Front. Environ. Sci. Eng. 2022, 16(5): 63
https://doi.org/10.1007/s11783-022-1542-7

SHORT COMMUNICATION

On the potential of iPhone significant location data to characterize individual mobility for air pollution health studies

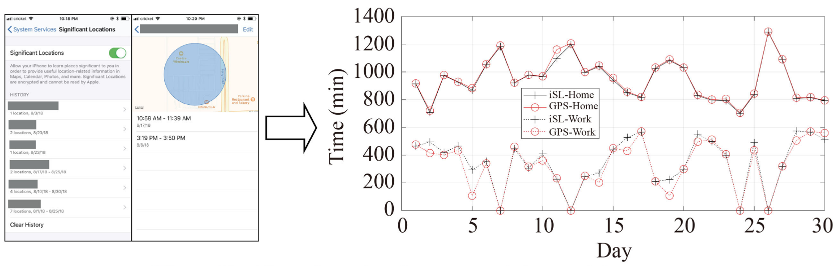
Eastman Elizabeth[1], A. Stevens Kelly[2], Ivey Cesunica[3], Haofei Yu (✉)[1]

1 Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL 32816, USA
2 School of Public Administration, University of Central Florida, Orlando, FL 32801, USA
3 Department of Civil and Environmental Engineering, University of California-Berkeley, Berkeley, CA 94720, USA

HIGHLIGHTS

- We evaluated the accuracy of iPhone data in capturing time-activity patterns.
- iPhone data captured the most important microenvironments and time spent in them.
- iPhone data also accurately captured daily exposure to ambient PM pollution.
- A considerable fraction of the population in the USA may have iPhone data available.
- iPhone data has great potential in air pollution health studies.

GRAPHIC ABSTRACT



iPhone Significant Location data captured time-activity pattern

ARTICLE INFO

Article history:
Received 16 December 2021
Revised 4 April 2022
Accepted 6 April 2022
Available online 28 April 2022

Keywords:
Air pollution exposure
Human mobility
iPhone
Significant Location
Smartphone data

ABSTRACT

In many air pollution health studies, the time-activity pattern of individuals is often ignored largely due to lack of data. However, a better understanding of this location-based information is expected to decrease uncertainties in exposure estimation. Here, we showcase the potential of iPhone's Significant Location (iSL) data in capturing the user's historical time-activity patterns in order to estimate exposure to ambient air pollutants. In this study, one subject carried an iPhone in tandem with a reference GPS tracking device for one month. The GPS device recorded locations in 10 second intervals while the iSL recorded the time spent in locations the subject visited frequently. Using GPS data as a reference, we then evaluated the accuracy of iSL data in capturing the subject's time-activity patterns and time-weighted air pollution concentration within the study time period. We found the iSL data accurately captured the time the subject spent in 16 microenvironments (i.e. locations the subject visited more than once), which was 93% of the time during the study period. The average error of time-weighted aerosol optical depth value, a surrogate of particle pollution, is only 0.012%. To explore the availability of iSL data among iPhone users, an online survey was conducted. Among the 349 surveyed participants, 72% of them have iSL data available. Considering the popularity of iPhones, iSL data may be available for a significant portion of the general population. Our results suggest iSL data have great potential for characterizing historical time-activity patterns to improve air pollution exposure estimation.

1 Introduction

Exposure to ambient air pollution is known to be associated with a wide range of adverse health impacts (Bernstein et al., 2004), and millions of pre-mature deaths across the globe are attributable to air pollution exposure in a single year (GBD 2016 Risk Factors Collaborators, 2017). To investigate the associations between pollution exposure and health outcomes, and to design effective policies to mitigate such impacts, it is crucial to appropriately quantify human exposure to air pollution.

✉ Corresponding author
E-mail: haofei.yu@ucf.edu

Special Issue—Frontier Progresses from Chinese-American Professors of Environmental Engineering and Science (Responsible Editors: Xing Xie, Jinkai Xue & Hongliang Zhang)

One major source of uncertainty in accurately estimating an individual's air pollution exposure is the movement of the individual, or time-activity patterns (Bell et al., 2018). Neglecting time-activity patterns may introduce significant uncertainties in exposure estimation because ambient concentrations of many common air pollutants change spatiotemporally (Apte et al., 2017). Such mismatch may lead to potential errors and/or biases in exposure estimation, which could then bias subsequent statistical analyses (Park and Kwan, 2017).

