GPS-based Microenvironment Tracker (MicroTrac) Model to Estimate Time-Location of Individuals for Air Pollution Exposure Assessments: Model Evaluation in Central North Carolina

Michael S. Breen,[1] Thomas C. Long,[2] Bradley D. Schultz,[1] James Crooks,[3] Miyuki Breen,[3] John E. Langstaff,[4] Kristin K. Isaacs,[1] Yu-Mei Tan,[1] Ronald W. Williams,[1] Ye Cao,[2] Andrew M. Geller,[5] Robert B. Devlin,[3] Stuart A. Batterman,[6] Timothy J. Buckley[1]

[1]National Exposure Research Laboratory, US EPA, RTP, NC, USA

[2]National Center for Environmental Assessment, US EPA, RTP, NC, USA

[3]National Health and Environmental Effects Research Laboratory, US EPA, RTP, NC, USA

[4]Office of Air Quality Planning and Standards, US EPA, RTP, NC, USA

[5]Immediate Office of the Assistant Administrator, US EPA, RTP, NC, USA

[6]Environmental Health Sciences, University of Michigan, Ann Arbor, MI, USA

Corresponding author: Michael Breen; 109 T.W. Alexander Drive, Mail Code: E205-02, Research Triangle Park, NC 27709, USA; Tel: (919) 541-9409; Fax: (919) 541-9444; Email: Breen.Michael@epa.gov

Running Title: GPS-based Microenvironment Tracker

Key words: GPS, time-activity, microenvironment, air pollution, health, exposure

Abstract

A critical aspect of air pollution exposure assessment is the estimation of the time spent by individuals in various microenvironments (ME). Accounting for the time spent in different ME with different pollutant concentrations can reduce exposure misclassifications, while failure to do so can add uncertainty and bias to risk estimates. In this study, a classification model, called MicroTrac, was developed to estimate time of day and duration spent in eight ME (indoors and outdoors at home, work, school; inside vehicles; other locations) from global positioning system (GPS) data and geocoded building boundaries. Based on a panel study, MicroTrac estimates were compared to 24 h diary data from nine participants, with corresponding GPS data and building boundaries of home, school, and work. MicroTrac correctly classified the ME for 99.5% of the daily time spent by the participants. The capability of MicroTrac could help to reduce the time-location uncertainty in air pollution exposure models and exposure metrics for individuals in health studies.

.

Introduction

Many epidemiologic studies have found associations between air pollutant concentrations measured at central-site ambient monitors and adverse health outcomes.[1] Using central-site concentrations as exposure surrogates, however, can lead to exposure misclassification due to time spent in various microenvironments (ME) with pollutant concentrations that can be substantially different from central-site concentrations. [2,3] This exposure misclassification can lead to uncertainty and bias to risk estimates.[2,3] To reduce exposure misclassification, we are developing an air pollution exposure model for individuals (EMI) in health studies.[4-6] The EMI predicts personal exposures based on outdoor concentrations, meteorology, questionnaire information (e.g., building characteristics, occupant behavior related to building operation and indoor sources), and time-location information. This study describes a critical aspect of EMI: the development and evaluation of a classification model, called MicroTrac, that estimates time of day and duration spent by individuals in eight ME (indoors and outdoors at home, work, school; inside vehicles; other locations) based on global positioning system (GPS) data and geocoded (geographic coordinates expressed as latitude and longitude) boundaries of buildings.

Exposure models can account for the variations in the time people spend in different locations by using time-weighted pollutant concentrations in each ME.[7] For population-level exposure assessments, exposure models rely on databases of time-activity diary information from other exposure studies,[8-10] such as the Consolidated Human Activity Database.[11] For individual exposure assessments, diaries from the study participants can be used.[4,12,13] However, diaries have limitations, including burden on participants, inaccuracies due to recall and reporting errors, and missing data.

To address the limitations of diaries, there is an increasing use of common mobile electronic devices such as smartphones, which often have embedded GPS receivers, and dedicated GPS dataloggers to collect personal time-location information.[14] Some advantages of GPS include automated logging, high time resolution, and an electronic format that does not require manual coding of handwritten diaries. However, manual processing of GPS data to determine time spent in different ME is limited due to several challenges, including (1) datasets that are large (potentially thousands of data points per person per day) and multidimensional (location, speed, time, satellite signal quality), (2) missing data due to no GPS signal reception while inside certain (e.g., steel/concrete) buildings, (3) GPS spatial inaccuracies due to temporal and spatial variations in the satellite geometry (i.e., spatial distribution of satellites used),[15,16] (4) localized transient spatial errors due to signal reflection (multipath errors) from nearby objects (e.g., water surfaces, buildings, hills, trees),[17] and (5) difficulty discriminating among certain ME (e.g., most detached homes, townhomes, and low-rise apartments in the United States are wooden structures with no substantial indoor/outdoor differences in satellite signal strength). The lack of a consistent and comprehensive solution to these problems has limited the use of GPS in personal exposure and health studies.[18] To address these limitations, we developed MicroTrac, an automated classification model for GPS data.

Using MicroTrac to determine the time spent in different indoor and outdoor locations can improve exposure estimates. For outdoor air pollutant concentrations C_{out} assumed to be at steady-state conditions (i.e., short-term changes of concentrations are considered negligible compared with long-term average concentrations), the steady-state exposure E_{true} can be described by:

$$E_{true} = f_{in} F_{inf} C_{out} + (1-f_{in})C_{out} \qquad (1)$$

where f_{in} is the fraction of time spent indoors and F_{inf} is the fraction of C_{out} that enters and remains airborne indoors (i.e., infiltration factor).[7] Setting F_{inf}=0.56 based on a reported median value for airborne particles (diameter=2.5 µm) for homes,[7] E_{true} for people who spend 30% (f_{in}=0.3) and 100% (f_{in}=1.0) of their time indoors are 0.87 and 0.56 times C_{out}, respectively. Using central-site air pollutant concentrations as an exposure surrogate, the exposure $E_{central}$ is C_{out}, which yields relative exposure differences ($|E_{central} - E_{true}| / E_{true}$) of 15% and 79% for f_{in}=0.3 and 1.0, respectively. This scenario analysis demonstrates that exposure differences are greater for people who spend more time indoors, and using MicroTrac to account for the time-location of individuals can substantially improve exposure assessments.

