GPS-based time-activity classification for personal exposure to air pollutant concentrations

Enki Yoo

Dept. of Geography,
University at Buffalo, SUNY

June 27, 2014

Exposure to air pollutants

- "Outdoor air pollution, a leading environmental cause of cancer deaths" [source: WHO, 10/17/2013]

Exposure to air pollutants

- "Outdoor air pollution, a leading environmental cause of cancer deaths" [source: WHO, 10/17/2013]
- Jeff's asthma attack

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.
- personal exposure estimates

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.
- personal exposure estimates
  1. continuous air quality

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.
- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.
- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.
  2. time-location-activity

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.
  2. time-location-activity
     - traditionally based on questionnaire and travel diaries

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.
  2. time-location-activity
     - traditionally based on questionnaire and travel diaries
     - GPS enables to keep track of activities of people at finer resolution in space and time.

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.
  2. time-location-activity
     - traditionally based on questionnaire and travel diaries
     - GPS enables to keep track of activities of people at finer resolution in space and time.
     - can be summarized into significant microenvironment (ME)s, such as home, work place, other indoor, in-vehicle, outdoor (Ott:82)

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.
  2. time-location-activity
     - traditionally based on questionnaire and travel diaries
     - GPS enables to keep track of activities of people at finer resolution in space and time.
     - can be summarized into significant microenvironment (ME)s, such as home, work place, other indoor, in-vehicle, outdoor (Ott:82)
  3. health outcomes

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.
  2. time-location-activity
     - traditionally based on questionnaire and travel diaries
     - GPS enables to keep track of activities of people at finer resolution in space and time.
     - can be summarized into significant microenvironment (ME)s, such as home, work place, other indoor, in-vehicle, outdoor (Ott:82)
  3. health outcomes
     - acute health outcomes: biomarkers of exposure collected from breath, urine, hair, nails, or blood or fluids

Personal exposure assessment

- associations between air pollutant concentrations and adverse health outcomes found in epidemiological studies may be misleading.

- personal exposure estimates
  1. continuous air quality
     - ambient air pollution data obtained from central monitoring stations, spatially coarse and homogeneous assumption within the boundary within which population data are collected.
     - personal monitoring, which has been script-based and expensive, and potentially intervene the daily routine of individuals as well as privacy.
  2. time-location-activity
     - traditionally based on questionnaire and travel diaries
     - GPS enables to keep track of activities of people at finer resolution in space and time.
     - can be summarized into significant microenvironment (ME)s, such as home, work place, other indoor, in-vehicle, outdoor (Ott:82)
  3. health outcomes
     - acute health outcomes: biomarkers of exposure collected from breath, urine, hair, nails, or blood or fluids
     - accumulated health outcomes in longitudinal studies

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors
   - portable personal device can be used with limitations, such as feasibility, participant burden, high cost

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors
   - portable personal device can be used with limitations, such as feasibility, participant burden, high cost
2. mobility of individuals

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors
   - portable personal device can be used with limitations, such as feasibility, participant burden, high cost

2. mobility of individuals
   - recall bias and reliability in self-reported diary or questionnaires (Crosbie:06).

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors
   - portable personal device can be used with limitations, such as feasibility, participant burden, high cost

2. mobility of individuals
   - recall bias and reliability in self-reported diary or questionnaires (Crosbie:06).
   - GPS equipped devices can be used in combination with questionnaires and travel diaries, which calls for raw GPS data transformation into relevant information, such as "the time spent at a specific ME at a certain location".

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors
   - portable personal device can be used with limitations, such as feasibility, participant burden, high cost

2. mobility of individuals
   - recall bias and reliability in self-reported diary or questionnaires (Crosbie:06).
   - GPS equipped devices can be used in combination with questionnaires and travel diaries, which calls for raw GPS data transformation into relevant information, such as "the time spent at a specific ME at a certain location".
   - GPS positions often "*drift*" around an actual location (Nathery:14).