Many methods have been used in the past to collect individual time-activity data to investigate how individual movements impact exposure to air pollution. Some of the methods involve manual data recording, such as a personal interview and travel diary (Klepeis et al., 2001). These approaches are subject to potential recall bias and compliance issues and may be intractably expensive for large sample populations. Individual location data can also be collected using dedicated GPS devices or specifically designed smartphone applications (Glasgow et al., 2016). These approaches perform well for prospective studies where future time-activity data are needed, but not for retrospective studies that depend on historical data.

With rapid technology advancement, smartphones running on Android Operating Systems (OS) and iOS are becoming globally pervasive (Yu et al., 2018). Nearly all smartphone users' location data are being collected and stored by OS and/or different applications. In a previous study, we showcase the potential of Google Maps Location History (GMLH) data in characterizing an individual's historical time-activity pattern (Yu et al., 2019). In the current study, we will demonstrate another promising yet largely overlooked source for historical time-activity data, the iPhone Significant Location (iSL) data.

iPhone is a class of smartphones designed by Apple Inc. (Cupertino, CA. USA). iPhone runs on the iOS operating system, and are one of the most popular smartphones on the market. As of May 2021, iPhones represent 52% of the smartphone market in the USA (Statista, 2021). All iPhones are equipped with a feature called "Significant Locations". If "Significant Locations" is enabled, iOS will continuously collect information on the locations the user may have visited frequently, and store the data on the user's iPhone. For this study, we also refer to "Significant Locations" as a microenvironment, a term commonly used in health studies.

In this exploratory study, we compared iSL data collected from a single individual with reference GPS data to evaluate the ability of iSL in capturing: 1) all microenvironments the subject visited during the study period; 2) the duration of time the subject spent at each microenvironment; and 3) the impact of neglecting time-activity patterns on the subject's air pollution exposure estimates. Further, we also conducted an online survey to explore the availability of iSL data among iPhone users in the USA, and the acceptability of using iSL data for academic research purposes. The main purpose of this study is not to directly apply iSL data in air pollution health studies, but to understand the potential of iSL data in air pollution health research.

2 Materials and methods

This work is intended to be a proof-of-concept study. Between July 31 and August 29, 2018, a single subject carried an iPhone 6s in tandem with a dedicated GPS logger (BT-Q1000XT, Qstarz International Co., Ltd., Taipei, China) for a one month time period. We configured to GPS logger to record coordinates (latitude and longitude) every 10 seconds. As an example, a portion of the collected iSL data are provided in Fig. 1(a) (the names of some cities are redacted for privacy reasons). Detailed information on one significant location (a retail store) is provided in Fig. 1(b). A portion of the collected GPS data are provided in Fig. 1(c), with corresponding iSL marked with circles. To capture a baseline scenario for data collection, the subject also carried a second smartphone for daily use. After completion of the data collection, we took screenshots of all pages of the iSL data (Fig. 1). Because we are not aware of any tools that are capable of extracting raw iSL

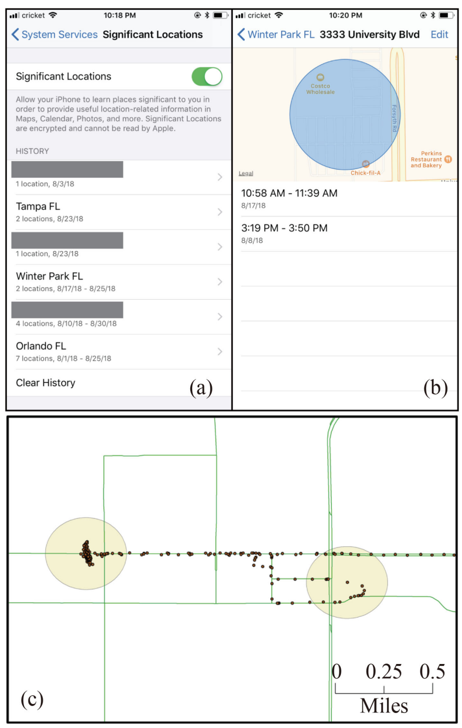


Fig. 1 Example iSL data stored on the subject's iPhone (a). An example of one significant location (b), and a portion of the collected GPS data with corresponding iSL shown in circles (c).

data, we manually identify the latitude and longitude of all iSL locations, and record all start and end times of iSL visits.

