Science of the Total Environment 506–507 (2015) 518–526

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Integrating smart-phone based momentary location tracking with fixed site air quality monitoring for personal exposure assessment

Jason G. Su *, Michael Jerrett, Ying-Ying Meng, Melissa Pickett, Beate Ritz

Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley 94720-7360, United States

HIGHLIGHTS

- Integrating location tracking and air quality monitoring to estimate personal exposure.
- Application of "topics models" to aggregate data in space-time and reduce data noise.
- Application of Deletion/Substitution/Addition modeling technique to avoid over-fitting.
- Identified the usefulness of using WiFi network only for personal location tracking.
- Identified typical issues associated with location tracking through smart phones.
- Personal exposure could be substantially different from home addressed based exposure.

ARTICLE INFO

Article history:
Received 20 August 2014
Received in revised form 21 October 2014
Accepted 4 November 2014
Available online 28 November 2014

Edited by: Prof. P. Kassomenos

Keywords:
Smart phone
Momentary location tracking
Personal exposure
Land use regression
Fixed site monitoring
Traffic-related air pollution

ABSTRACT

Epidemiological studies investigating relationships between environmental exposures from air pollution and health typically use residential addresses as a single point for exposure, while environmental exposures in transit, at work, school or other locations are largely ignored. Personal exposure monitors measure individuals' exposures over time; however, current personal monitors are intrusive and cannot be operated at a large scale over an extended period of time (e.g., for a continuous three months) and can be very costly. In addition, spatial locations typically cannot be identified when only personal monitors are used. In this paper, we piloted a study that applied momentary location tracking services supplied by smart phones to identify an individual's location in space–time for three consecutive months (April 28 to July 28, 2013) using available Wi-Fi networks. Individual exposures in space–time to the traffic-related pollutants Nitrogen Oxides (NO_X) were estimated by superimposing an annual mean NO_X concentration surface modeled using the Land Use Regression (LUR) modeling technique. Individual's exposures were assigned to stationary (including home, work and other stationary locations) and in-transit (including commute and other travel) locations. For the individual, whose home/work addresses were known and the commute route was fixed, it was found that 95.3% of the time, the individual could be accurately identified in space–time. The ambient concentration estimated at the home location was 21.01 ppb. When indoor/outdoor infiltration, indoor sources of air pollution and time spent outdoors were taken into consideration, the individual's cumulative exposures were 28.59 ppb and 96.49 ppb, assuming a respective indoor/outdoor ratio of 1.33 and 5.00. Integrating momentary location tracking services with fixed-site field monitoring, plus indoor–outdoor air exchange calibration, makes exposure assessment of a very large population over an extended time period feasible.

1. Introduction

Traffic-related air pollution (TRAP) is a major contributor to urban air pollution (Health Effects Institute, 2010). Epidemiological evidence identifies TRAP as a risk factor for adverse health outcomes, including preterm and low birth weight (Ghosh et al., 2012; Lee et al., 2013; Lee et al., 2008; Wilhelm et al., 2012), respiratory disease formation and exacerbation (Jerrett and Finkelstein, 2005; Spiric et al., 2012; Zhu et al., 2012), cardiovascular disease (Langrish et al., 2012; Raaschou-Nielsen et al., 2012) and premature mortality (Jerrett et al., 2011; Jerrett et al., 2009). Key pollutants of health concern emitted by vehicles include fine particulate matter, ultrafine particles, nitrogen oxides, diesel soot, and a variety of other gas- and particle-phase air contaminants (Kampa and Castanas, 2008). Data from government monitoring or special-purpose designed networks are usually modeled to derive air pollution surfaces so subjects within a study region can be

* Corresponding author. Tel.: +1 510 643 0102; Fax: +1 510 642 5815.
E-mail address: jasonsu@berkeley.edu (J.G. Su).

http://dx.doi.org/10.1016/j.scitotenv.2014.11.022
0048-9697/

assigned an exposure estimate for health outcome assessments. These air pollution modeling techniques include inverse distance weighting (IDW) and kriging (Brauer et al., 2008; Mercer et al., 2011), land use regression (LUR) modeling (Su et al., 2008; Su et al., 2010; Su et al., 2009b), spatiotemporal models such as Bayesian Maximum Entropy (De Nazelle et al., 2010; de Nazelle and Serre, 2006) and dispersion models (Beevers et al., 2012; Gulliver and Briggs, 2011; Lepeule et al., 2011). Typically, the residential address of a subject is used to assign air pollution exposures based on ambient concentrations, while the mobility of the subject and indoor–outdoor air exchange is ignored. Travel surveys, such as the California Household Travel Survey (http://www.californiatravelsurvey.com) and National Household Travel Survey (http://nhts.ornl.gov), provide detailed information on where a household travels; however, they do not identify spatially where people travel. Some research used modes of travel to study exposure from air pollution, typically through personal monitoring (Briggs et al., 2008; Sexton et al., 2007). These approaches are usually expensive to carry out and difficult to apply to a large population. Because of data storage limitations, they cannot be applied for extended periods of time, such as several months of continuous monitoring, to reflect activity patterns and exposure scenarios of individuals.