MicroTrac supports the recommendations of the National Research Council (NRC) report on exposure science in the 21st century[19] to link personal GPS and accelerometry (motion sensors) data from mobile electronic devices with exposure and lung dosimetry models, respectively. The NRC report recommends applying these sensors and models to reduce exposure and dose misclassifications for health studies, and to play a critical role in processing the large data from ubiquitous sensing networks, which collect personal exposure information using citizen scientists.

In this paper, we describe the development and evaluation of the MicroTrac. We first describe the panel study used to collect GPS data and create time-location diaries. We then describe the MicroTrac algorithm and method used for evaluation.

Methods

Time-Location Panel Study

A panel study consisting of nine participants was conducted by the National Exposure Research Laboratory of the U.S. Environmental Protection Agency (EPA). The participants lived in central North Carolina and worked at the EPA campus in Research Triangle Park, North Carolina. Each participant carried a GPS data logger (model BT-Q1000XT; Qstarz International, Taipei, Taiwan) for a continuous 24 h period. Seven participants collected GPS data on a workday (five in summer, two in fall), and two participants collected GPS data on a non-workday (one in summer, one in fall).

Before each 24 h deployment, the GPS memory was cleared using QTravel software (version 1.2; Qstartz International, Taipei, Taiwan) and the battery was fully charged. The GPS was programmed using QTravel to sample every 5 sec and to collect the time, position (latitude, longitude), speed, number of satellites used (NSAT), and position dilution of precision (PDOP, dimensionless value ≥ 1 that indicates accuracy of GPS position due to the satellite geometry; larger spatial distributions of satellites used yield smaller PDOP and more accurate positions).[16] GPS data were acquired and each sample was electronically marked in the GPS memory as either a scheduled or waypoint GPS sample. A scheduled GPS sample was collected automatically based on the programmed settings. A waypoint GPS sample was collected manually by pressing the waypoint button on the GPS, which was used to create time-location diaries. When transitioning between two ME, the participants pressed the waypoint button and manually recorded their corresponding starting and ending ME. The sampled data (approximately 17,280 scheduled samples per participant and 13-34 waypoint samples that varied across participants) were stored in the GPS memory during the 24 h sampling period, and then downloaded and stored using QTravel into two types of GPS files: a keyhole markup language (KML) file to view the GPS tracks as overlays in Google Earth (version 6.1.0.5001;

Google, Mountain View, CA, USA), and a text file for the classification algorithm described below.

The time-location diaries were used to determine the time of day and duration that participants spent in eight ME. The ME are: (1) indoors at the participant's home (Home-In); (2) outdoors near the participant's home (Home-Out); (3) indoors at the participant's workplace (Work-In); (4) outdoors near the workplace (Work-Out); (5) indoors at the school of the participant's children (School-In); (6) outdoors near the school (School-Out); (7) inside a vehicle (In-Vehicle); and (8) Other. Any time spent inside a vehicle, even if at Home-Out, Work-Out, or School-Out, was considered to be In-Vehicle. These eight ME are the same ME used by MicroTrac.

The accuracy of the time-location diaries (i.e., times when a participant transitioned between two ME) was verified manually for each participant's 24 h GPS data. For each waypoint GPS sample collected when entering a building that blocked GPS signal reception (e.g., work), the KML files, which overlay the scheduled and waypoint GPS samples in Google Earth, were used to verify that the waypoint sample occurred near the building boundary. For each waypoint GPS sample collected when entering or leaving a vehicle, the text files, which chronological list the scheduled and waypoint GPS samples, were used to verify that the waypoint sample occurred when speeds changed from driving speeds to walking speeds (e.g., In-Vehicle to Home-Out) or vice versa (e.g., Home-Out to In-Vehicle). Any suspected diary errors were discussed with the participant. If any diary error was confirmed, new 24 h GPS and diary data were collected.

Microenvironment Tracker Algorithm (MicroTrac)

We developed and evaluated an algorithm to determine which one out of the eight ME corresponds to the location of an individual at each GPS sampling time. Below, we describe the classification model, and then the temporal filtering of GPS speed samples, identification of GPS samples with poor signal quality (PSQ), and segmentation of building boundaries from aerial images. We then describe the method for evaluation of MicroTrac.

Microenvironment Classification Model

Our model is based on the time-course of GPS position (POS), speed (SPD), and signal quality (NSAT, PDOP); and geocoded boundaries of building rooftops for participant homes, workplaces, and schools. The model consists of eight parameters with seven parameters assigned values without using GPS data (i.e., no model fitting), and one parameter (PDOP threshold) assigned a value based on GPS data. We first describe the classification algorithm for time intervals with GPS samples, and then describe the algorithm for time intervals with missing GPS samples. The classification model was written and evaluated using MATLAB software (version R2011b; Mathworks, Natick, MA, USA).

Classification with GPS Samples and Building Boundaries

The details of the classification model are shown in the decision tree (Figure 1A) and described in the Supplementary Information. In summary, to classify a GPS sample as Home-In, there are three decision tree paths, which are unique pathways starting at the model inputs and ending at the classified ME. For the first decision tree path, the model determines whether the GPS position is within the home building boundary. To account for GPS spatial errors and since people tend to spend more time indoors than outdoors,[20] the model includes a 5 m spatial buffer for the home building boundary. The 5 m spatial buffer was assumed to be two times the GPS accuracy (2.5 m) specified by the manufacturer (model BT-Q1000XT; Qstarz International, Taipei, Taiwan). To account for transient GPS spatial errors greater than 5 m, the model includes a 15 s temporal buffer to determine whether any GPS position within 15 s is inside the spatial-buffered building boundary. Since the temporal buffer can introduce misclassifications when a person transitions from indoors to outdoors, a reasonably short duration (15 s) was assumed for the temporal buffer.