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors
   - portable personal device can be used with limitations, such as feasibility, participant burden, high cost

2. mobility of individuals
   - recall bias and reliability in self-reported diary or questionnaires (Crosbie:06).
   - GPS equipped devices can be used in combination with questionnaires and travel diaries, which calls for raw GPS data transformation into relevant information, such as "the time spent at a specific ME at a certain location".
   - GPS positions often "*drift*" around an actual location (Nathery:14).
     - GPS spatial errors at darkness: satellite blocking, temporal and spatial variations in satellite geometry, signal reflection, etc.

Misclassification of personal exposure

Potential exposure misclassifications due to

1. air quality measurements
   - the lack of spatial coverage of central-site ambient monitors
   - portable personal device can be used with limitations, such as feasibility, participant burden, high cost

2. mobility of individuals
   - recall bias and reliability in self-reported diary or questionnaires (Crosbie:06).
   - GPS equipped devices can be used in combination with questionnaires and travel diaries, which calls for raw GPS data transformation into relevant information, such as "the time spent at a specific ME at a certain location".
   - GPS positions often "*drift*" around an actual location (Nathery:14).
     - GPS spatial errors at darkness: satellite blocking, temporal and spatial variations in satellite geometry, signal reflection, etc.
     - potential autocorrelation in space and time

Objectives

We aim to

1. develop a GPS-based time-activity classification method for air pollution study and validate the performance of the proposed method.

Objectives

We aim to

1. develop a GPS-based time-activity classification method for air pollution study and validate the performance of the proposed method.
2. assess uncertainty in the exposure estimates due to estimated time-activity patterns.

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes
- outputs

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes
- outputs
  - a type of microenvironments (MEs): home, workplace, other indoor, in-vehicle, outdoor

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes
- outputs
  - a type of microenvironments (MEs): home, workplace, other indoor, in-vehicle, outdoor
  - spatial location (the centroid of the place and the size)

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes
- outputs
  - a type of microenvironments (MEs): home, workplace, other indoor, in-vehicle, outdoor
  - spatial location (the centroid of the place and the size)
  - time spent at each ME

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes
- outputs
  - a type of microenvironments (MEs): home, workplace, other indoor, in-vehicle, outdoor
  - spatial location (the centroid of the place and the size)
  - time spent at each ME
- model performance: discrepancy with travel diary

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes
- outputs
  - a type of microenvironments (MEs): home, workplace, other indoor, in-vehicle, outdoor
  - spatial location (the centroid of the place and the size)
  - time spent at each ME
- model performance: discrepancy with travel diary
  - type/number of MEs where time activity occurred

Methods

Rule-based classification: an automated classification method to transform raw GPS data into time-location-activity.