To estimate the duration of time the subject spent at each location based on GPS logger data, we draw a 300 meter radius circle on the center of identified location (Fig. 1(c)). We compared the center of all iSL circles with GPS coordinates of the corresponding location the subject visited, and found 300 meters to be the maximum spatial error of iSL data in this study. If a 10-second interval GPS data point was located within a circle, we assume the subject stayed at the corresponding location for the entire 10-second time period. We then count the number of GPS data points in each circle, and estimate the time (in minutes) the subject spent at the corresponding location. The iSL data specifically listed the start and end time of visits.

To evaluate the impact of neglecting time-activity pattern on the subject's pollution exposure estimates, we applied a similar method from our previous study. We retrieved Multiangle Implementation of Atmospheric Correction Aerosol Optical Depth (AOD) data (1 km resolution) during the corresponding study time period (Lyapustin et al., 2011), and extracted AOD data based on GPS and iSL locations separately. We separately estimated daily time-weighted AOD by aggregating AOD data at all locations for the corresponding day using the time the subject spent at different locations as weights. For comparison purposes, we also normalized the AOD estimates by setting the retrievals from the subject's home location to equal one.

To explore the fraction of iPhone users in the USA that have iSL data available from their smartphones, we conducted an online survey on the Amazon Mechanical Turk platform. The survey participants are restricted to iPhone users in the USA. Each participant is asked to specify their gender, age, and whether the significant location feature is turned on on their iPhone. The participants also have the option to provide comments on potential concerns if iSL data is used for academic research purposes. The survey typically took participants 2 minutes to complete, and each participant received $1 to compensate for their participation. This survey was approved by the Institutional Review Board at the University of Central Florida, USA.

3 Results and discussion

During the one-month time period, the subject visited 25 microenvironments located in eight cities (the names of three cities and one microenvironment, a retail store, are shown in Fig. 1), and iSL data captured 16 of them. An example of the collected data is shown in Fig. 1. Among the nine microenvironments not captured, the subject only made brief visits to four of them, including a gas station and pick-up/drop-off locations for local businesses. Four additional microenvironments are restaurants, and the last microenvironment is a grocery store, for which the subject did not bring the iPhone during the visit.

Although the subject's iSL data failed to capture nine microenvironments, it captured the most important 16 microenvironments where the subject spent the vast majority (93%) of their time (Table 1). In addition, the daily times (in minutes) the subject spent at home and work locations as captured by GPS and iSL data, are provided in Fig. 2(a). iSL data performs well at the subject's home and work locations, and captured the day-to-day variation of times the subject spent at the two major microenvironments. As shown in Fig. 2(a), iSL data performs particularly well at the subject's home location (mean error of 0.8%), and less favorable at the work location (mean error of 6.5%).

For the online survey, we received 349 responses, among them 73% (n = 223) indicated that the significant location feature is turned on in their iPhone. Among the 249 comments received, 48% (n = 140) participants indicated that they are not concerned if their iSL data were used for academic research purposes, 22% (n = 65) indicated unwillingness to share data, and 2% (n = 6) requested more information to make the decision. We note that 62% of the sample population (n = 218) were male, and 57% (n = 199) were between 25–34 years old. As an exploratory study, we do not anticipate the sample population to be representative of USA population (Ipeirotis, 2010).

Figure 2(b) provides normalized and time-weighted AOD as estimated using GPS and iSL data, which were used as a surrogate of ambient particle pollution here due to lack of data. The estimated time-weighted AOD using iSL were similar to the those estimates using reference GPS data (Fig. 2). Based on GPS data, the subject's daily time-weighted AOD varied between 0.95 to 1.12. The average error of estimates using iSL is only 0.012%. It's worth noting that, for comparison purposes, AOD is used as a surrogate of ambient particle pollution in this study, and AOD values at the subject's home location is set to 1.