For the second decision tree path, a GPS sample is classified as Home-In when the GPS position is within 1 km of home and the GPS sample has PSQ, which can occur while indoors. The 1 km distance from home was assumed based on a reasonable surrounding area of home. To account for large transient spatial errors in the GPS position from multipath conditions that occur near structures that reflect GPS signals (e.g., tall buildings), the model uses a 15 s temporal buffer of the GPS position and PSQ data.

For the third decision tree path, a GPS sample is classified as Home-In when the GPS position is within 1 km of home, the GPS filtered speed (FSPD) is less than 18 km/h, and GPS sampling time is when there is no natural light outdoors (DARKNESS; period between astronomical dusk and dawn). The DARKNESS condition accounts for any GPS spatial errors that may occur when the GPS receiver is not moving for extended periods of time (e.g., sleeping). To account for multipath errors that can produce large transient spatial errors and large positive speed spikes, the FSPD condition is examined after the temporal-buffered GPS position and PSQ conditions. The 18 km/h speed threshold for the classifying as In-Vehicle was assumed based on an attempt to include slow moving vehicles (i.e., vehicle speeds slightly greater than 18 km/h) and to exclude people walking, running, and cycling. We assumed the typical speeds for walking, running, and cycling are less than 18 km/h.

For the work and school ME that have segmented building boundaries, the three paths described above for the home ME (Home-In, Home-Out) are used. One exception is the DARKNESS condition, which is not included for the work and school MEs.

If a GPS sample is not classified as a home, work, or school ME, the sample is classified as Other when PSQ_{15s} or FSPD<18 km/h. Otherwise, the GPS sample is classified as In-Vehicle.

Classification with Missing GPS Samples

The details of the classification model for missing data are shown in the decision tree (Figure 1B). When the GPS device does not receive a sufficiently strong signal from four or more satellites, no GPS sample is recorded. Since GPS signals can be attenuated by different building materials (e.g., concrete/steel), the model classifies a time interval with missing GPS samples as either Home-In, Work-In, School-In, or Other. The model first identifies any missing GPS samples by calculating the time difference between each pair of consecutive GPS samples. The number of missing GPS samples between consecutive GPS samples is the time difference divided by the GPS sampling period (5 s), then minus one. The model then classifies all consecutive missing GPS samples as the same ME. To classify a time interval with missing GPS samples as Home-In, the model determines whether any GPS sample within 60 s before the time interval with missing GPS samples is classified as Home-In or Home-Out. The 60 s duration was assumed for missing GPS samples based on a reasonable period that can account for possible misclassifications due to multipath errors immediately before satellite reception is lost when entering certain types of buildings. As shown in Figure 1B, a similar method is used to classify a time interval with missing GPS samples as Work-In or School-In. A time interval with missing GPS samples is classified as Other when no GPS sample within 60 s before the time interval is classified as Home-In, Work-In, or School-In.

Temporal Filtering of GPS Speed Samples

A GPS sample is classified as In-Vehicle based on exceeding a speed threshold. Since GPS speeds are at or near zero during brief periods due to stop lights, traffic, and other factors, we applied temporal filtering to the GPS speed data to remove negative transient speed spikes. The GPS speed is filtered across the entire time-course of GPS samples with a temporal filter.[21] The filter was designed to remove negative speed spikes with durations of approximately 2 min or less. The 2 min duration was assumed for the temporal filter based on reasonable waiting periods at traffic lights. The details of the filter are described in the Supplementary Information. This automatic filtering process is implemented in a conservative manner to produce an enhanced speed time-course with reduced negative transient spikes, while leaving any positive transient speed spikes and overall speeds relatively undisturbed.

Identification of GPS Samples with Poor Signal Quality

The PSQ from objects that obstruct the signal from satellites and decrease NSAT can occur outdoors near large tall structures (e.g., dense clusters of trees, buildings, hills) and indoors within steel/concrete buildings. Also, PSQ can occur when the time-varying positions of the satellites used by the GPS are not well distributed across the sky (i.e., poor satellite geometry), which increases PDOP.[16] For our classification algorithm, a GPS sample is considered PSQ when NSAT $\leq$ 4 or PDOP > 3.0. The NSAT threshold was set to 4 since a minimum of 4 satellites are needed to determine positions. The PDOP threshold was set to 3.0 since measured PDOP were consistently less than 2.5 under good signal quality conditions (NSAT > 8). When PSQ is detected, the GPS sample is classified as the indoor ME (Home-In, Work-In, School-In or Other) that corresponds to the location (home, work, school, or other) of the previously classified GPS sample.

Segmentation of Building Boundaries

To discriminate between GPS positions indoors and outdoors at home, work, and school, we created geocoded boundaries for these three types of buildings. In this panel study with nine participants, building boundaries were marked for nine homes (eight detached homes, one apartment), one workplace (five story office complex with five connected buildings), and two schools (one story detached buildings visited by two participants to drop off and pick up their children). The outline of each building's rooftop was manually segmented using the "Add Path" tool in Google Earth. For the evaluation of MicroTrac, the GPS tracks were not visible during segmentation. In Google Earth, the tilt angle was set for a view perpendicular to the ground, and the image zoom was adjusted to achieve a large display of the rooftop and a clear view of the rooftop edges. Before segmentation, the buildings were identified in the geocoded aerial images of Google Earth by entering the building addresses provided by the participants into Google Earth, and verified by using the KML GPS files to overlay the GPS tracks (displays placemarks for the GPS positions and line segments connecting the placemarks in chronological order) on the Google Earth images. After the buildings were identified and any GPS track overlays were removed, the rooftop boundaries were segmented and stored as KML building files for the classification model described below.