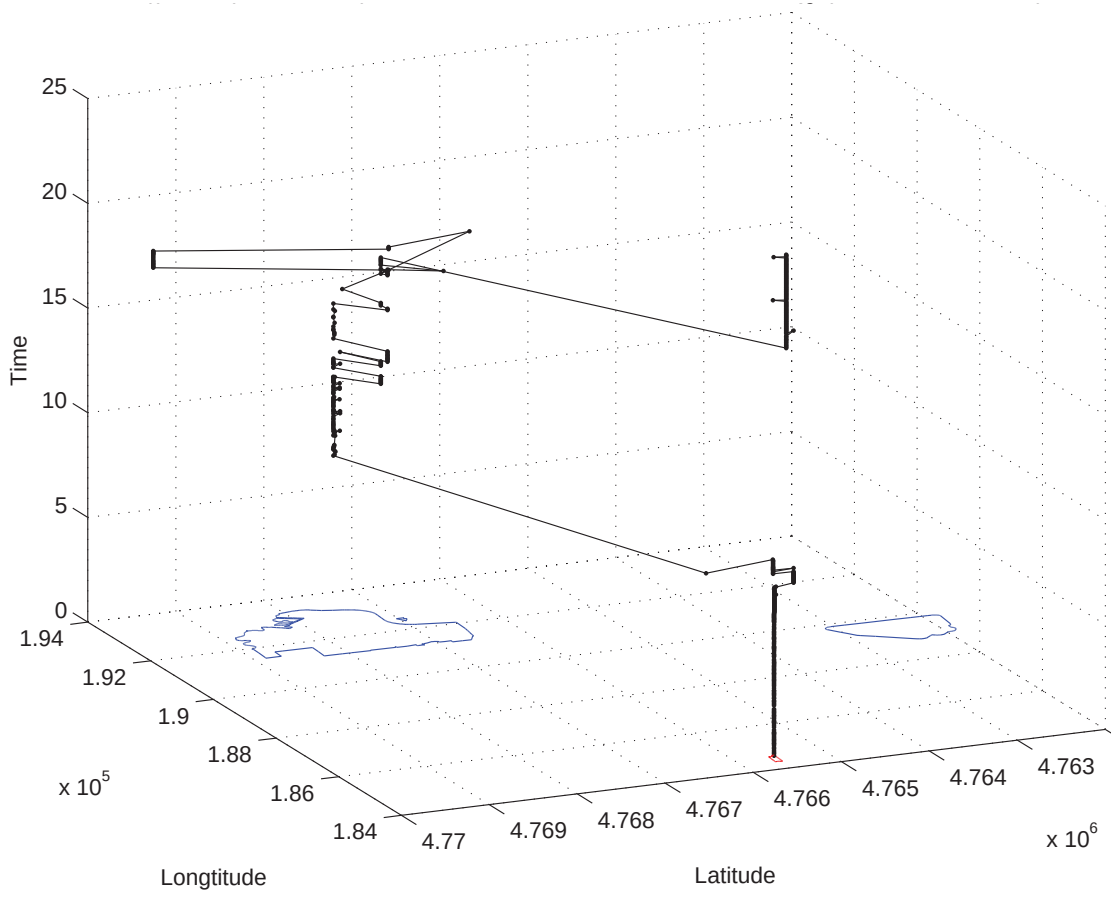
- input data
  - GPS recordings (longitude, latitude, and time/day)
  - base questionnaire inform home and work place address for each participant as well as occupation and gender.
  - spatial location of geocoded home and work place addresses
  - parcel boundaries and property classes
- outputs
  - a type of microenvironments (MEs): home, workplace, other indoor, in-vehicle, outdoor
  - spatial location (the centroid of the place and the size)
  - time spent at each ME
- model performance: discrepancy with travel diary
  - type/number of MEs where time activity occurred
  - time spent at each ME

Input Data

- Travel diary

ID	Date	StartTime	EndTime	Location	Address	City	State
A009	1/27/12	0:01:00	11:15:00	Home	406 Linwood Avenue	Buffalo	New York
A009	1/27/12	11:35:00	23:59:59	Home	406 Linwood Avenue	Buffalo	New York
A009	1/29/12	0:01:00	9:55:00	friend	130 Emerson Drive	Amherst	New York
A009	1/29/12	9:55:00	10:09:00	in-vehicle			
A009	1/29/12	10:10:00	10:25:00	Home	406 Linwood Avenue	Buffalo	New York
A009	1/29/12	10:35:00	11:25:0				

- Baseline questionnaire: age, gender, occupation, …
- Time-location information from a smartphone app.



Rule-based classification algorithm

1. Home ME: focus on GPS points at dark (12:00 and 6:00)

Rule-based classification algorithm

1. Home ME: focus on GPS points at dark (12:00 and 6:00)

   - spatial search: locate any point within 1 km from geocoded home address (78% covered).

Rule-based classification algorithm

1. Home ME: focus on GPS points at dark (12:00 and 6:00)

   - spatial search: locate any point within 1 km from geocoded home address (78% covered).
   - spatio-temporal cluster search:

Rule-based classification algorithm

1. Home ME: focus on GPS points at dark (12:00 and 6:00)

   - spatial search: locate any point within 1 km from geocoded home address (78% covered).
   - spatio-temporal cluster search:
     - if consecutive GPS points are located within 2 km from home address, they are classified as a home ME, but they are considered 'other indoor ME' if the separation distance is more than 2 km.

