 + \beta_{Minor\ Arterial\ Income} * Length\ on\ Minor\ Arterial_i * Income\ over\ \$75{,}000$$
$$+ \beta_{Parks\ Evening} * Length\ in\ Park_i * Evening + \beta_{Commercial\ Employed}$$
$$* Length\ along\ Commercial\ Land_i * Employed + \beta_{Collecter\ Age\ 45} * Length\ on\ Collector_i$$
$$* Age\ over\ 45 + \beta_{Office\ Age\ 25} * Length\ along\ Office\ Land_i * Age\ under\ 25$$
$$+ \beta_{Walkway\ Age\ 25} * Length\ along\ Walkways_i * Age\ under\ 25 + \beta_{PS} * \ln(PS_i) + \ln\left(\frac{k(i)}{q(i)}\right)$$

Interaction Model Results

	Coefficient*
Length (m)	-0.0198
Length (m) x Female	-0.00788
Turns	-0.724
Length with sidewalk on both sides of the road	0.0073
Number of signalized intersections	0.729
Length on minor arterial roads as a student	-0.00333
Length on major or minor arterial roads when under age 25	0.00337
Length on minor arterial roads when income >$75,000/yr	0.0029
Length along parks after 4PM	-0.0214
Length along commercial land use when employed	-0.00911
Length on collector roads when over age 45	-0.00319
Length along office land use when under age 25	-0.0143
Length along walkways when under age 25	0.0145
ln(PS)	1.39
Log-likelihood (Null)	-1488.946
Log-likelihood (Model)	-742.723
Rho squared	0.501
N	776

* all coefficients significant at p<0.05

Interaction Model Distance Trade-off

Attribute	Distance Equivalent	
	Male	Female
Per additional..		
Turn	+37	+26
Signalized Intersection	-37	-26
Change in perceived distance along..		
Sidewalk both sides	-37%	-26%
Minor arterial as a student	17%	12%
Arterial road as a person under 25	-17%	-12%
Minor arterial road as a person with income over $75k/yr	-15%	-10%
Collector road as a person over 45	+16%	+12%
Park land use after 4 PM	+108%	+77%
Commercial land use as a employed person (full or part-time)	+46%	+33%
Office land use as a person under 25	+72%	+52%
Walkway land use as a person under 25	-73%	-52%

Interaction Model Distance Trade-off

Attribute	Distance Equivalent	
	Male	Female
Per additional..		
Turn	+37	+26
Signalized Intersection	-37	-26
Change in perceived distance along..		
Sidewalk both sides	-37%	-26%
Minor arterial as a student	17%	12%
Arterial road as a person under 25	-17%	-12%
Minor arterial road as a person with income over $75k/yr	-15%	-10%
Collector road as a person over 45	+16%	+12%
Park land use after 4 PM	+108%	+77%
Commercial land use as a employed person (full or part-time)	+46%	+33%
Office land use as a person under 25	+72%	+52%
Walkway land use as a person under 25	-73%	-52%

Interaction Model Distance Trade-off

Attribute	Distance Equivalent	
	Male	Female
Per additional..		
Turn	+37	+26
Signalized Intersection	-37	-26
Change in perceived distance along..		
Sidewalk both sides	-37%	-26%
Minor arterial as a student	17%	12%
Arterial road as a person under 25	-17%	-12%
Minor arterial road as a person with income over $75k/yr	-15%	-10%
Collector road as a person over 45	+16%	+12%
Park land use after 4 PM	+108%	+77%
Commercial land use as a employed person (full or part-time)	+46%	+33%
Office land use as a person under 25	+72%	+52%
Walkway land use as a person under 25	-73%	-52%

Interaction Model Distance Trade-off

Attribute	Distance Equivalent	
	Male	Female
Per additional..		
Turn	+37	+26
Signalized Intersection	-37	-26
Change in perceived distance along..		
Sidewalk both sides	-37%	-26%
Minor arterial as a student	17%	12%
Arterial road as a person under 25	-17%	-12%
Minor arterial road as a person with income over $75k/yr	-15%	-10%
Collector road as a person over 45	+16%	+12%
Park land use after 4 PM	+108%	+77%
Commercial land use as a employed person (full or part-time)	+46%	+33%
Office land use as a person under 25	+72%	+52%
Walkway land use as a person under 25	-73%	-52%

Interaction Model Distance Trade-off

Attribute	**Distance Equivalent**	
	Male	**Female**
Per additional..		
Turn	+37	+26
Signalized Intersection	-37	-26
Change in perceived distance along..		
Sidewalk both sides	-37%	-26%
Minor arterial as a student	17%	12%
Arterial road as a person under 25	-17%	-12%
Minor arterial road as a person with income over $75k/yr	-15%	-10%
Collector road as a person over 45	+16%	+12%
Park land use after 4 PM	+108%	+77%
Commercial land use as a employed person (full or part-time)	+46%	+33%
Office land use as a person under 25	+72%	+52%
Walkway land use as a person under 25	-73%	-52%

