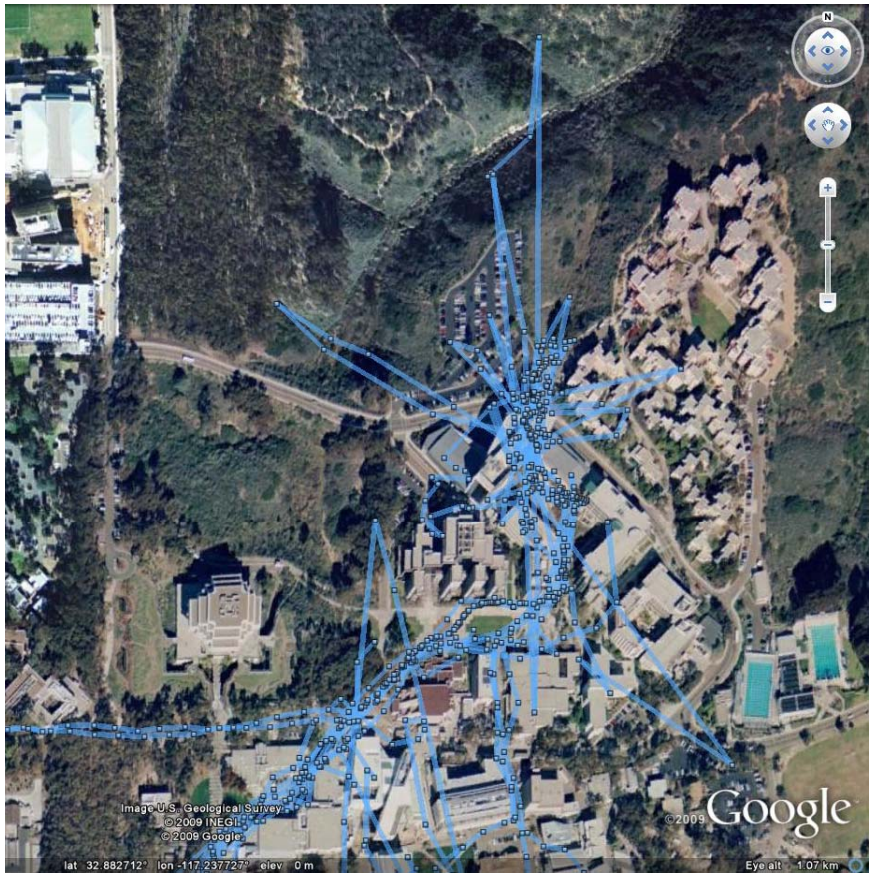
Data processing & cleaning

- Cleaning in physical activity studies
  - Often limited wear time so not needed
  - Not well reported
- Processing in Transportation studies
  - More sophisticated
  - Use GIS data
  - Mostly car based

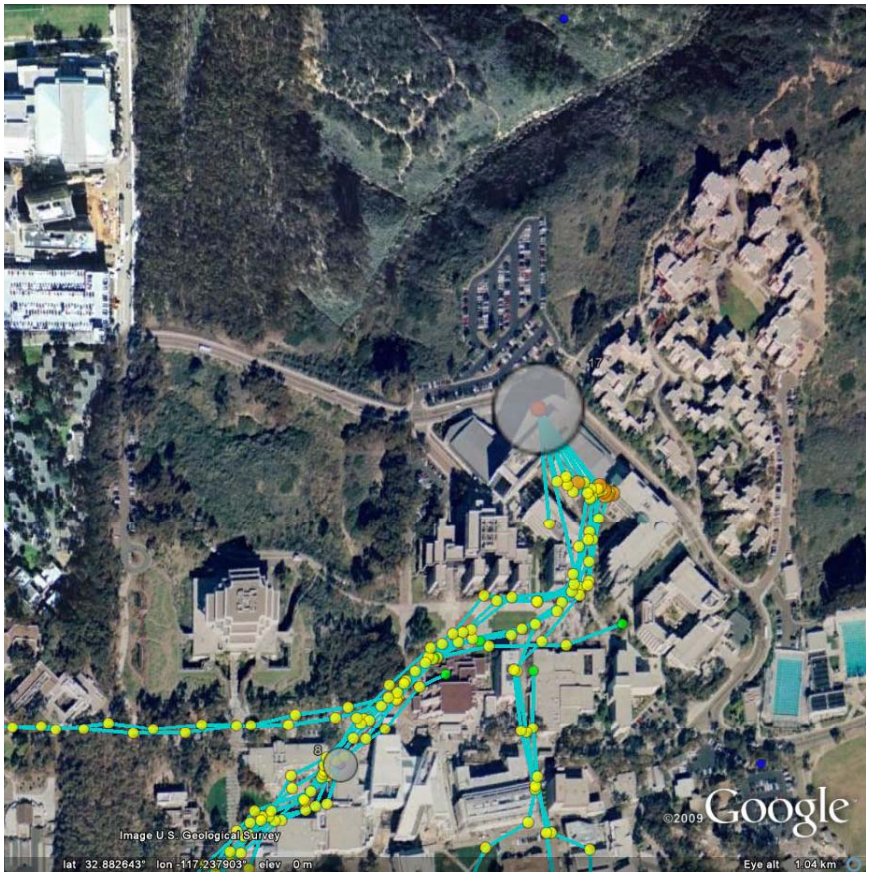
PALMS

- Physical Activity Location Measurement System
- Data processing is key (& expensive)
  - Want to avoid reinventing wheel
- Need a system that is flexible & user friendly
- Support a community of users & developers
- **http://ucsd-palms-project.wikispaces.com**

GPS Data Processing



Before – Multi-story buildings generate noise



After – Noise removed
Yellow = walking Orange = paused
Gray circles = relative time at location

GPS Variables

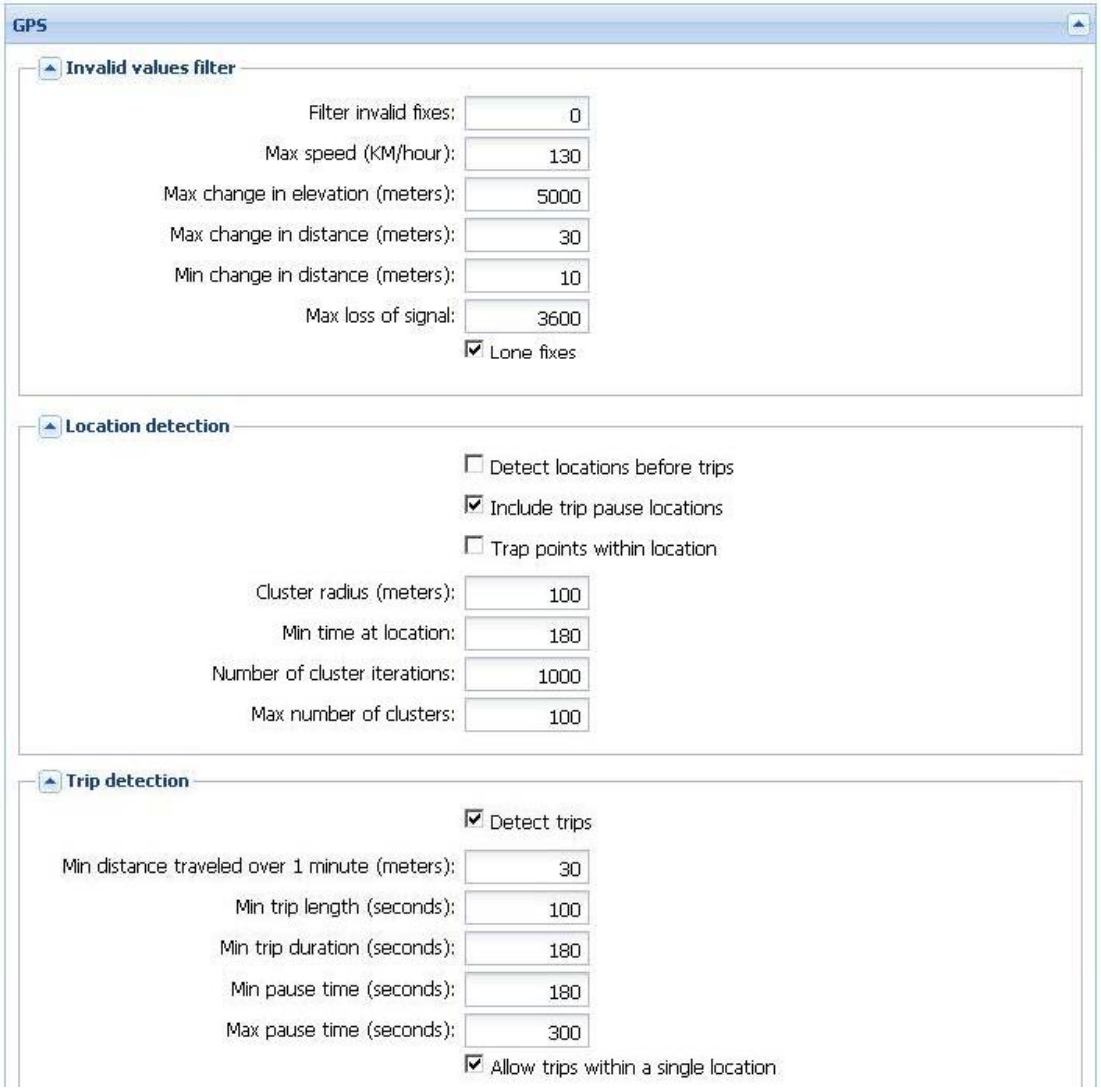
Inputs	Outputs
Date / Time	Date / Time, Duration
Latitude / Longitude	Latitude / Longitude Distance traveled, Direction (bearing)
Elevation	Elevation, Elevation delta, Grade
Satellite info (optional)	Indoors or outdoors / Fix type
Derived by algorithms	Trips: Start, end, pauses Duration, Length, Speed (Avg, Max) Estimated mode of transportation
	Locations: Time spent at location, Time away from location, Trips to/from location Number of visits to location, frequency

PALMS' Multi-step Approach to Processing

1) Pre-process trackpoints
2) Detect and mark as indoor or outdoor
3) Filter invalid and duplicate trackpoints
4) Detect trips
5) Find location clusters & mark locations
6) Reconsider trips
7) Classify and number trips
8) Capture trackpoints at locations (if desired)
9) Average speed and elevation (if desired)
10) Align to desired interval