Rule-based classification algorithm

1. Home ME: focus on GPS points at dark (12:00 and 6:00)

   - spatial search: locate any point within 1 km from geocoded home address (78% covered).
   - spatio-temporal cluster search:
     - if consecutive GPS points are located within 2 km from home address, they are classified as a home ME, but they are considered 'other indoor ME' if the separation distance is more than 2 km.
     - a single untagged point located away from home ($\geq$ 1.5km) with the time spent at the location is more than 30 min. is assigned as 'other indoor ME'. Any single untagged point with time spent less than 30 min. located within 2km from home is considered a 'home ME', but it is considered 'GPS error' if the distance to home is greater than 2km.

Rule-based classification algorithm

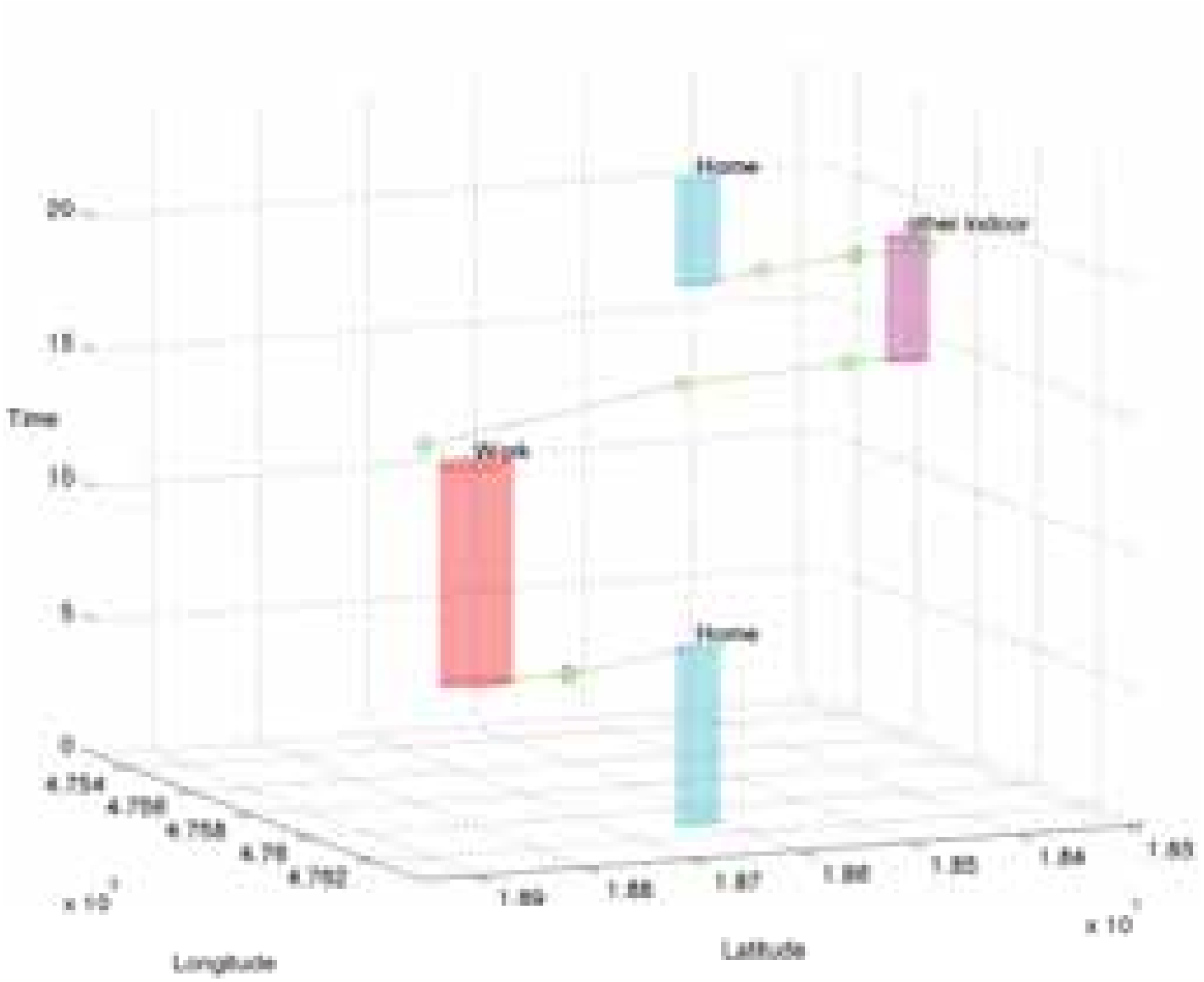
1. Home ME: focus on GPS points at dark (12:00 and 6:00)