Table 1 Time (in minutes) the subject spent at 16 microenvironments as estimated based on iPhone Significant Location (iSL) data and GPS logger data

Microenvironment ID	GPS (min)	iSL (min)	Microenvironment ID	GPS (min)	iSL (min)
1	28297	28536	9	161	166
2	109	112	10	10567	9876
3	306	314	11	114	123
4	46	49	12	61	65
5	72	79	13	101	106
6	17	25	14	71	78
7	120	235	15	108	123
8	54	64	16	49	20

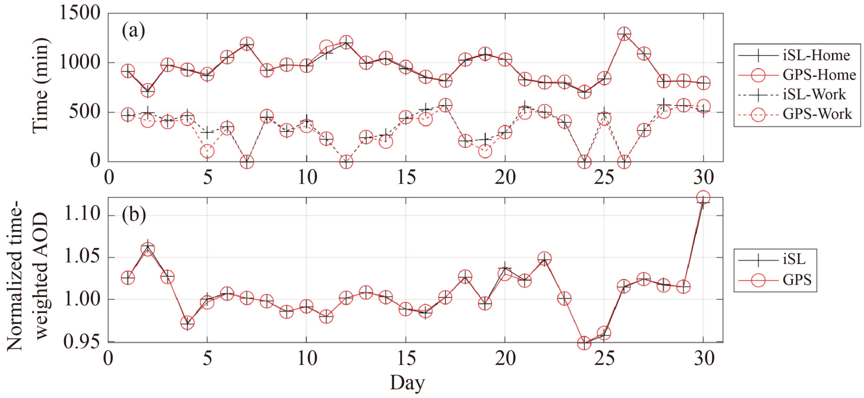


Fig. 2 The estimated time (minutes) the subject stayed at work and home during the study period, based on estimated GPS and iSL data (a); and daily variations of time-weighted AOD data, as estimated using iSL and GPS data. Data were normalized to AOD retrieved at the subject's home location (b).

Additionally, due to its nature, AOD values can not be directly and proportionally converted to ground-level particle concentrations. Therefore, we did not estimate a "percentage of improvement" if iSL data is used for exposure estimation purposes, as it would be misleading. We plan to use actual pollution concentration data to quantitively evaluate how iSL data can be used to improve exposure estimates in future studies.

Our results suggest that iSL data captured the most important microenvironments where the subject spent most of their time, as well as the daily variations of time spent at different locations. Using iSL data, the subject's daily time-weighted AOD, a surrogate for exposure to ambient PM, closely approximates those estimated using reference GPS data. Limited evidence from our exploratory survey suggest iSL data could be available for more than 70% of iPhone users, which translates to approximately 36% of the USA population (assuming 52% iPhone ownership (Statista, 2021)). Above results suggest that iSL data has great potential for characterizing an individual's historical time-activity pattern for not only air pollution health studies, but also for other research where time-activity pattern is important.

However, there remain many obstacles before iSL can be readily applied for characterizing individual time-activity patterns. The first obstacle is public acceptability, particularly privacy concerns (Goodwin, 2021). In our exploratory research, 22% of participants indicated that they are not willing to share their iSL data due mostly to privacy concerns. It is worth mentioning that some comments we received contain strong language, which highlights the sensitivities of iSL data. Although the general public's willingness to participate in passive mobile data collection is generally positive (Keusch et al., 2019), studies that focus specifically on iSL data remain scarce.

Another obstacle is the difficulty in data extraction. In the case of Google location data (GMLH), data was uploaded to the "cloud" by the user's smartphone and a user can download their raw GMLH data directly (Yu et al., 2019). iSL data, on the other hand, are stored locally on a user's iPhone. The user is able to view their iSL data but no options are available for the user to extract their raw iSL data. We are also not aware of any publicly available software/program that are capable of extracting iSL data on an iPhone running the latest iOS. If a user performed a local backup using an outdated version of iTunes (a software that can be used to manage iPhones), the local backup file may contain retrievable iSL data. Further, it is technically possible to match iSL images with maps to extract the center of a location, although we are not aware of the existence of such software. Apple provides developer tools, such as the Significant-Change Location Service, that enables background collection of iSL-like data prospectively (but not retrospectively).