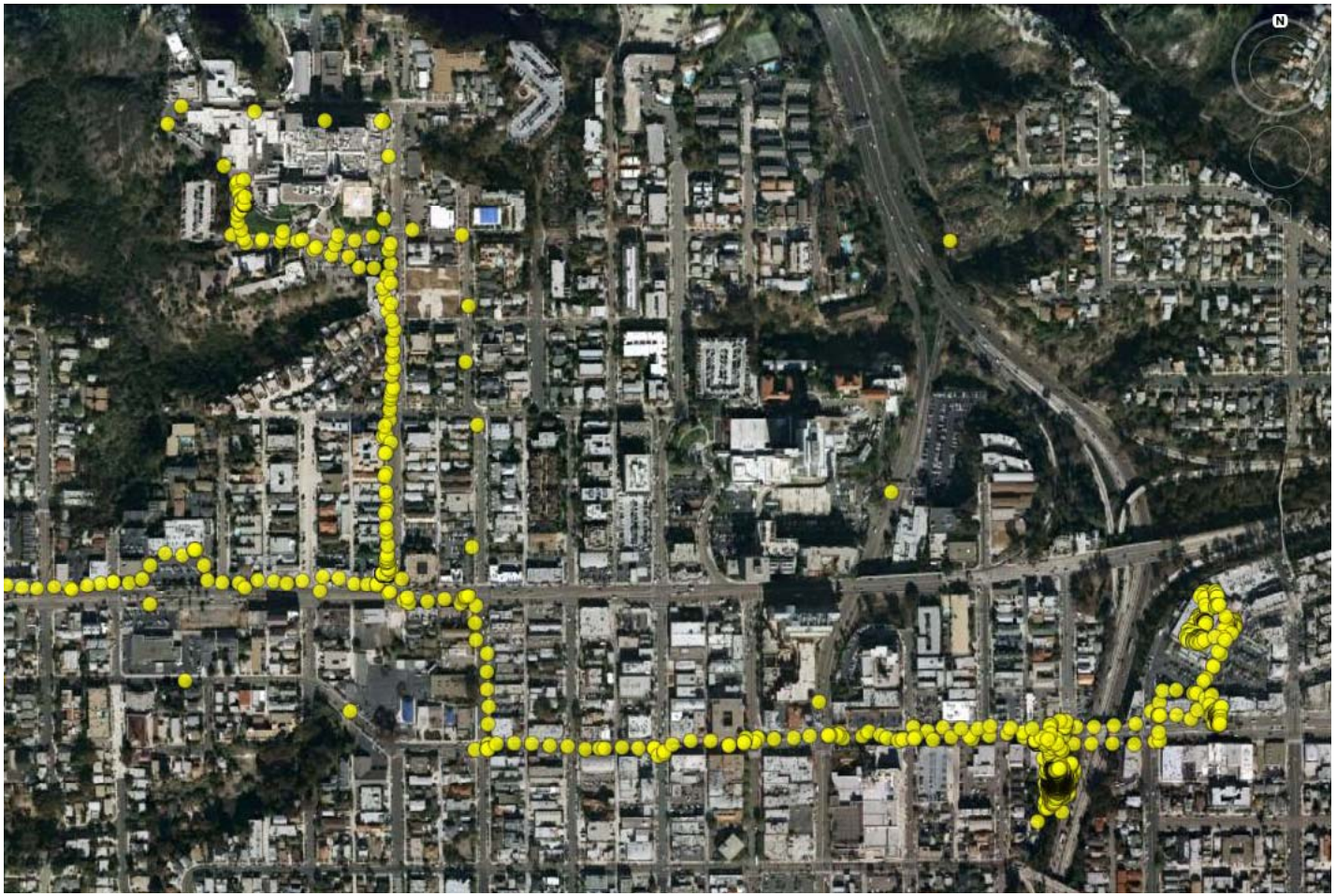
GPS Processing: Parameters, Cutoffs & Options

- All processing steps have parameters.
- In PALMS, all default parameters are visible and adjustable.
- Because of the multiple steps, parameter values of the earlier steps impact the results of downstream processing.



GPS Processing in Action

Raw Data



Example:

one subject's neighborhood

collected over 3 days

15 second epoch

GPS Processing in Action

Indoor / Outdoor Detection

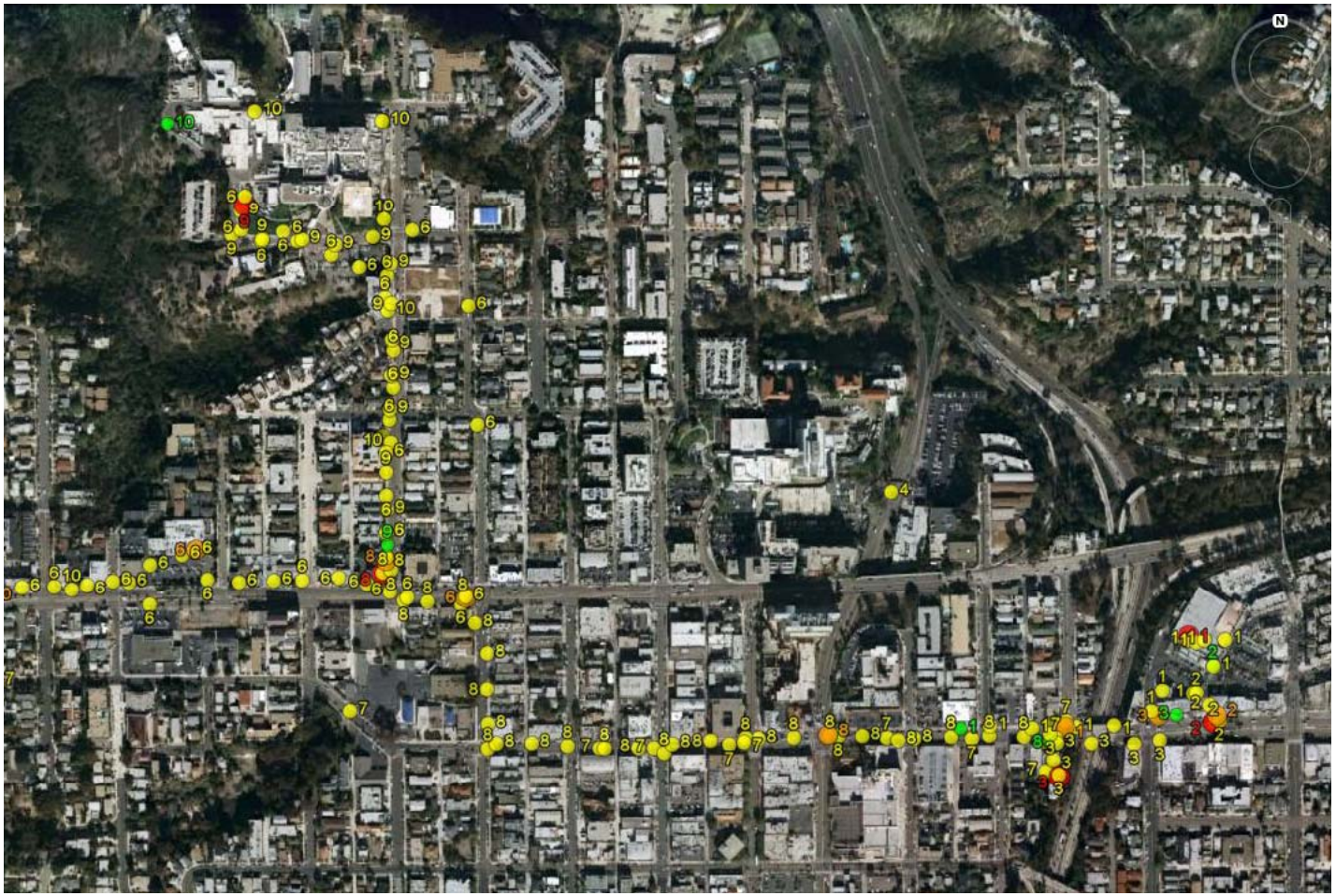
30 second epoch

Indoors

Outdoors

GPS Processing in Action

Trip Detection



Starting point

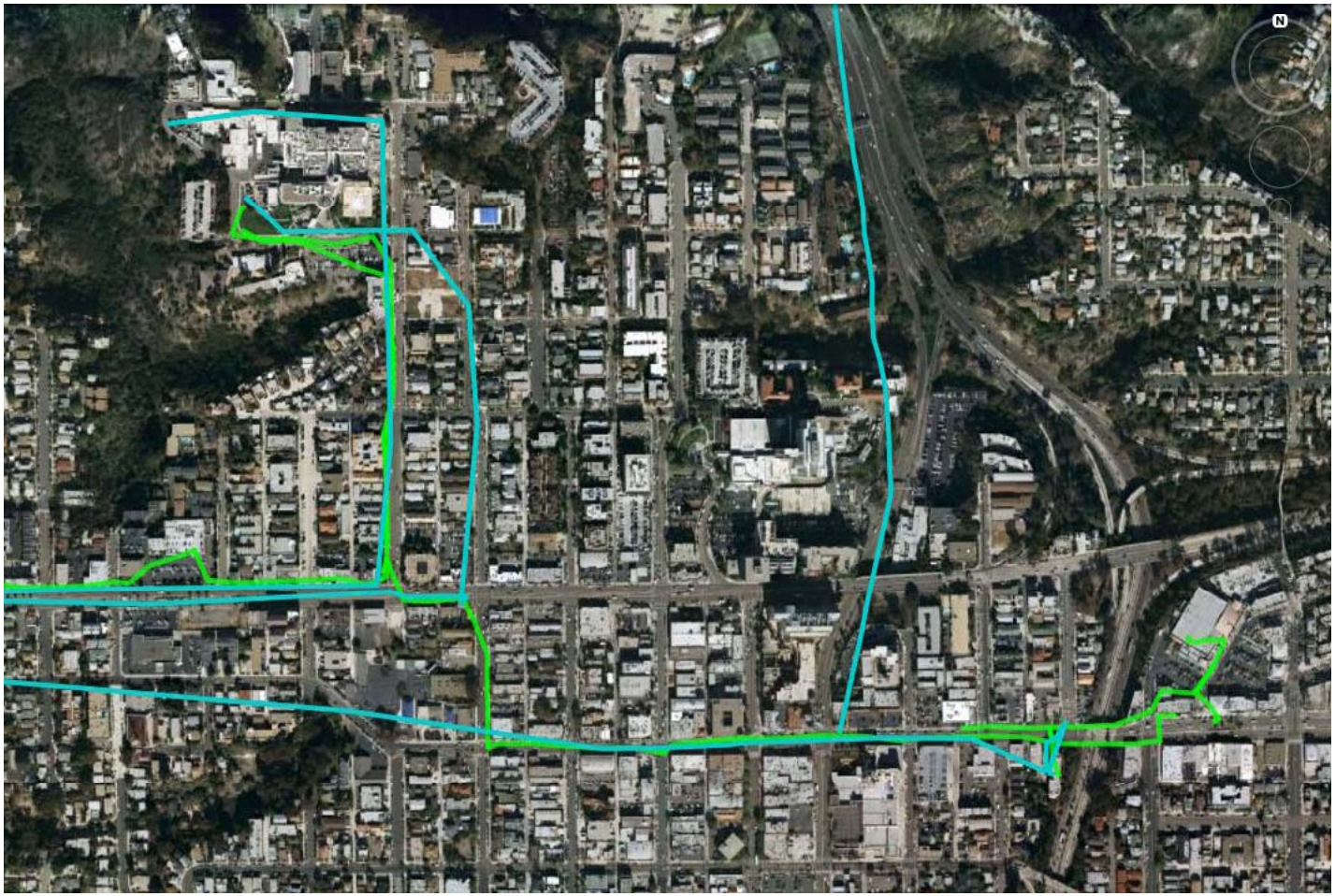
Ending point

In motion

Paused

GPS Processing in Action

Mode of Transportation



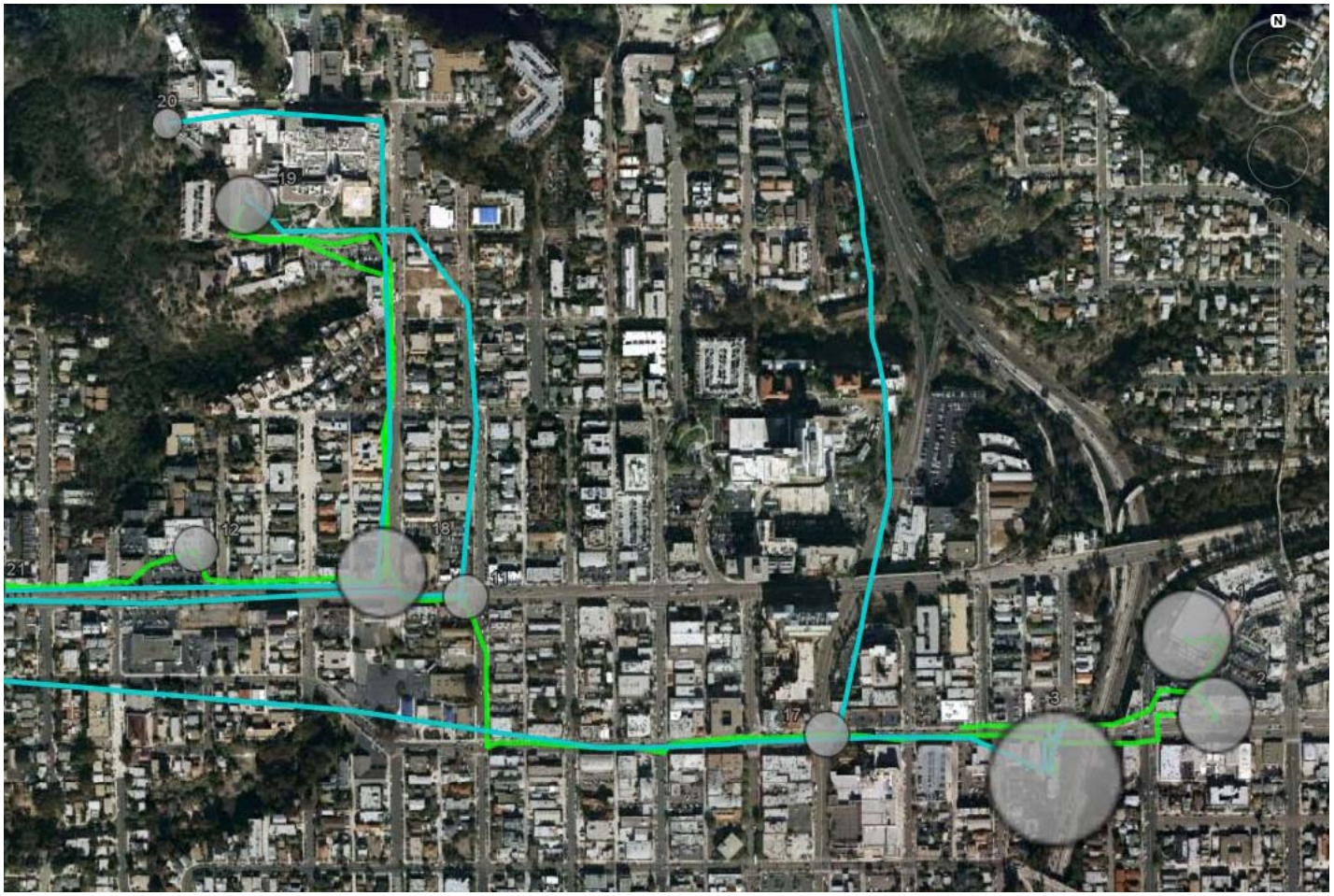
Vehicle

Pedestrian

Bicycle

GPS Processing in Action

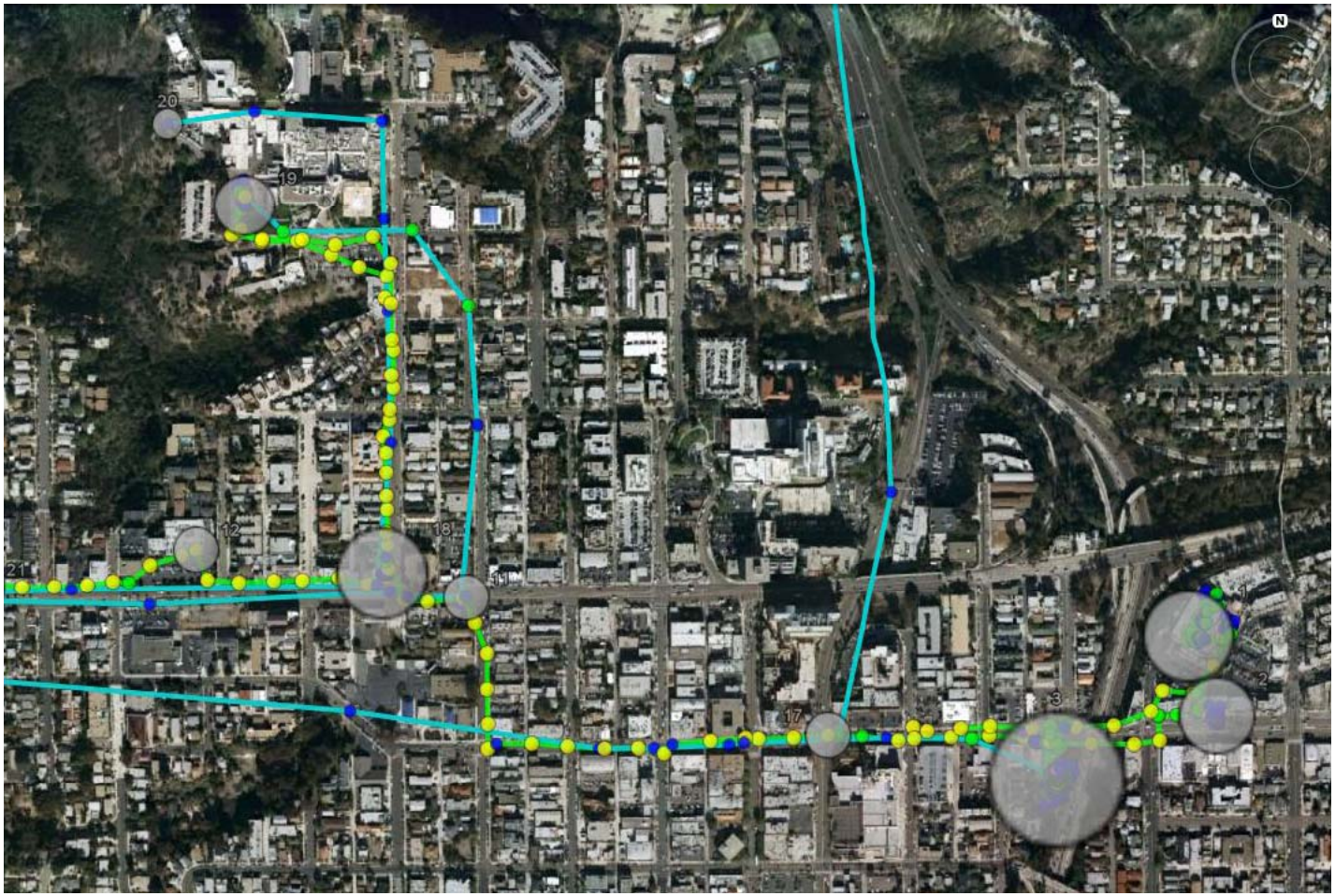
Location Detection



Size relative to amount of time at location

GPS Processing in Action

Combine with Accelerometer Data



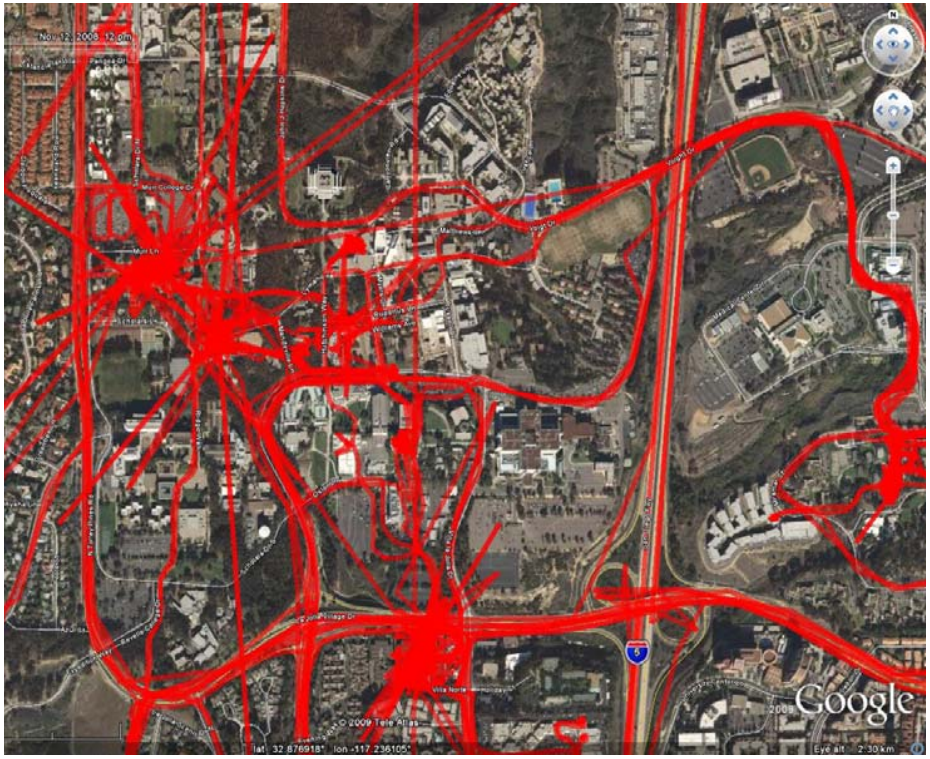
Sedentary

Light

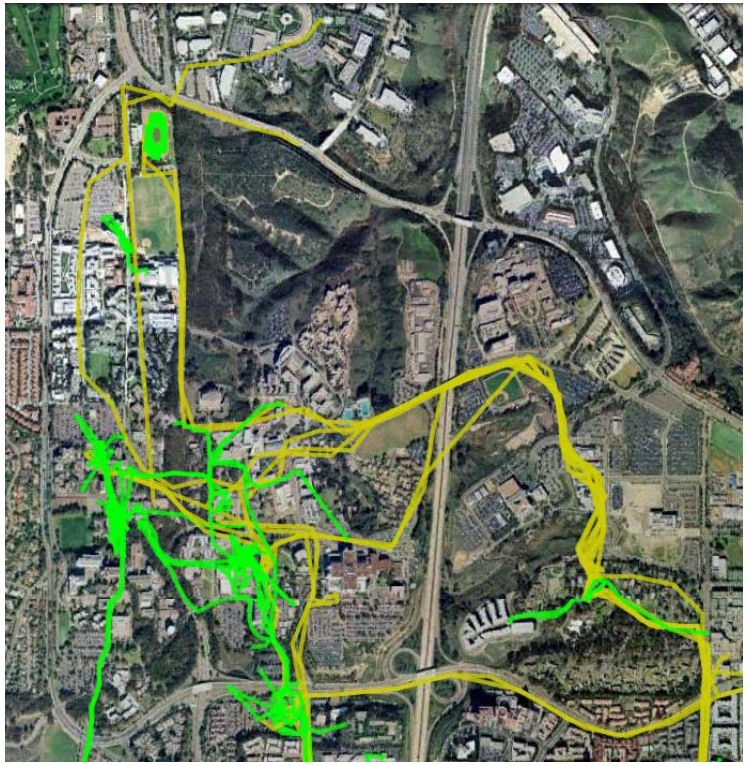
Moderate

GPS Processing in Action

Identifying walking and bicycling trips on campus



Before: All trackpoints

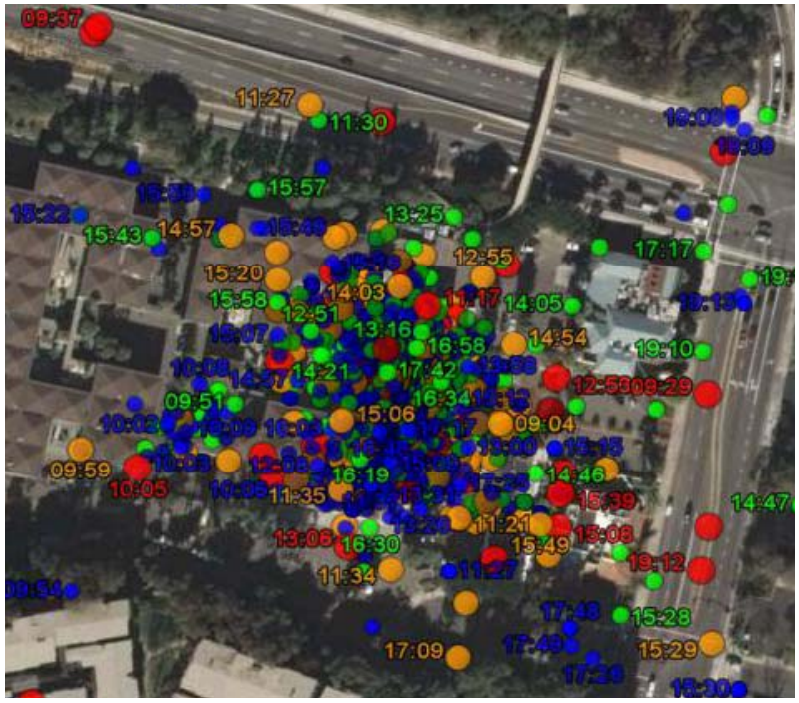


After: **Walking** / **Bicycling**

Any questions, all questions are good!

GPS Filtering (optional)

- Refine grossly invalid trackpoints (bad fixes)
- Refine redundant trackpoints (non-movement)
- Refine extraneous data points (jitter)



Before

After

Color coded by speed

GPS Filtering Algorithms

- Filter lone-fixes – do not filter first and last fixes
- Determine if trackpoint is valid
  - Check for excessive speed (> X)
  - Check for excessive change in elevation
  - Check for excessive distance traveled between trackpoints
- Determine if trackpoint is redundant
  - Check for minimum change in distance between trackpoints (redundant)
  - Check for forward / backwards movement (jitter)
- If invalid or redundant, delete trackpoint from vector and update derived values in adjacent trackpoint

GPS Indoor/Outdoor Detection

GPS Indoor/Outdoor Detection

- Qstarz GPS provides
  - Number of satellites in view
  - Number of satellites used to compute location
  - "Signal strength" of each satellite
- Outdoors = stronger signals and more satellites
  - Mark outdoors if total signal strength > threshold
  - Mark outdoors if ratio of SatUsed to SatView > threshold
- Problems to be addressed:
  - Outdoor points can be marked as indoors when traveling through area of poor coverage
  - Indoor points can be marked as outdoors if threshold values are not set correctly (typically occurs under wooden roofs)
  - In-vehicle points typically marked as indoors

GPS Trip Detection

- Detects starting and stopping locations of trips
- Detects short pauses during trips
- Note: start point variations due to time to acquire first fix.