6. Route Generation Analysis

Route Generation Analysis

Generation scenario	Average probability of drawing observed route	Probability of drawing observed route at least once
Biased around shortest path	21.3%	53.7%
Biased around least cost	20.8%	52.1%
Biased around calibrated least cost	21.2%	51.9%

Route Generation Analysis

	Value	
Number of trips where…		
Least cost probability >= shortest path probability	584	75%
Calibrated least cost probability >= shortest path probability	572	74%
Calibrated least cost probability >= least cost probability	578	74%
Average route length where…		
Least cost probability >= shortest path probability	960.2	
Least cost probability < shortest path probability	810.1	
Calibrated least cost probability >= shortest path probability	991.9	
Calibrated least cost probability < shortest path probability	730.0	
Calibrated least cost probability >= least cost probability	985.7	
Calibrated least cost probability < least cost probability	740.2	

Route Generation Analysis

	Value	
Number of trips where…		
Least cost probability >= shortest path probability	584	75%
Calibrated least cost probability >= shortest path probability	572	74%
Calibrated least cost probability >= least cost probability	578	74%
Average route length where…		
Least cost probability >= shortest path probability	960.2	
Least cost probability < shortest path probability	810.1	
Calibrated least cost probability >= shortest path probability	991.9	
Calibrated least cost probability < shortest path probability	730.0	
Calibrated least cost probability >= least cost probability	985.7	
Calibrated least cost probability < least cost probability	740.2	

Route Generation Analysis

Table - Average percent difference compared to observed route

	Length	Turn	Signalized Intersection	Sidewalk Both
Shortest	6.6%	-1.37	-0.29	13.3%
General Cost	6.9%	-1.33	-0.26	14.5%
Calibrated Cost	7.6%	1.27	-0.01	29.9%

Route Generation Analysis

Table - Average percent difference compared to observed route

	Length	Turn	Signalized Intersection	Sidewalk Both
Shortest	6.6%	-1.37	-0.29	13.3%
General Cost	6.9%	-1.33	-0.26	14.5%
Calibrated Cost	7.6%	1.27	-0.01	29.9%

Calibrated route generation method had longer routes, routes with more turns, and more travel on streets with complete sidewalks

Route Generation Analysis

- Route generation biased around shortest path had the highest probability of generating observed route
- Calibrated route generation methods were more likely to generate observed route for longer routes

7. Conclusions

Conclusions

- Smartphone GPS data proved to be viable source for pedestrian route choice
- Distance, turns, complete sidewalk, and signalized intersections are significant factors
- Calibrating the stochastic route choice generator made generating the observed route more likely for longer routes

Limitations/Future Work

- GPS accuracy was too low to determine detailed trip behaviour
- Trip purpose not collected
- Stochastic route choice generation works well but may generate very random routes
- Multiple observations influence results
- Land use measure could be improved with observational data

8. References

References

City of Toronto. (2016). Built Environment - Environmental Health - Toronto Public Health | City of Toronto. Retrieved January 28, 2017, from http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=d06e23bf6d481410VgnVCM10000071d60f89RCRD

Dalumpines, R., & Scott, D. M. (2011). GIS-based Map-matching: Development and Demonstration of a Postprocessing Map-matching Algorithm for Transportation Research In *ResearchGate* (pp. 101–120). https://doi.org/10.1007/978-3-642-19789-5_6

Frejinger, E. (2007). Random Sampling of Alternatives in a Route Choice Context. *ResearchGate*. Retrieved from https://www.researchgate.net/publication/37456566_Random_Sampling_of_Alternatives_in_a_Route_Choice_Context

Frejinger, E., & Bierlaire, M. (2009). Route Choice Analysis: Data, Models, Algorithms and Applications. In *The 12th International Conference on Travel Behaviour Research*. Retrieved from https://infoscience.epfl.ch/record/152428/files/IATBR09_EMMA.pdf

Frejinger, E., Bierlaire, M., & Ben-Akiva, M. (2009). Expanded Path Size attribute for route choice models including sampling correction. In *International Choice Modelling Conference 2009*. Retrieved from http://www.icmconference.org.uk/index.php/icmc/icmc2009/paper/view/52

Guo, Z. (2009). Does the pedestrian environment affect the utility of walking? A case of path choice in downtown Boston. *Transportation Research Part D: Transport and Environment*, *14*(5), 343–352.

Harding, C., Zhang, Y., & Miller, E. J. (2015). Multiple purpose tours and efficient trip chaining in Toronto: an analysis of the effects of land use and transit provision on mode choice and trip chaining using smartphone data. In *ResearchGate*. Retrieved from https://www.researchgate.net/publication/281455393_Multiple_purpose_tours_and_efficient_trip_chaining_in_Toronto_an_analysis_of_the_effects_of_land_use_and_transit_provision_on_mode_choice_and_trip_chaining_using_smartphone_data