Evaluation of MicroTrac Performance

To quantitatively evaluate MicroTrac, we compared the estimated ME at each sampling time to its corresponding actual ME, as reported in the time-location diaries. To assess the daily differences between the actual and estimated time spent in each ME, we calculated the duration and percentage of day in each ME. To evaluate the model error for each ME, we determined the number of samples correctly identified as positive (true positive, TP) and negative (true negative, TN), and incorrectly identified as positive (false positive, FP) and negative (false negative, FN). We also identified the misclassified ME for each FP and FN. We then calculated the true positive fraction (TPF=TP/(TP+FN)) and false positive fraction (FPF=FP/(TN+FP)) to determine the sensitivity (TPF, proportion of actual positives correctly classified) and specificity (1-FPF, proportion of actual negatives correctly classified). The number of FP and specificity provide an assessment of the model's overestimation. The number of FN and sensitivity provide an indication of the model's underestimation. We also calculated the accuracy ((TP+TN)/(TP+TN+FP+FN)) and positive predictive value (PPV=TP/(TP+FP)) for each ME.

RESULTS

Summary statistics for each participant are provided for the day type, time spent in each ME, duration for missing GPS data and reason for missing data (i.e., GPS signal obstruction from building or time to reacquire satellite signal) in each ME, ME with occurrences of PSQ, and the eight types of locations (restaurant, store, park, friend's home, movie theater, doctor's office, library, fitness club) where participants spent time in the Other ME (Table 1). For workdays, there were long periods with missing GPS data at Work-In due to building obstruction of signal, and shorter periods of missing GPS data at Work-Out and In-Vehicle due to time for GPS to reacquire signal after leaving buildings that obstructed the signal. While at Other, missing GPS data occurred while at restaurants, stores, movie theater, and doctor's office. While at Home-In and School-In, there were no missing GPS data, but Home-In had several occurrences of GPS samples with PSQ. For the GPS data logger, the battery life (without recharging) and memory capacity were sufficient for each participant's 24 h sampling period. Also, there were no diary errors observed when we manually verified the accuracy of the diaries.

A comparison of the estimated and actual percentages of day in each ME is shown for each participant (Figure 2). The largest differences between actual and estimated percentage of day were 3.3% (underestimation) at Home-In and 3.4% (overestimation) at Home-Out for one participant (Figure 2C). All other differences were less than or equal to 1.0%. The model always slightly overestimated percentage of day at Work-In, School-In, and Other with median differences of 0.3%, 0.3%, and 0.4%, respectively, due to the time needed to reacquire GPS signal (typically 2-4 min) after leaving buildings (e.g., work, stores) that block satellite signals. The model always slightly underestimated percentage of day In-Vehicle with median difference of 0.7%, which was due to vehicle traveling below the speed threshold at the start and end of each trip (e.g., entering and leaving parking lots), and time needed to reacquire GPS signal while In-Vehicle after leaving buildings with no satellite reception.

A comparison of misclassifications (FN for underestimation and FP for overestimation) for each ME is shown across all participants (Table 2). Three MEs (Home-In, Work-Out, In-Vehicle) had greater FN than FP (underestimation); the other five MEs had greater FP than FN (overestimation). There were misclassifications between Home-In and Home-Out, between Work-In and Work-Out, and between School-In and School-Out. For In-Vehicle, there were FN from the other ME, and one FP when Home-Out and School-Out. For Other, there were no FN, and FP when In-Vehicle.

We also evaluated the model by calculating the sensitivity, specificity, accuracy, and PPV across all participants for each ME (Table 2). Sensitivities and specificities less than 100% correspond to overestimation and underestimation of the ME, respectively. The lowest sensitivities were 60.4% and 73.5% at Work-Out and School-Out, respectively, while the other sensitivities were greater than 81.0%. The specificities were greater than or equal to 99.0%. The accuracy across all samples was 99.5%. The lowest accuracy was 98.9% both at Home-In and Home-Out. The lowest PPV was 63.0% at School-Out, and the highest PPV was 100.0% In-Vehicle.

We also compared the model performance on workdays and non-workdays. The sensitivities on workdays and non-workdays were 98.8% and 98.8% at Home-In, 47.1% and 87.8% at Home-Out, 86.8% and 89.3% for In-Vehicle, and 100.0% and 100.0% for Other, respectively.

DISCUSSION

Our goal was to develop and evaluate a model to classify GPS samples into eight ME from GPS data and building boundaries. The daily estimated ME closely correspond to the actual ME with a mean accuracy of 99.5%. These results demonstrate the capability of using GPS data with MicroTrac to estimate time spent in various ME, and support the feasibility of integrating MicroTrac into individual air pollution exposure models (e.g., EMI).[6] Since MicroTrac automates the processing of GPS data for ME classification, it could also provide a method to support the potentially large GPS data from widespread sensor networks of citizen scientists, as recommended by the NRC report on exposure science in the 21st century.[19]

We can compare the model used to classify GPS samples as indoors and outdoors with previously reported ones. In Adams et al.,[22] using a geocoded building boundary of a home to classify GPS samples as Home-In did not perform well (64.4% sensitivity). In Elgethun et al.,[23] boundaries of homes and each building entered by participants were used to classify as Home-In and Other-In, respectively. Boundaries of each yard at home were used to classify as Home-Out. The sensitivities were 84.8% (Home-In), 18.3% (Home-Out) and 45.6% (Other-In). In Wu et al.,[24] a rule-based classifier identified intervals of GPS samples with speeds less than 3 km/h for a minimum of 1 min (static clusters). A static cluster was then classified as indoors based on various criteria (e.g., time includes midnight, duration greater than 2 h, positions within 50 m of home). The sensitivities were 84.1% (indoors) and 51.7% (outdoors).

Our model has several novel features for classifying GPS samples as indoors and outdoors. First, MicroTrac uses 5 m spatial buffering of the building boundaries to account for the spatial inaccuracy of the GPS device. Second, our model uses a 15 s temporal buffer of GPS position and PSQ data to account for multipath conditions that occurs near structures that reflect GPS signals (tall buildings, dense clusters of trees). Third, for positions within 1 km of home and speeds less than 18 km/h, the astronomical dusk-to-dawn period is used to account for possible positional drift errors of GPS that can occur when the GPS is stationary for several hours (e.g., sleeping). Using these unique features, the sensitivities of MicroTrac for indoor ME were 98.8% (Home-In), 99.9% (Work-In), 93.1% (School-In); and for outdoor ME were 81.4% (Home-Out), 60.4% (Work-Out), and 73.5% (School-Out).