   - spatial search: locate any point within 1 km from geocoded home address (78% covered).
   - spatio-temporal cluster search:
     - if consecutive GPS points are located within 2 km from home address, they are classified as a home ME, but they are considered 'other indoor ME' if the separation distance is more than 2 km.
     - a single untagged point located away from home ($\geq$ 1.5km) with the time spent at the location is more than 30 min. is assigned as 'other indoor ME'. Any single untagged point with time spent less than 30 min. located within 2km from home is considered a 'home ME', but it is considered 'GPS error' if the distance to home is greater than 2km.
   - extend the search after dark by overlaying with GPS points with home clusters.

Rule-based classification algorithm, cont.

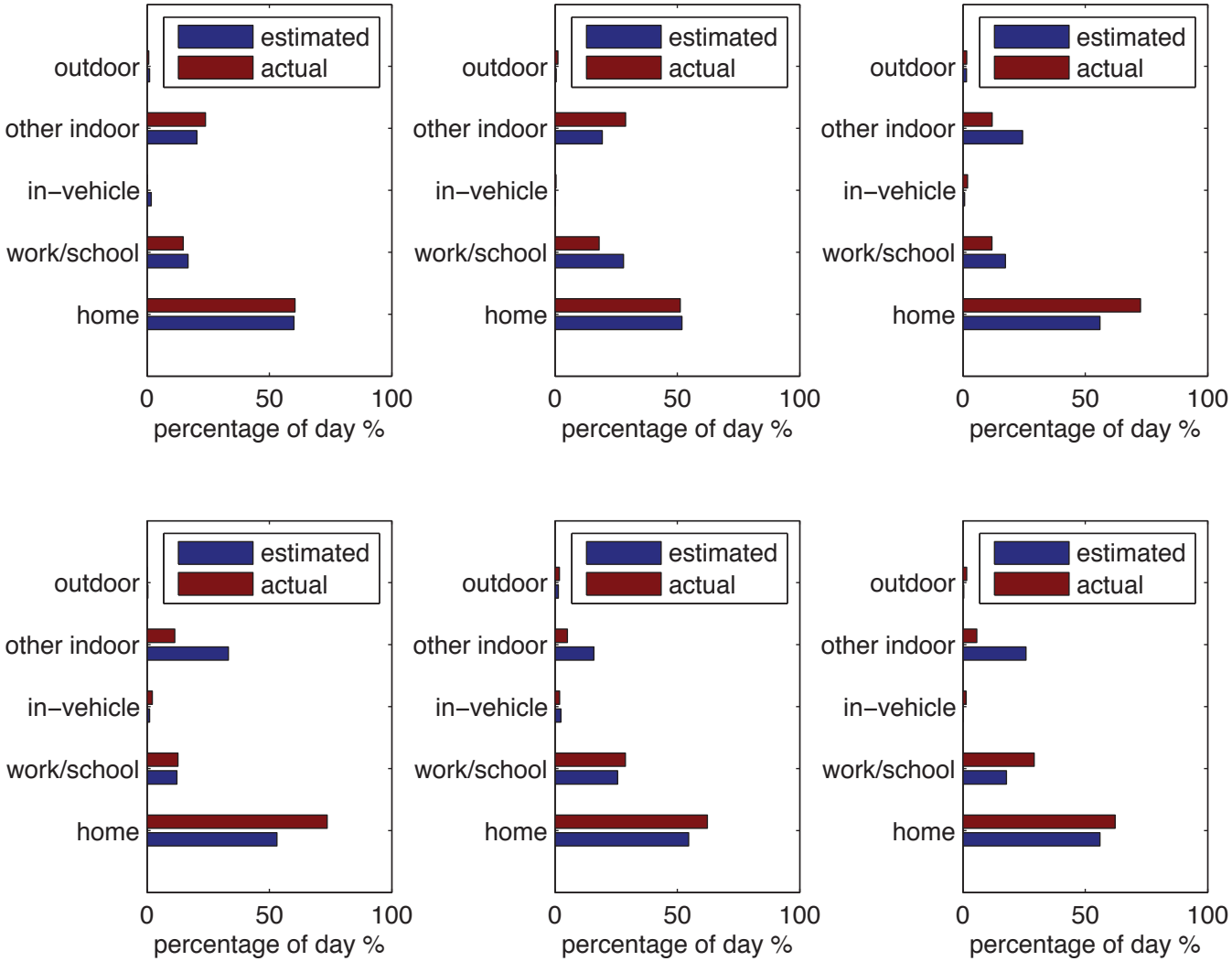
3. work place: there are students who live near or on campus.
   - any point is located within 1 km from the work place address.
   - any point located within the boundary of work places, such as school boundaries.
4. outdoor: any point belong to the known boundary of outdoor space, such as parks, playground, and vacant lands (delineated from GIS parcel map).
5. in-vehicle: based on the travel distance between two consecutive GPS recordings and the speed.
   - any pair of consecutive points with speed 15 mph and their travel distance per minute is $\geq$ 0.8 km.
   - any point in-transit between two different MEs when the speed is $>$ 10 mph.
6. other indoor: a spatiotemporal cluster of GPS points that do not belong to home, work place, nor in-transit.