There are also limitations to our study. First, the iSL data presented here are collected from a single individual. However, there is little evidence to suggest that iSL data from another individual would not perform as well in capturing the individual's time-activity patterns. Further, depending on local regulations, iSL data availability could differ among different countries or regions. Last, the length of retrospective availability is still unclear. Anecdotal evidence suggests that iSL data from several months ago may be removed, though the exact cut-off time is still unknown.

4 Conclusions

In this study, we evaluated the capability of iPhone Significant Location (iSL) data in capturing time-activity patterns by comparing iSL data collected from a single individual over a one month period against reference GPS logger data. We also conduct an online survey of over

300 participants to get a sense of user's willingness to share location data for research purposes. Since pollution exposure is largely dependent on time-activity patterns, finding available and historical location data for individuals is important for air pollution health studies. Our results show that iSL data captured the most important microenvironments where the subject spent most of their time (> 92%) during the study period. iSL data also performed well at capturing the daily variations of time spent at different microenvironments. These results suggest iSL data is able to reasonably record the spatial and temporal movement of an individual. For the purposes of estimating daily exposure to ambient PM pollution, the iSL data produces comparable results to pollution exposure estimates using other types of spatiotemporal location data, indicating that iSL data is suitable to use for air pollution health studies. Further, a considerable fraction of the general population in the USA may have iSL data available on their iPhone. However, privacy about sharing this data for research or other purposes remains a significant concern. Overall, the results suggest iSL data has great potential for characterizing an individual's historical time-activity patterns, and further studies are needed to investigate its utility for environmental health-based studies.

References

Apte J S, Messier K P, Gani S, Brauer M, Kirchstetter T W, Lunden M M, Marshall J D, Portier C J, Vermeulen R C H, Hamburg S P (2017). High-resolution air pollution mapping with Google street view cars: Exploiting big data. Environmental Science & Technology, 51(12): 6999–7008

Bell M L, Banerjee G, Pereira G (2018). Residential mobility of pregnant women and implications for assessment of spatially-varying environmental exposures. Journal of Exposure Science & Environmental Epidemiology, 28: 470–480

Bernstein J A, Alexis N, Barnes C, Bernstein I L, Bernstein J A, Nel A, Peden D, Diaz-Sanchez D, Tarlo S M, Williams P B (2004). Health effects of air pollution. Journal of Allergy and Clinical Immunology, 114(5): 1116–1123

GBD 2016 Risk Factors Collaborators (2017). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. Lancet, 390(10100): 1345–1422

Glasgow M L, Rudra C B, Yoo E H, Demirbas M, Merriman J, Nayak P, Crabtree-Ide C, Szpiro A A, Rudra A, Wactawski-Wende J, Mu L (2016). Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. Journal of Exposure Science & Environmental Epidemiology, 26(4): 356–364

Goodwin R (2021). How To Find, Track & Manage Your iPhone Location History: Everything You Need to Know. London: Know Your Mobile

Ipeirotis P G (2010). Demographics of mechanical turk. In: NYU Stern School of Business Research Paper Series. New York: New York University

Keusch F, Struminskaya B, Antoun C, Couper M P, Kreuter F (2019). Willingness to participate in passive mobile data collection. Public Opinion Quarterly, 83(S1): 210–235

Klepeis N E, Nelson W C, Ott W R, Robinson J P, Tsang A M, Switzer P, Behar J V, Hern S C, Engelmann W H (2001). The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. Journal of Exposure Science & Environmental Epidemiology, 11(3): 231–252

Lyapustin A, Wang Y, Laszlo I, Kahn R, Korkin S, Remer L, Levy R, Reid J (2011). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. Journal of Geophysical Research, D, Atmospheres, 116: D03211

Park Y M, Kwan M P (2017). Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. Health & Place, 43: 85–94

Statista (2021). Subscriber share held by smartphone operating systems in the United States from 2012 to 2021. New York: Statista

Yu H, Russell A, Mulholland J, Huang Z (2018). Using cell phone location to assess misclassification errors in air pollution exposure estimation. Environmental Pollution, 233: 261–266

Yu X, Stuart A L, Liu Y, Ivey C E, Russell A G, Kan H, Henneman L R F, Sarnat S E, Hasan S, Sadmani A, Yang X, Yu H (2019). On the accuracy and potential of Google Maps location history data to characterize individual mobility for air pollution health studies. Environmental Pollution, 252(Pt A): 924−930