Starting point

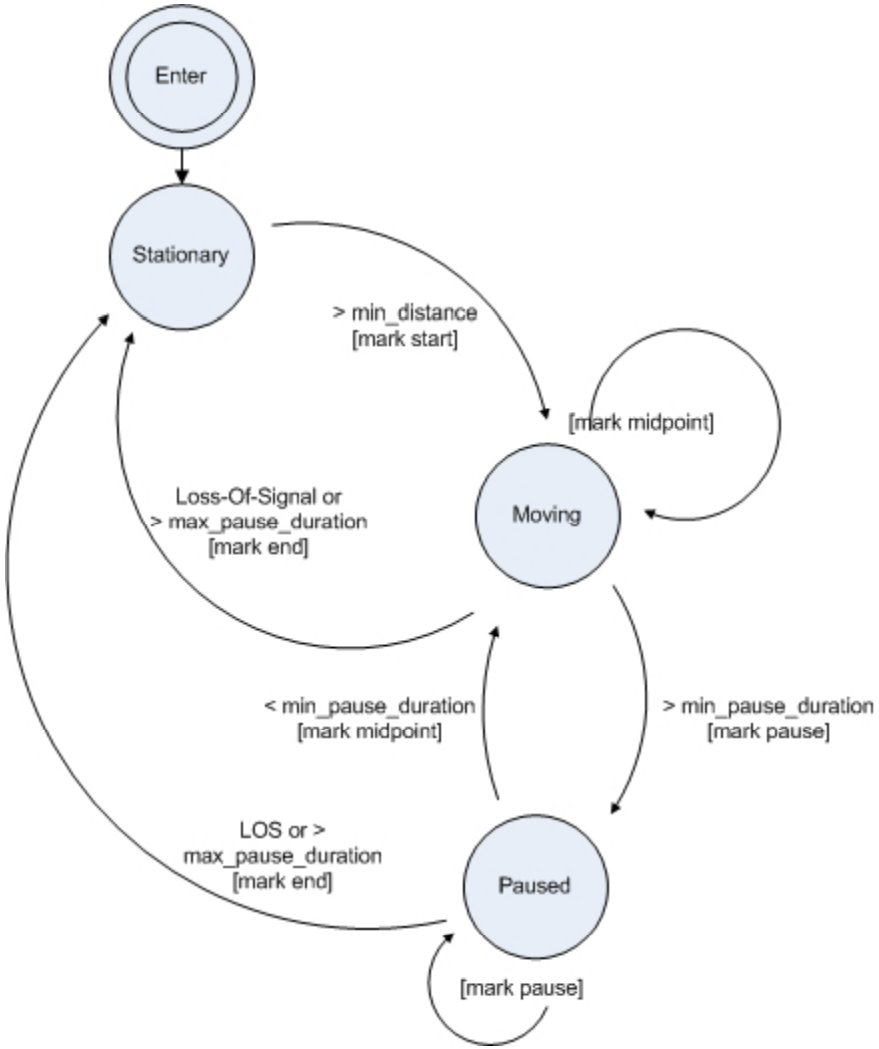
Ending point

In motion

Paused

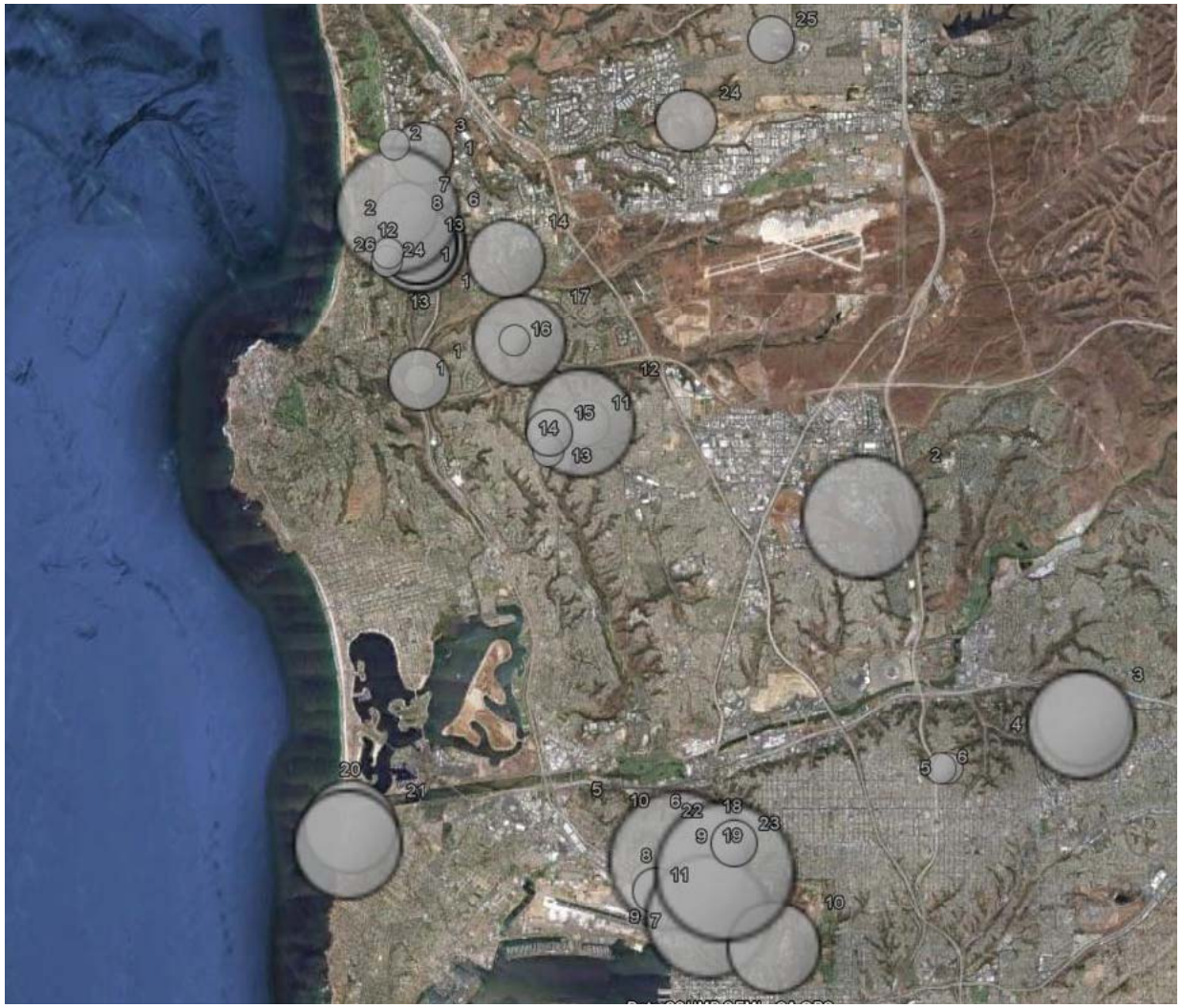
GPS Trip Detection

- Marks Trackpoints as either:
  - Stationary
  - Start points
  - Mid points
  - Pause points
  - End points
- Start point marked when distance traveled > threshold
- End points marked on loss of signal or when duration at point exceeds a time threshold
- Pause points when distance travel < threshold & duration at point within a time threshold

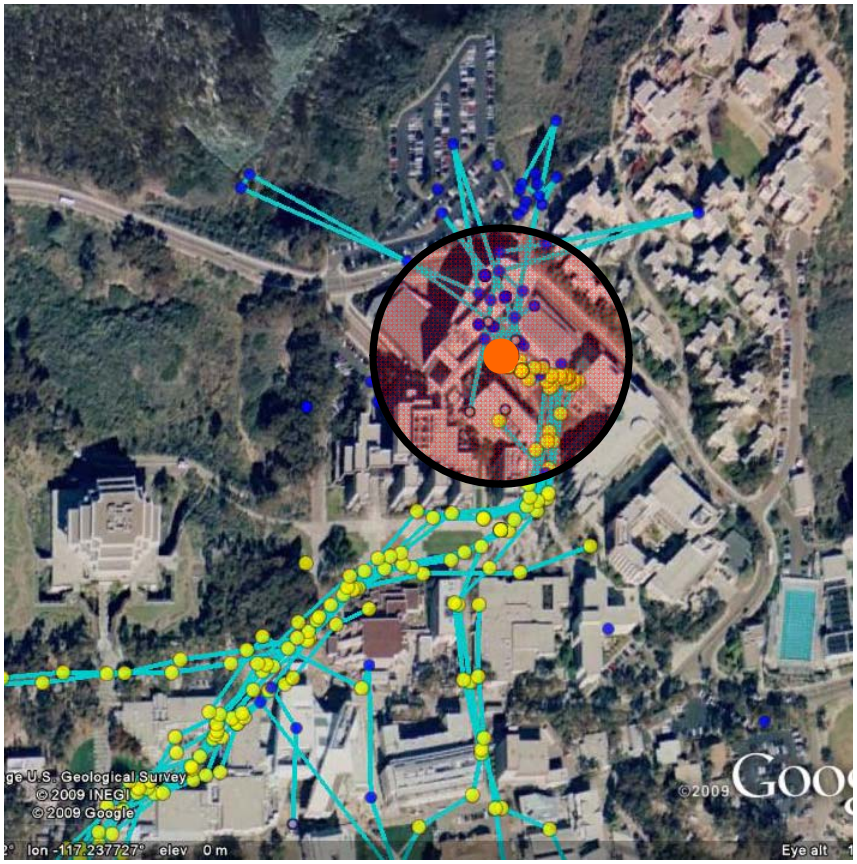


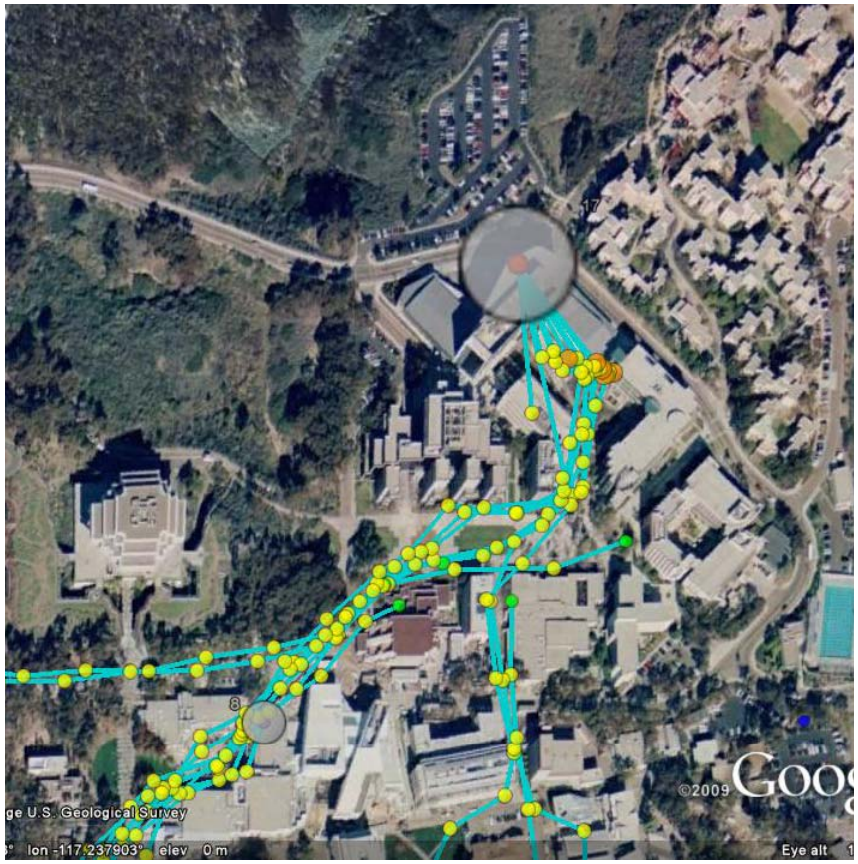
GPS Location Detection

- Detect clusters of activity, time spent, most visited, etc.



Location Capture (optional)





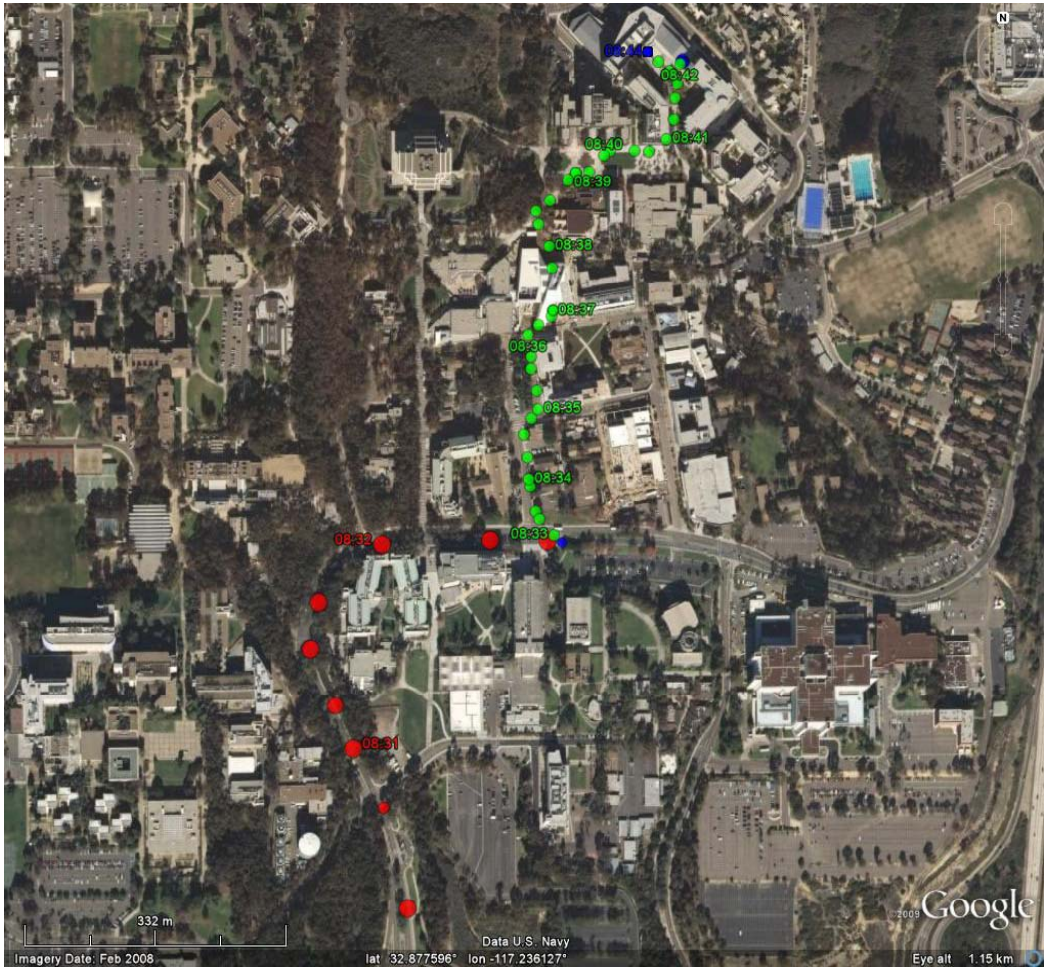
- Map all points (not part of a trip) within a radius to the location center.
- Map indoor points to the location center.