Participatory sensing is the process whereby individuals and communities use mobile phones and cloud services to collect and analyze systematic data (Estrin, 2010). Given the widespread availability of mobile phones, participatory sensing provides an opportunity for a paradigm shift in data gathering, especially for collecting time–space data at the individual level. Momentary location tracking services, such as Google Location Reporting & History, provide an optimal way to acquire high-fidelity, real-time location data through users' mobile phones. They support multi-modal localization, automatically switching between Global Positioning System (GPS), Wi-Fi and Global System for Mobile Communications (GSM) localization based on their availability (Constandache et al., 2009). User location data are continuously collected in the background to reduce power consumption. These momentary location tracking services make feasible the collection of mobility data for a large population over an extended time period.

In this study, we propose to estimate personal exposures in space–time through the combination of available Wi-Fi networks and fixed-site field exposure monitoring. Location histories were obtained from Google for three months, and space–time exposures were assigned based on an air pollution surface modeled through LUR, an approach that has been increasingly used in the past few years (Hoek et al., 2008). LUR is a relatively inexpensive and effective tool to model small area variations of air pollutant concentrations for epidemiological studies. With this combined experimental data, we try to address the following research questions: 1) Can smart phones be used to accurately track an individual's locations in space–time using Google's Location History? 2) Are the personal exposures estimated through smart phone momentary location tracking different from those based on home address alone?

2. Materials and methods

2.1. Momentary location tracking and data processing

We recruited a middle aged male researcher who commutes to work largely with a fixed-route and fixed modes of transport. The individual's space–time data from April 28 to July 28, 2013 (three months) were collected using smart phone momentary location tracking services provided by Google. Momentary location tracking by Google is a privacy-controlled application that, if enabled, allows Google to store a record of an individual location data through smart phones or other capable devices. Data from momentary location tracking can be retrieved and viewed through Google's Location History. Only the individual with the Google account can have access to the data collected by Google. Location History and location reporting can be switched on and off by the person who owns the Google account. The individual did not have a data plan from a mobile service provider, but had Wi-Fi networks at home and work, and obtained location data from free Wi-Fi networks outside of home and work. His space–time locations were thus provided by Wi-Fi, but not GPS. The individual provided the addresses of home and work, the means of commute, commute route, and the typical commute time spent each way. We used only one individual for the study, but a relatively long period for monitoring (three months in this study vs. one week in most literature) for the purpose of identifying all possible issues related to using smart phones for momentary location tracking using Google services. No data plan and GPS functions from smart phones were used in this study because of the interest in identifying the feasibility of momentary location tracking through public Wi-Fi networks, which have been expanding dramatically in the last few years.

For the location data collected, we first used a principal similar to topic models (Ferrari and Mamei, 2011) and aggregated the location history of the individual into two major categories: stationary and in-transit. The stationary locations included home, work and other stationary locations; in-transit locations included commute and other travel. The individual was considered to be stationary when the mapped location data were clustered around a location for more than 20 min (Mellegard et al., 2011). Since the individual was typically at work from Monday to Friday, we envisioned that those clustered locations were largely at the office or home addresses. Because of the availability of free Wi-Fi networks in some public places, the individual's locations might also be tracked outside of his home and office. The time spent in each stationary location was summed for each day.

While commuting, the individual's locations in space–time were interpolated based on (1) the time he left his home/office, (2) commute route and mode of transportation, and (3) the time he arrived at his home/office. We identified the individual's one way commute time to be close to one hour, with commuting by bike and BART (Bay Area Rapid Transit System) evenly splitting the commute time. The time he left his home/office each day was identified by the sudden disappearance of location tracking at his home/office for approximately 1 h; accordingly, the time he arrived at his home/office was detected when location tracking information reappeared after having been absent for almost 1 h. Spatial locations for the commute between office and home were interpolated using road segment centroids while biking and the six BART stations while on BART. We used BART stations as points of exposure since the exchange with ambient air while on a BART train was only available at BART stations. Time spent in each location point was linearly interpolated using the proportion of time spent traveling from one point to another during the one way trip (i.e., time of exposure). For other travel purposes, space in time interpolation was not conducted and exposure was not estimated because there was no information on mode of transportation other than for commute.

In an effort to reduce erroneous location reporting by cell phone towers, points that had a single location recording within a 1 km radius were reassigned. Locations were also reassigned if 8 of its 10 closest time stamp points (5 before- and 5 after-time stamp points) were in a location category other than the point being classified (e.g., home or work). These erroneous locations could occur if the phone had just been turned on or if the number of cell towers was not enough to triangulate its location. The location category of an erroneous point was reassigned to the location category of its nearest previous data point or the category to which ≥ 8 of the 10 neighboring points belonged. To address uncertainty or abnormal occasions, the mobile data were always processed the day after they were collected. Abnormal occasions included no location tracking information for an entire day, multiple disconnected spatial locations other than home/office and commute duration of more than 90 min.