In Adams et al.,[22] an alternative method classified GPS and personal temperature samples as Home-In and School-In for GPS positions within 30 m of the building centroid and for temperatures above 15.55 °C (60 °F). The sensitivities for indoor ME were 99.9% (Home-In), 99.8% (School-In); and for outdoor ME were 65.4% (Home-Out), and 84.6% (School-Out) during the winter in Colorado. Indoor/outdoor classification based on a simple temperature threshold is limited to days with substantial indoor-outdoor temperature differences,[22] and can have limited temporal resolution due to the thermal response time of the monitor following a temperature change. In Kim et al.,[25] NSAT was used for indoor/outdoor classification, and classified samples as Home-In when NSAT was less than 9 and positions were within 40 m of home. The sensitivities were 89.3% for Home-In and 86.4% for Other-In. In Tandon et al.,[26] the signal to noise ratio (SNR) was used for indoor/outdoor classification, and GPS samples were classified as outdoors when the total SNR of all satellites in view exceeded 250. The sensitivity was 82% for children outdoors at child care centers. For indoor/outdoor classification, we tried various thresholds based on indoor-outdoor temperature, NSAT, total SNR of satellites, but none were reliable. In Tandon et al.,[26] personal light samples were used for indoor/outdoor classification, and classified as outdoors for light intensities above 110 lux. The sensitivity was 74% for children outdoors at child care centers. We decided not to use a light sensor since wearing the device outside of clothing and uncovered for extended periods of time to avoid obstructing the light can be problematic, as described in Tandon et al.[26]

We can compare our method used to classify GPS samples as transit (i.e., when not at home, work, or school) with previously reported ones. In Adams et al.,[22] GPS samples were simply classified as transit when not classified at home or school with a sensitivity of 95.3%. In Elgethun et al.,[23] GPS samples were classified as transit when GPS speeds exceeded 18 km/h with a sensitivity of 29.6%. In Wu et al.,[24] GPS samples were classified into two types of transit ME (In-Vehicle, Out-Walking) based on GPS speed and geocoded roadway data. Moving periods were identified based on various criteria that include individual speeds above 15 km/h, consecutive samples with speeds above 2.5 km/h, and positions within 10 m of a roadway. Moving periods were then classified as In-Vehicle when second highest speed exceeded 10 km/h and median speed exceeded 5 km/h with a sensitivity of 72.1%; otherwise, Out-Walking with a sensitivity of 68.4%. In Kim et al.,[25] GPS samples classified as outdoors (based on NSAT threshold) were further classified as transit when GPS speeds exceed 9 km/h with a sensitivity of 45.3%. In our model, MicroTrac classified GPS samples as In-Vehicle when filtered speeds exceed 18 km/h, and obtained a sensitivity (87.6%) higher than previously reported ones.

Unlike previous reports, our model compares speeds to a threshold only after evaluating positions with a spatial buffer (GPS position is within 1km of a building) and a temporal buffer (within 15 s), which helps reduce misclassifications due to any large speed errors from multipath interference that can occur soon before entering a large concrete/steel building. In addition, the temporal filtering of the GPS speed samples can reduce misclassifications while In-Vehicle by accounting for the reduced speed or stopping of the vehicle due to various conditions (e.g., traffic congestion, traffic signals, stop signs, intersections of roads with high traffic volume). The conservatively implemented temporal filter can effectively eliminate negative transient speed spikes, while leaving positive transient speed spikes and the overall speeds across time relatively unaffected. The enhanced filtered speed time course allows for reduced number of misclassifications since the removal of negative speed spikes can reduce the number of false negatives while In-Vehicle.

We can also compare our model used to classify intervals with missing GPS data with previously reported ones. In Adams et al.,[22] intervals with missing GPS data were classified as Home-In or School-In for sampling times during pre-defined home and school periods, respectively. Otherwise, the intervals with missing GPS data were classified as the same ME as the previously classified GPS sample immediately before satellite reception was lost. In Elgethun et al.,[23] intervals with missing GPS data were classified as Home-In. Our model uses a 15 s temporal buffer for the previously classified GPS samples before satellite reception was lost. The temporal buffer is a key feature of our model since misclassifications can occur soon before satellite reception is lost due to multipath errors at the entrance of large buildings. A temporal buffer can help account for these multipath errors and reduce the misclassifications of intervals with missing GPS data.

Our model can be practically implemented for various applications. First, MicroTrac can be integrated within exposure models (e.g., EMI) to estimate exposure metrics for epidemiologic analyses and risk assessments.[6] Second, using MicroTrac with personal GPS devices, movement sensors (e.g., accelerometers), air pollutant monitors, and health monitors in exposure and health effect studies will allow scientists to link the location and activity of study participants with air pollution concentrations and health effects. Using smartphones with these data collection capabilities will facilitate and expand the use of MicroTrac in these studies, and will support community applications of MicroTrac such as alerting susceptible populations (e.g., asthmatics) to behavior or activities that may compromise their health. Since the manual segmentation of the building boundaries does not require any specialized training and the Google Earth software is free and publicly available, MicroTrac could be used by both researchers and citizen scientists.

Finally, MicroTrac's ability to classify time spent inside vehicles can be used to correct physical activity information from accelerometers when inside moving vehicles.

MicroTrac could also be applied to improve the time-activity pattern data used for population-level exposure assessments. With a high percentage of the US population using GPS-enabled smartphones, large sets of GPS data collected with low participant burden could be classified in various ME by MicroTrac to increase the sample size and update the older diary data in the time-activity databases (e.g., Consolidated Human Activity Database),[11] which are used for population-level exposure assessments. These updates are needed for regions with substantial time-activity pattern changes due to various factors such as large economic, demographic, or population changes. Also, the high accuracy of MicroTrac can help improve the accuracy of the time-activity databases that have been developed from diaries with possible recall and reporting errors.