GPS Trip Post Processing

- Reconsider trips
  - Remove trips where total distance < threshold
  - Remove trips where total duration < threshold
  - Remove trips contained within one location
  - Remove trips totally indoors
- Number trips
- Classify mode-of-transportation based on average speed cutoffs

Mode of Transportation

- Classify trips as walking, running, bicycle, vehicle
- Average trip speed used as initial classifier
- Current implementation needs a pause point to detect mode changes within a trip (walk – car – walk)
- Assisted by use of accelerometer and/or heart rate data (if available)



Validation study

- To test the sensitivity & specificity of PALMS algorithms
- Data set that can be used to test improvements in calculations
- Data set that can be used as a demonstration set
- Protocols that can be used by other researchers

Aggregating data

- Creating meaningful variables from millions of data points
- # trips, distance travelled
- Time spent in locations & modes
- Daily & event based summaries

Integration with GIS (Geographic Information System)

- Aggregating GPS data
  - Select in time and space
- Include environmental characteristics
  - Green space, sport facilities, cycle routes, land use mix, intersection density, etc.
- Animate activity and movement

Active Living Research

Building the Evidence to Prevent Childhood Obesity and Support Active Communities

1 school, 1 schoolday (8am-2pm), 151 kids, ≈ 220,000 points

Active Living Research

Building the Evidence to Prevent Childhood Obesity and Support Active Communities

1 school recess (9.30-9.45am), 151 kids, ≈ 9,000 points

Active Living Research

Building the Evidence to Prevent Childhood Obesity and Support Active Communities

1 school recess, 4 grade levels

Active Living Research

Building the Evidence to Prevent Childhood Obesity and Support Active Communities

1 school recess, activity intensity

Active Living Research

Building the Evidence to Prevent Childhood Obesity and Support Active Communities

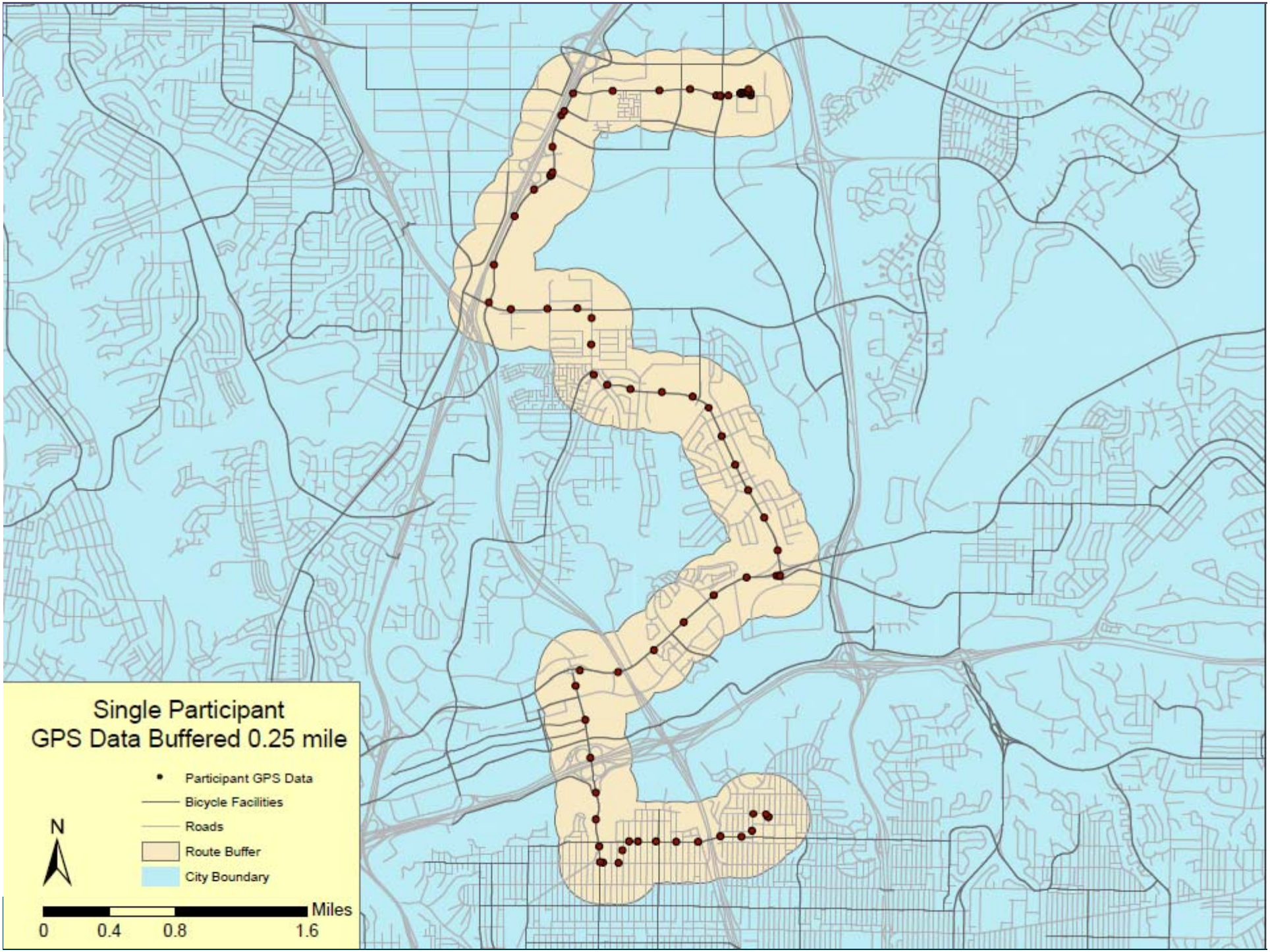
1 school recess, vigorous & very vigorous activity

Include environmental characteristics

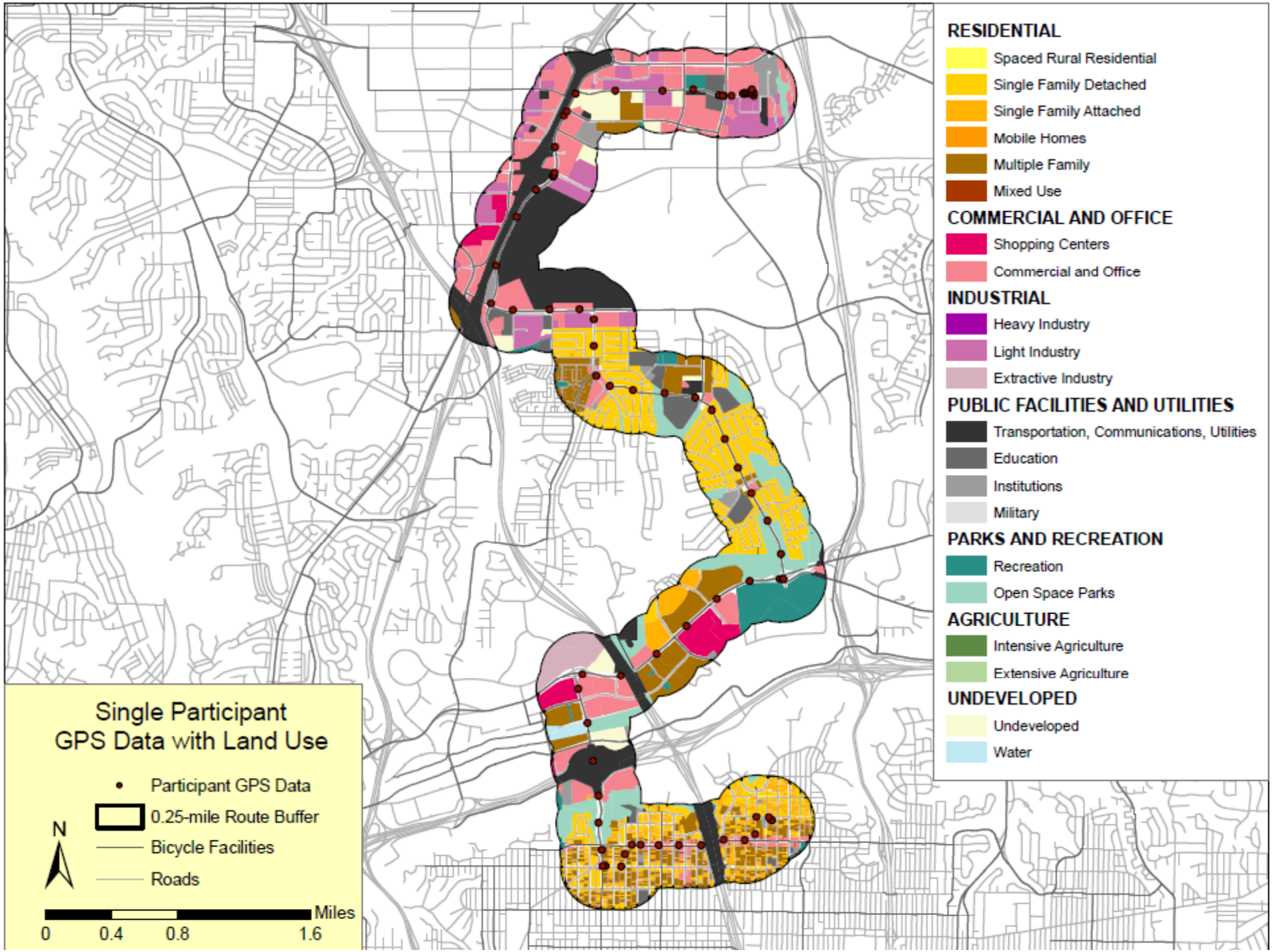
- Green space, sport facilities, cycle routes, land use mix, intersection density, etc.
- Analysis only as good as the data you're using
- People don't look at administrative borders; behavior doesn't stop at county or city border

Exposure buffers around routes

- Slides from Sherry Ryan- SDSU


Single Participant
GPS Data Buffered 0.25 mile
Participant GPS Data
Bicycle Facilities
Roads
Route Buffer
City Boundary
N
Miles
0
0.4
0.8
1.6