Estimated time-activity patterns



Performance evaluation

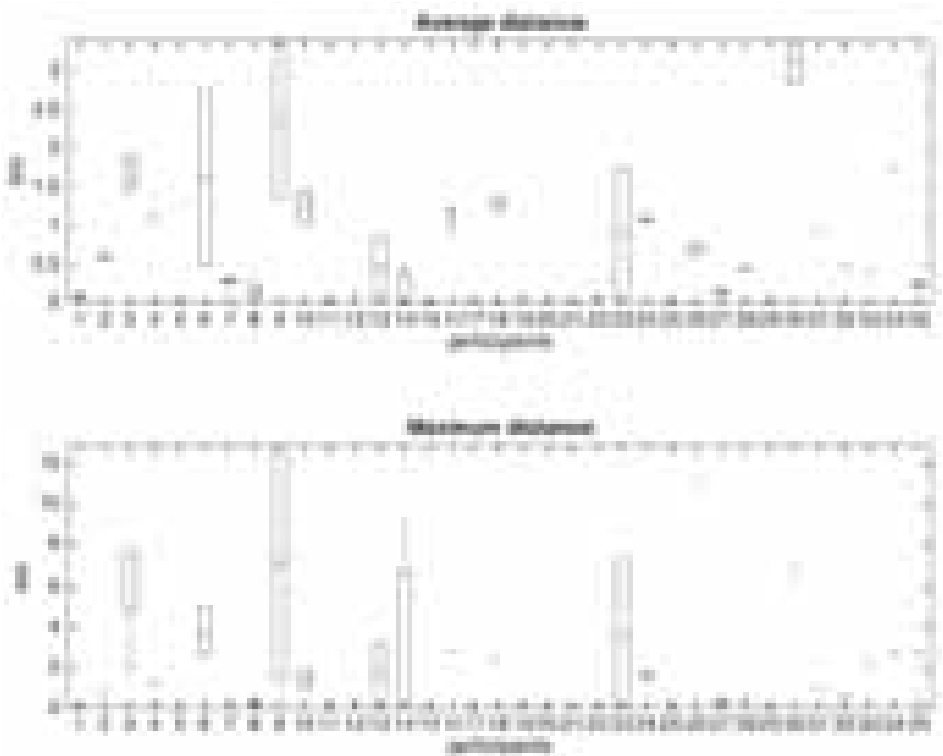
- comparison of GPS-based classification outputs (into five categories) with travel diaries
- on three randomly assigned days in weekend and weekday