Our model evaluation was based on the time-location of adult participants on workdays and non-workdays, which live in single family homes and a low-rise apartment building, and work in a multi-story office building that are all located in suburban areas. We expect similar results in homes, schools, and workplaces with similar building characteristics and located in suburban or rural areas. The ability of MicroTrac to predict the time-location of individuals in urban areas with high-density high-rise buildings, and individuals with more dynamic location patterns than working adults (e.g., children) needs to be investigated. To address these limitations, we plan to perform additional model evaluation using other panel studies, such as the Near-Road Exposures and Effects of Urban Air Pollutants Study (NEXUS) in Detroit, Michigan with 139 school-age children with asthma.[4] In our study, we evaluated the model with data in central North Carolina since we plan to apply MicroTrac for cohort health studies with adult participants living and working in the same suburban areas.

There are some limitations to our model. First, the model cannot account for time spent outdoors within 1 km radius of home between astronomical dusk and dawn due to the DARKNESS condition (e.g., walking in neighborhood during the night). However, the model does account for time spent inside vehicles within 1 km radius of home between dusk and dawn. Second, outdoor home locations less than 5 m from edge of rooftop (e.g., decks, patios) are included within the 5 m buffer of the segmented building boundary and cannot be distinguished from the indoor living space of the home. Third, attached structures with a roof (e.g., attached garages, porches) often cannot be distinguished in aerial images from the indoor living space of a home, and are included within the segmented building boundary. Fourth, we were unable to classify GPS samples as Other-In and Other-Out, but combined these two ME into one (Other). In addition, the model does not use geocoded roadway data to determine time spent on specific roads (e.g., interstate highways). MicroTrac could be modified to incorporate this additional information. However, this would substantially increase the model's complexity, limit the use of the model to those with specialized expertise and software (e.g., geographic information systems), and is beyond the scope of this study. Finally, the manual segmentation of boundaries for the buildings of interests (e.g., home, work, and school) may not be feasible for large cohort studies (e.g., 100,000 children in the National Childrens' Study).[27] In these cases, it is possible that automated image segmentation algorithms could be implemented.[28] Even with these limitations, MicroTrac is an improvement from previously reported methods, and its few input requirements can facilitate its use for various applications.

The pilot study used to evaluate MicroTrac has some limitations. The panel study of nine participants is not large, and all participants were working adults that lived in central North Carolina. We plan to further evaluate MicroTrac with larger cohort studies, which include: children with asthma that are living in Detroit,[4] Michigan, and older adults with cardiovascular disease that are living in North Carolina.

There are some key strengths of the pilot study. The GPS data are from a prospective panel study using real-world activity patterns, instead of scripted activities. Also, the participant diaries used to evaluate MicroTrac are high quality since the participants understood the study goals, followed a strict protocol, and used the clock on the GPS device to record electronically the time when transitioning to a different ME. In addition, the accuracy of the diaries was verified manually. Obtaining high quality diaries can be a substantial challenge for large cohort studies.

We conclude that our study demonstrates the feasibility of using MicroTrac to estimate time of day and duration spent in eight ME from GPS data and building boundaries. Results show that for seven workdays and two non-workdays, the estimated and actual time spent in the ME closely corresponds. This capability could help reduce the time-location uncertainty in air pollution exposure models used to predict exposure metrics for individuals in health studies and for citizen scientists. MicroTrac could also help improve the time-activity databases used for population-level exposure assessments.

ACKNOWLEDGMENTS

We thank Jennifer Richmond-Bryant and Karen Wesson for review comments and helpful suggestions. Miyuki Breen was supported by the North Carolina State University/Environmental

Protection Agency Cooperative Training Program in Environmental Sciences Research, Training Agreement CT833235-01-0 with North Carolina State University. Ye Cao was supported by the Oak Ridge Institute for Science and Education (ORISE)/Environmental Protection Agency Research Participation Training Program [Interagency Agreement DW-89-92298301-0]. Procedures involving humans were conducted in accordance with U.S. EPA Order 1000.17 Change A1 (Policy and Procedures on Protection of Human Research Subjects). These procedures were reviewed and approved by the US EPA Human Subjects Research Review Official. Although this manuscript was reviewed by the U.S. Environmental Protection Agency and approved for publication, it may not necessarily reflect official Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

SUPPLEMENTARY INFORMATION

The supplementary information includes additional details on the classification model with GPS samples and building boundaries, and the temporal filtering of GPS speed samples.

REFERENCES

1. Integrated Science Assessment for Particulate Matter (Final Report); EPA/600/R-08/139F; United States Environmental Protection Agency: Washington, DC, 2009.

2. HEI panel on the health effects of traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects. HEI Special Report 17. Health Effects Institute, Boston, MA, 2010.

3. National Research Council. Epidemiology and Air Pollution. The National Academies Press: Washington, DC, 1985.

4. Vette A, Burke J, Norris G, Landis M, Batterman S, Breen M et al. Community Action Against Asthma Steering Committee. The near-road exposures and effects of urban air pollutants study (NEXUS): study design and methods. Sci. Total. Environ. 2013; 448: 38-47.

5. Breen MS, Breen M, Williams RW, Schultz BD. Predicting residential air exchange rates from questionnaires and meteorology: model evaluation in central North Carolina. Environ. Sci. Technol. 2010; 44: 9349-9356.

6. United States Environmental Protection Agency. Exposure Model for Individuals (EMI). Available at http://www.epa.gov/heasd/emi.

7. Wallace L, Willams R. Use of personal-indoor-outdoor sulfur concentrations to estimate the infiltration factor and outdoor exposure factor for individual homes and persons. Environ Sci Technol 2005; 39: 1707-1714.

8. Burke JM, Zufall MJ, Ozkaynak H. A population exposure model for particulate matter: case study results for PM2.5 in Philadelphia, PA. J Expo Anal Environ Epidemiol 2001; 11: 470-489.

9. United States Environmental Protection Agency. Total Risk Integrated Methodology (TRIM) - Air Pollutants Exposure Model Documentation (TRIM.Expo / APEX, Version 4.4) Volume I:

User's Guide, EPA-452/B-12-001a; Office of Air Quality Planning and Standards, US EPA: Research Triangle Park, NC. Available at: http://www.epa.gov/ttn/fera/human_apex.html.