RESIDENTIAL
Spaced Rural Residential
Single Family Detached
Single Family Attached
Mobile Homes
Multiple Family
Mixed Use
COMMERCIAL AND OFFICE
Shopping Centers
Commercial and Office
INDUSTRIAL
Heavy Industry
Light Industry
Extractive Industry
PUBLIC FACILITIES AND UTILITIES
Transportation, Communications, Utilities
Education
Institutions
Military
PARKS AND RECREATION
Recreation
Open Space Parks
AGRICULTURE
Intensive Agriculture
Extensive Agriculture
UNDEVELOPED
Undeveloped
Water
Single Participant
GPS Data with Land Use
Participant GPS Data
0.25-mile Route Buffer
Bicycle Facilities
Roads
N
Miles
0
0.4
0.8
1.6

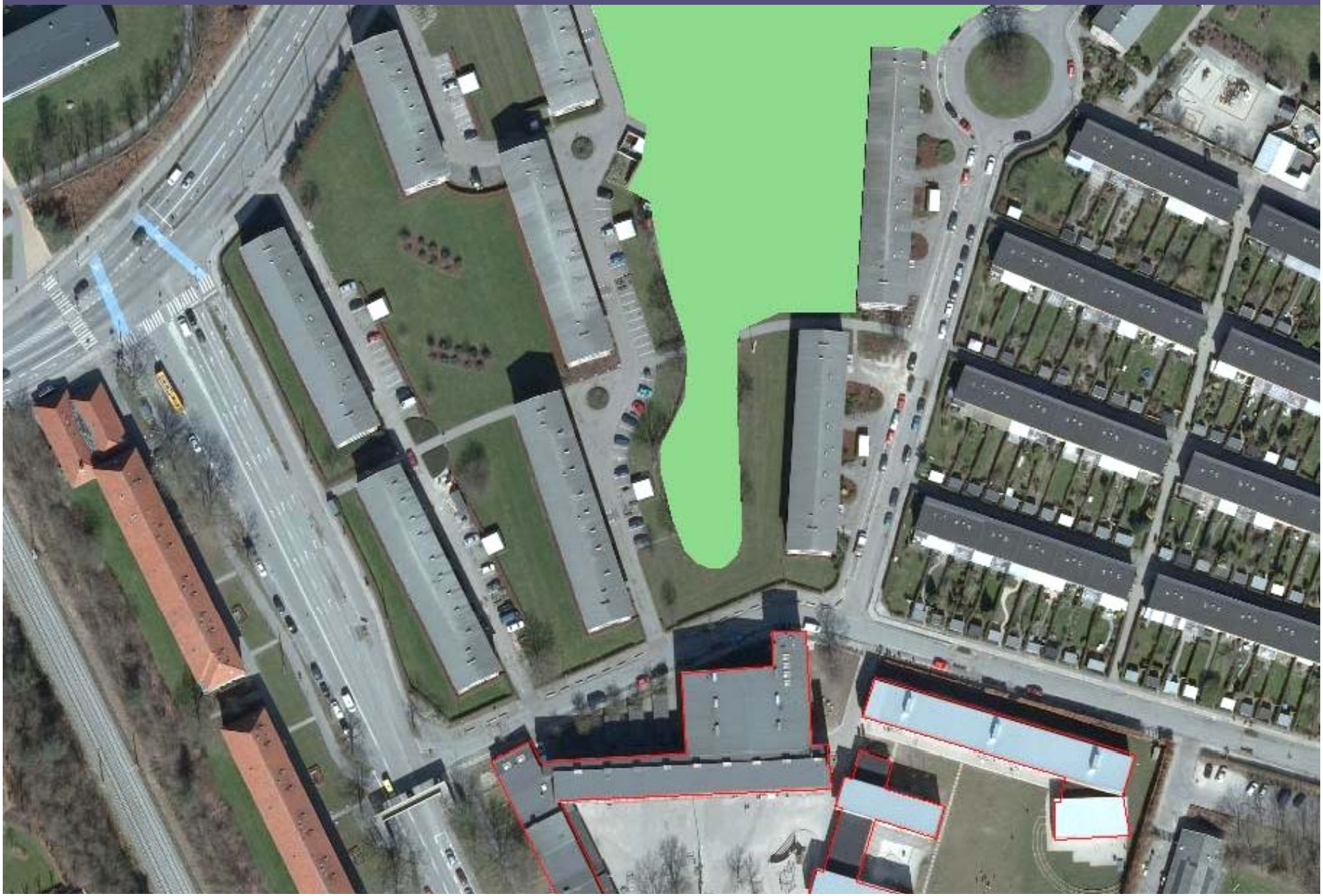

RESIDENTIAL
Spaced Rural Residential
Single Family Detached
Single Family Attached
Mobile Homes
Multiple Family
Mixed Use
COMMERCIAL AND OFFICE
Shopping Centers
Commercial and Office
INDUSTRIAL
Heavy Industry
Light Industry
Extractive Industry
PUBLIC FACILITIES AND UTILITIES
Transportation, Communications, Utilities
Education
Institutions
Military
PARKS AND RECREATION
Recreation
Open Space Parks
AGRICULTURE
Intensive Agriculture
Extensive Agriculture
UNDEVELOPED
Undeveloped
Water
Single Participant
GPS Data with Land Use
Participant GPS Data
0.25-mile Route Buffer
Bicycle Facilities
Roads
N
Miles
0
0.4
0.8
1.6

Active Living Research

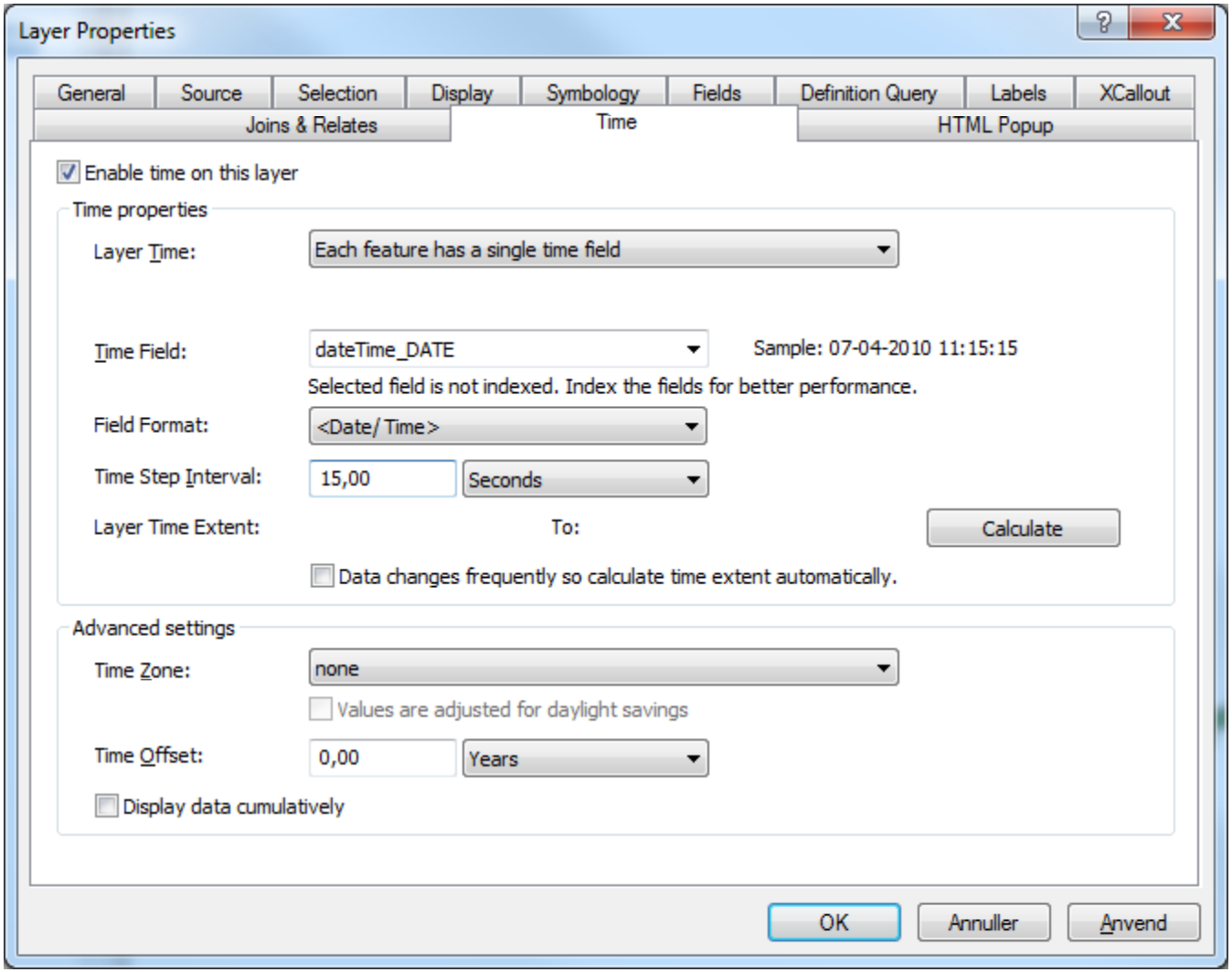
Building the Evidence to Prevent Childhood Obesity and Support Active Communities

Active Living Research

Building the Evidence to Prevent Childhood Obesity and Support Active Communities



Enable time field - new in ArcGIS 10



Active Living Research

Building the Evidence to Prevent Childhood Obesity and Support Active Communities