Spatial discrepancy between GPS points and diary at darkness

Travel dairies vs. GPS spatial errors at darkness (12:00 to 6:00 am)

- a total of 35 participants with matching GPS records (a total of 70 days)
- 51% of participants have two days of records (1 to 4 days)
- 90 % of participants reported one time-activity record
- spatial errors: difference between GPS recordings and some known places (home, other indoor, work place) from travel diary/questionnaire.
  - average mean distance 0.67 km (up to 3.4 km)
  - average maximum distance 1.9 km (up to 13 km).



Uncertainty propagation in personal exposure assessment

large spatial discrepancy between DTA and GPS points

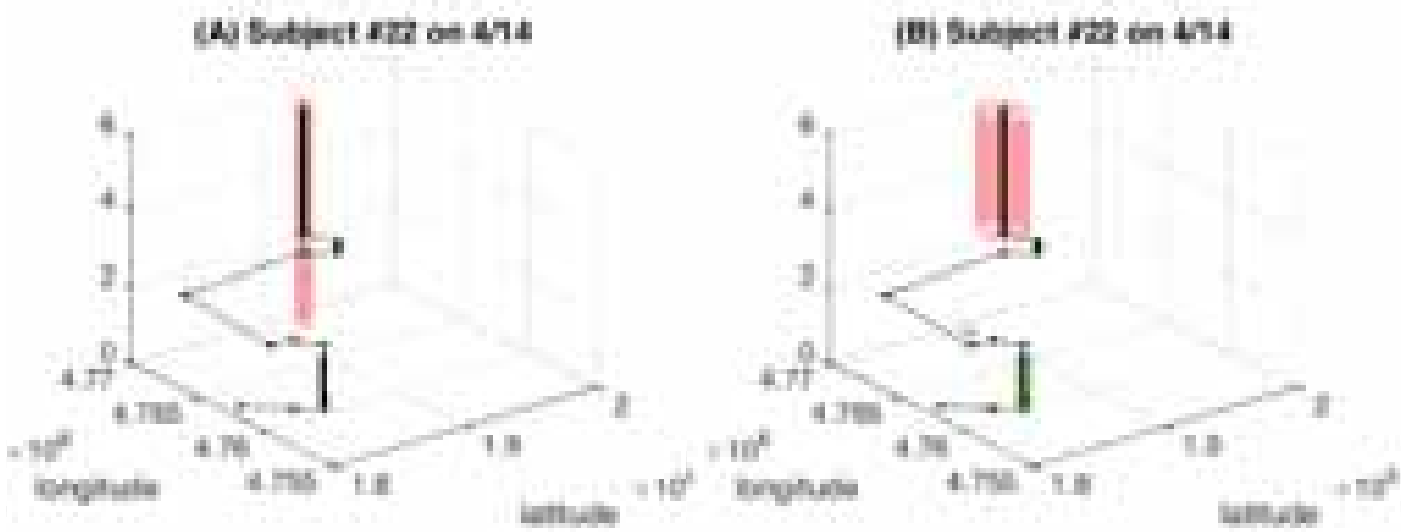
misclassification in time-activity patterns

misclassification in personal exposure estimates

- the quality of time-activities estimated from GPS-based classification method depends on the accuracy of GPS data.
- inaccurate GPS points are likely to yield misclassified time-activity patterns.
- misclassified time-activities yield misclassification of personal exposure.