10. United States Environmental Protection Agency. Total Risk Integrated Methodology (TRIM) - Air Pollutants Exposure Model Documentation (TRIM.Expo/APEX, Version 4.4) Volume II: Technical Support Document, EPA-452/B-12-001b; Office of Air Quality Planning and Standards, US EPA, Research Triangle Park, NC. Available at: http://www.epa.gov/ttn/fera/human_apex.html.

11. McCurdy T, Glen G, Smith L, Lakkadi Y. The national exposure research laboratory's Consolidated Human Activity Database. J Expo Anal Environ Epidemiol 2000; 10: 566-578.

12. Williams R, Suggs J, Rea A, Leovic K, Vette A, Croghan C et al. The Research Triangle Park particulate matter panel study: PM mass concentration relationships. Atmos Environ 2003**;** 37: 5349-5363.

13. Williams R, Rea A, Vette A, Croghan C, Whitaker D, Stevens C et al. The design and field implementation of the Detroit Exposure and Aerosol Research Study. J Expo Sci Environ Epidemiol 2009; 19: 643-659.

14. Chaix B, Meline J, Duncan S, Merrien C, Karusisi N, Perchoux C et al. GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference? Health & Place 2013; 21:46-51.

15. Beekhuizen J, Kromhout H, Huss A, Vermeulen R. Performance of GPS-devices for environmental exposure assessment. J Expo Sci Environ Epidemiol 2013; 23:498-505.

16. Misra P, Burke BP, Pratt MM. GPS performance in navigation, Proceedings of the IEEE, 1999; 87: 65-85.

17. GE L, Han S, Rizos C. Multipath mitigation of continuous GPS measurements using an adaptive filter. GPS Solutions, 2000**;** 4: 19-30.

18. Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. Environ Health Perspect 2004; 11: 1007-1015.

19. National Research Council. Exposure science in the 21st century: a vision and a strategy. The National Academies Press: Washington, DC, 2012.

20. Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P et al. The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants. J Expo Anal Environ Epidemiol 2001; 11: 231-252.

21. Breen MS, Breen M, Butts K, Chen L, Saidel GM, Wilson DL. MRI-guided thermal ablation therapy: model and parameter estimates to predict cell death from MR thermometry images. Ann Biomed Eng 2007; 35: 1391-1403.

22. Adams C, Riggs P, Volckens J. Development of a method for personal, spatiotemporal exposure assessment. J Environ Monit 2009; 11: 1331-1339.

23. Elgethun K, Yost M, Fitzpatrick C, Nyerges T, Fenske R. Comparison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time-location patterns. J Expo Sci Environ Epidemiol 2007; 17: 196-206.

24. Wu J, Jiang C, Houston D, Baker D, Delfino R. Automated time activity classification based on global positioning system (GPS) tracking data. Environ Health 2011; DOI: 10.1186/1476-069X-10-101.

25. Kim T, Lee K, Yang W, Yu S. A new analytical method for the classification of time-location data obtained from the global positioning system (GPS). J Environ Monit 2012; 14: 2270-2274.

26. Tandon P, Saelens B, Zhou C, Kerr J, Christakis D. Indoor versus outdoor time in preschoolers at child care. Am J Prev Med 2013; 1:85-88.

27. Landrigan P, Trasande L, Thorpe L, Gwynn C, Lioy P, D'Alton M et al. The National Children's Study: a 21-year prospective study of 100000 american children. Pediatrics 2006; 118: 2173-2186.

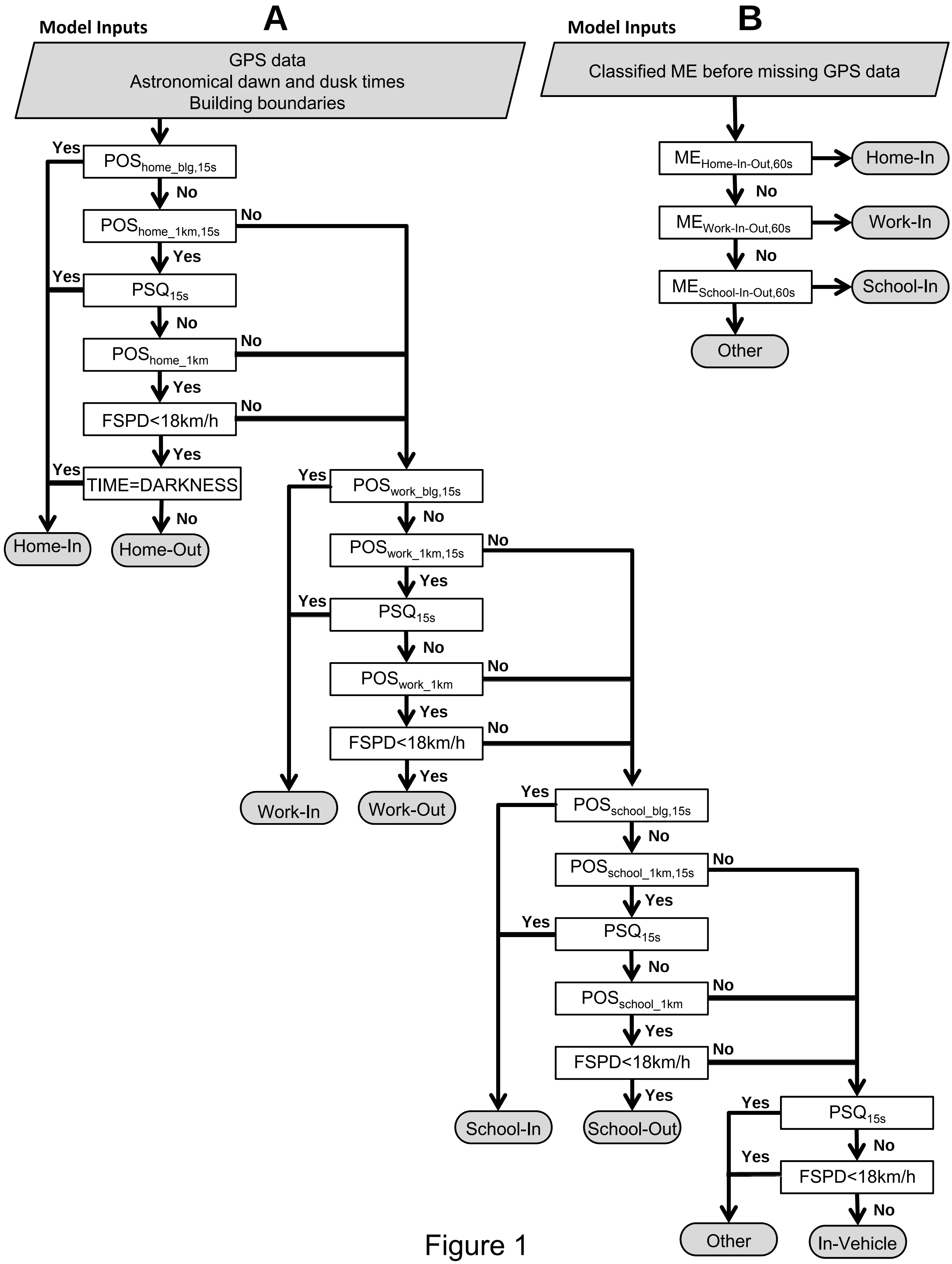
28. Chen D, Zhang L, Li J, Liu R. Urban building roof segmentation from airborne lidar point clouds. Int J Remote Sensing 2012; 33: 6497-6515.