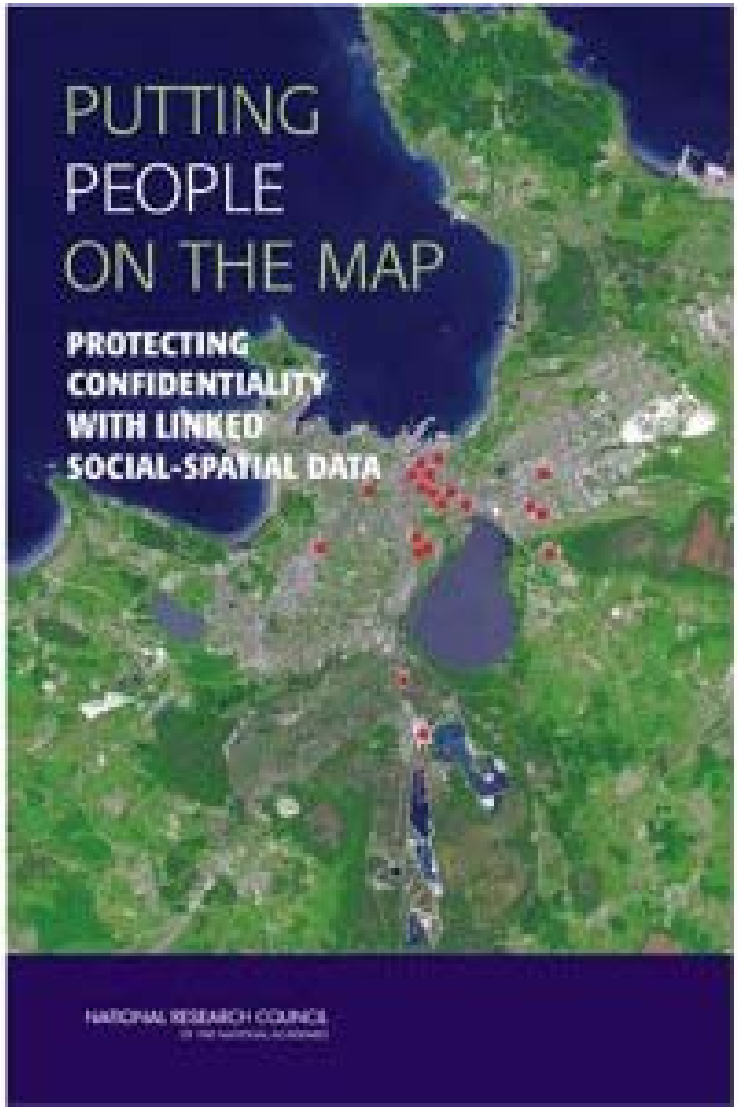
1 school recess, four 5th grade students

How to do all this in GIS?

- Helps to have GIS expert on team
- Developing step by step guide
- Developing a PALMS ArcGIS toolbar

Limitations/caution

- Privacy issues
- Staff time/cost
- Data size & quality
- Sample sizes
- New analytical techniques



Summarize

- Why it is challenging
- Why it is worth it
- What you need to do to get good data
- What you need to do to make sense of it all
- What next?

GPS HRN

The GLOBAL POSITIONING SYSTEMS in Health Research Network

www.gps-hrn.org

AUT UNIVERSITY

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

Beyond 2011

New Paradigms to Improve Physical Activity & Nutrition

www.beyond2011.org.nz



Nau mai haere mai

We are pleased to invite you to join us for the

2011 ISBNPA Post Conference Satellite Meeting in Queenstown, New Zealand

Rydges Lakeland Resort Hotel Queenstown, 38 - 54 Lake Esplanade, Queenstown. 21-23 June 2011

Ko te pae tawhiti whaia kia tata, ko te pae tata whakamaua kia tīna!

- EXTRAS if needed

- Test accuracy of algorithm to detect
  - Trips, Modes, Locations, Indoor/Outdoor
- Test Conditions
  - Cold/warm start
  - Urban canyon vs open space
  - 4 different modes of transportation
    - walking, cycling, bus riding, driving
  - Effects of pauses
    - transition modes and pauses within some tours
  - Data collected at 5, 15, and 30 second epochs
  - Device comparison
    - GPS: DG-100 and Qstarz BT1000XT

Methods

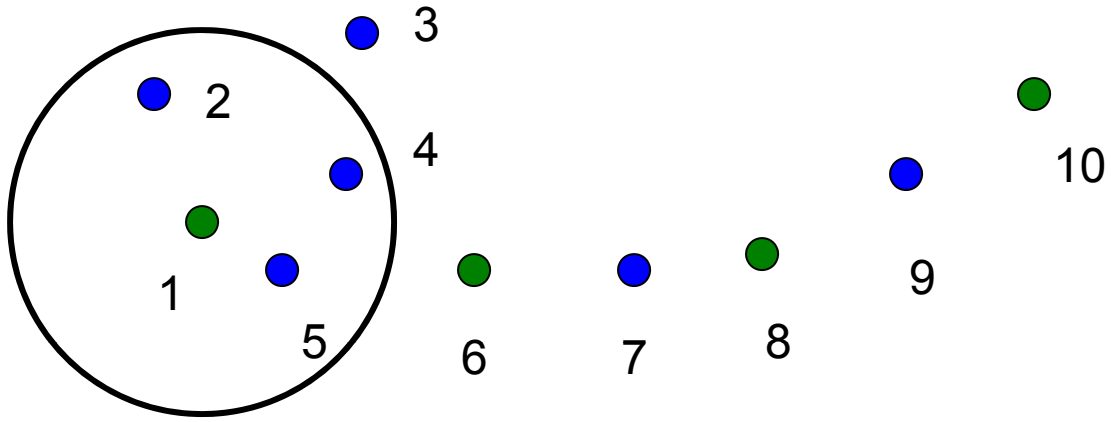
- Trained raters wear
  - Activity meters: GT3X, ActiTrainer, GT3X+
- Carry 2 boards each with 6 devices in backpack
  - Cold start/warm start
  - 2 device models
  - 3 epochs
- Follow a set protocol of trips & pauses
- 4 trip examples per condition, 500+ trips
- Note start time for each movement change

CW PHS
CENTER FOR WIRELESS & POPULATION HEALTH SYSTEMS

Analyses

- **Validation Questions**
  - E.g. How 'good' is PALMS at detecting walking trips?
- **Challenges**
  - Not independent observations – nested data
  - Determining unit of analysis
  - Not naturalistic sampling of trips
- **Determine the referent – define a 'hit' or 'miss'**
  - Percentage of rows with correctly identified modes of transportation in a day
  - Number of trips correctly identified in a day
- **Summary measures of agreement**
  - Sensitivity, specificity, Kappa – assumes independent observations
- **Predictors of agreement – multi-level models (e.g. GLM)**
  - IVs – Trip type, position in tour, trip number of the day, urban/open, etc.

GPS Filtering: Illustration

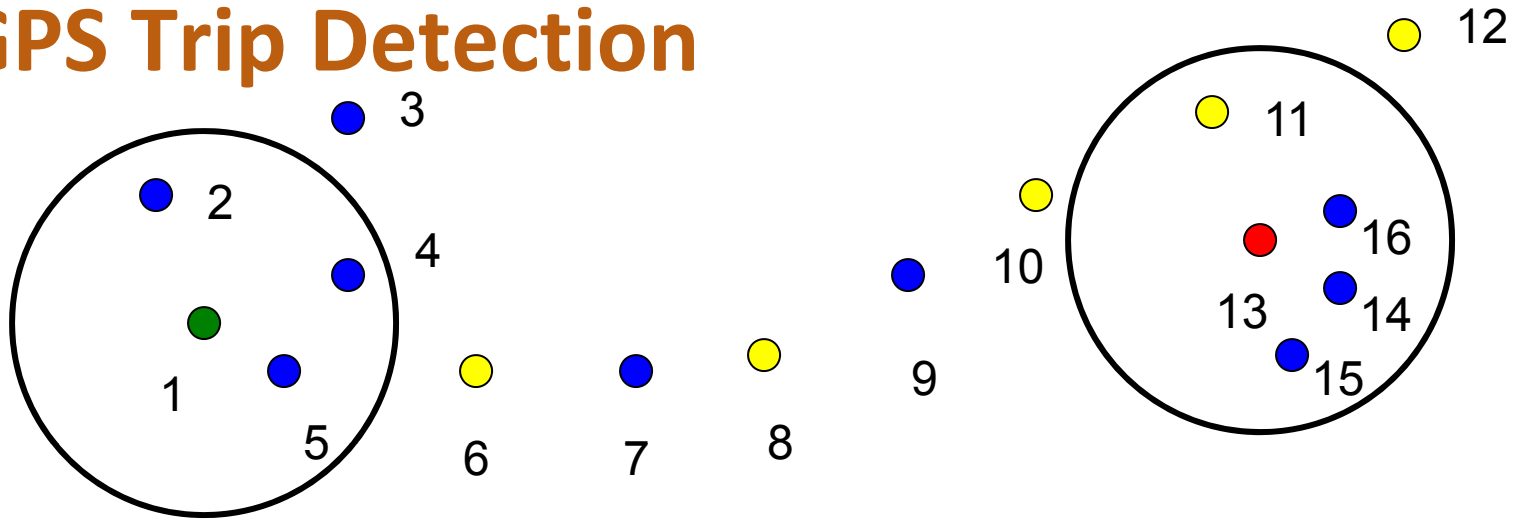


If D(n,n+1) < minimum_distance, then remove n+1

If D(n,n+2) < minimum_distance, then remove n+1

Point 1 - kept (first fix)
Point 2 - delete (< min)
Point 3 - delete (forward and backwards 1-3-4)
Point 4 - delete (< min)
Point 5 - delete (< min)
Points 6, 8,10 - kept as valid
Points 7, 9 - delete (< min)

GPS Trip Detection



- Points 1-5 previously filtered to one stationary point (point 1)
- Point 6 detected as start of trip based on distance traveled
  - mark point 1 as start of trip, mark point 6 as midpoint
- Points 7,9 previously filtered as redundant
- Points 8,10,11,12 identified as midpoints on the trip based on distance traveled
- Point 13-16 previously filtered to one stationary point (point 13)
- Point 13 detected as end point (time at point exceeded)

GPS Location Detection: PALMS implementation

- If trackpoint = First fix or Last fix or Start point or End point or Pause point, then add to location list
- If time at a trackpoint > threshold, then add to location list
- Once all trackpoints have been processed, find adjacent locations (distance between locations < threshold)
- For those adjacent locations, determine which location has the greatest amount of time spend at that location
- Collapse nearby locations to that location
- Number locations
- For each trackpoint, determine if it is within the location and if so, assign the location number to the trackpoint.

GPS Pre-processing: Derived Values

Given two fixes F1, F2 with time, lat, lon, ele

Distance = great_circle(F1,F2)

Duration = F2.time – F1.time

Elevation Delta = F2.ele – F1.ele

Speed = Distance / Duration

Grade = Elevation Delta / Distance
(change of 6m over 100m = 6% grade)

GPS Pre-processing

- Calculates distance, duration, speed, elevation-delta, grade and bearing between adjacent trackpoints

- Detects loss-of-signal
  - Defined as the duration between trackpoints exceeding X minutes (gaps in time)
  - Marks first, last and lone fixes

- Calculates SNR (signal to noise ratio) of satellites in view and satellites used to obtain fix (for Qstarz GPS)