Illustration of uncertainty propagation

Misclassified time-acivities (MEs)



(A) GPS points with travel diary (home ME)

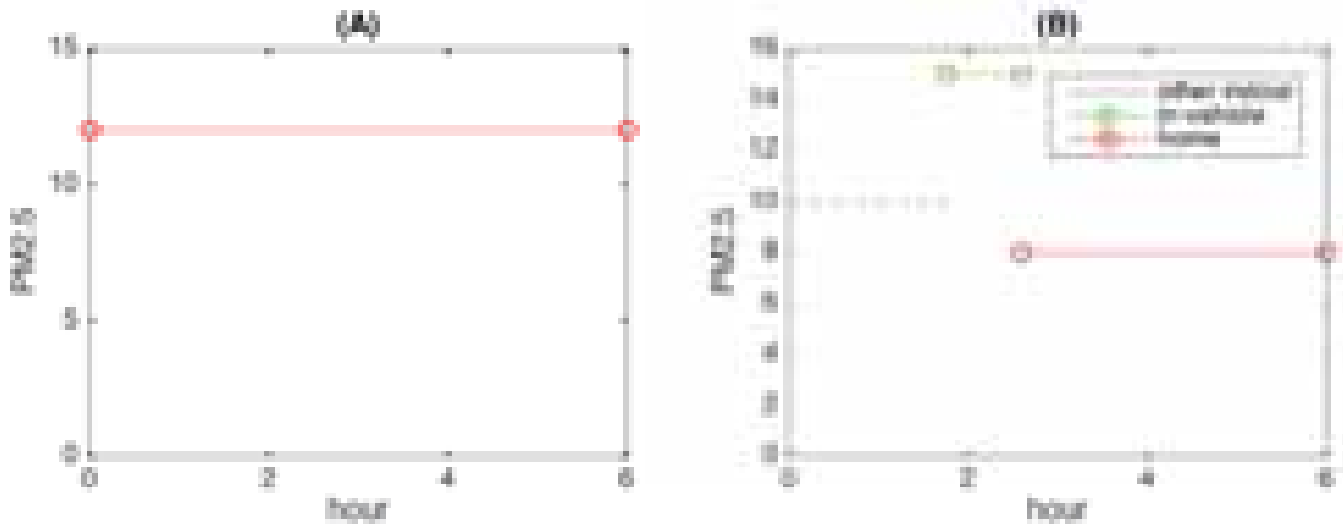
(B) GPS points with time-activities estimated from classification method

*red : home ME

*green: other indoor

Illustration of uncertainty propagation (II)

Misclassification of personal exposure



(A) home-based exposure estimate

(B) GPS-based time-activity classification based exposure estimate

Conclusions

1. we developed a GPS-based time-activity classification method, which potentially replace labor-intensive manual matching procedure.

2. we validate the proposed time-activity classification method with travel diaries.

3. we found substantial spatial discrepancy between travel diary and GPS recordings at dark, which can be somewhat alleviated by the proposed classification method but potentiallyly leads to misclassification in estimated time-activity patterns and exposure estimates.

Future work

1. to add multiple environmental sensors to enhance the capability to classify individuals' time-activity patterns and improve the accuracy.
   - waypoint data sampling by participants' manual input (Breen 14)
   - temperature data (Nethery:14), and humid, wind, ultraviolet, light, dust information to classify GPS samples into various types of MEs.
2. to evaluate the proposed model with a larger number of participants whose background information is diverse and representative.
3. to improve the time-activity pattern data used for population-level exposure assessments.
   - can be applied to large sets of GPS data collected from smartphones
   - to increase sample size and update older diary data to reflect the changes in population and result in time-activity patterns of large population with less errors.
4. to enhance the quality of travel diaries and questionnaire.

Acknowledgements

- Collaborators L. Mu and C. Rudra (Public health), M. Demirbas and A. Rudra (Computer science and engineering), A. Szpiro (Biostatistics)
- Students M. Glasgow (Public health), X. Jiang (Geography), N. Pramod (Computer science and engineering)
- Support from NIH R21 (5R21ES017826)