FIGURE LEGENDS

Figure 1. Decision tree of classification model for GPS samples and building boundaries (A), and for time intervals with missing GPS samples (B). For classification of GPS samples (A), decisions for home ME (Home-In, Home-Out) include: any GPS position within 15 s inside 1 km radius from centroid of home ($POS_{home\_1km,15s}$), any GPS position within 15 s inside home building boundary with 5 m buffer ($POS_{home\_blg,15s}$), time interval between astronomical dusk and dawn (DARKNESS), any sample within 15 s with poor signal quality (PSQ_{15s}), current position inside 1 km radius of home (POS_{home_1km}), and current filtered speed (FSPD) < 18 km/h. For work ME (Work-In, Work-Out), decisions include: any position within 15 s inside 1 km radius from centroid of work (POSwork_1km,15s), any position within 15 s inside work building boundary with 5 m buffer ($POS_{work\_blg,15s}$), any sample within 15 s with poor signal quality (PSQ_{15s}) when number of used satellites ≤ 4 or position dilution of precision > 4, current position inside 1 km radius from centroid of work (POS_{work_1km}), and current filtered speed (FSPD) < 18 km/h. For school ME (School-In, School-Out), decisions include: any position within 15 s inside 1 km radius from centroid of school ($POS_{school\_1km,15s}$), any position within 15 s inside school building boundary with 5 m buffer ($POS_{school\_blg,15s}$), any sample within 15 s with poor signal

quality (PSQ_{15s}), current position inside 1 km radius from centroid of school (POS_{school_1km}), and current filtered speed (FSPD) < 18 kph. For Other and In-Vehicle ME, decisions include: any sample within 15 s with poor signal quality (PSQ_{15s}), and current filtered speed (FSPD) < 18 kph. For classification of time intervals with missing GPS samples (B), decisions include: any ME within 60s before missing time interval that is classified as Home-In or Home-Out ($ME_{Home\text{-}In\text{-}Out,60s}$), Work-In or Work-Out ($ME_{Work\text{-}In\text{-}Out,60s}$), School-In or School-Out ($ME_{School\text{-}In\text{-}Out,60s}$).

Figure 2. Estimated and actual percentage of day in the eight ME for each participant (A-I). The nine participants (A-I) correspond to participants 1-9 shown in Table 1, respectively. Percentage values are shown for each bar for quantitative comparison between estimated and actual differences.



A
Model Inputs
GPS data
Astronomical dawn and dusk times
Building boundaries
POS home_blg,15s
POS home_1km,15s
PSQ15s
POS home_1km
FSPD<18km/h
TIME=DARKNESS
Home-In
Home-Out
POS work_blg,15s
POS work_1km,15s
PSQ15s
POS work_1km
FSPD<18km/h
Work-In
Work-Out
POS school_blg,15s
POS school_1km,15s
PSQ15s
POS school_1km
FSPD<18km/h
School-In
School-Out
PSQ15s
FSPD<18km/h
Other
In-Vehicle
Yes
No
B
Model Inputs
Classified ME before missing GPS data
ME Home-In-Out,60s
ME Work-In-Out,60s
ME School-In-Out,60s
Home-In
Work-In
School-In
Other

Figure 1

Table 1. Microenvironment characteristics by participant and duration of missing GPS data

Participant	Day type	Time spent in microenvironment (duration of missing GPS data)[a] (h)								
		Home-In	Home-Out	Work-In	Work-Out	School-In	School-Out	In-Vehicle	Other	Total[m]
1	Workday	13.18[b]	0.08	9.51[b] (9.46)[c]	0.06 (0.04)[d]	0.26	0.06	0.86 (0.03)[d]	0.00	24.01 (9.53)
2	Workday	12.31[b]	0.15	7.73[b] (7.71)[c]	0.30 (0.05)[d]	0.00	0.00	1.18	2.35[b,e,f]	24.01 (7.76)
3	Workday	12.93[b]	0.24	9.53[b] (9.36)[c]	0.08 (0.03)[d]	0.00	0.00	0.90	0.37[b,f,g] (0.05)[c,d]	24.05 (9.44)
4	Workday	15.41	0.00	5.34[b] (5.20)[c]	0.25 (0.06)[d]	0.00	0.00	0.99 (0.04)[d]	2.03[e,f] (0.24)[c]	24.02 (5.54)
5	Workday	14.65	0.02	7.56[b] (7.40)[c]	0.12 (0.03)[d]	0.20	0.12	1.34 (0.04)[d]	0.00	24.01 (7.47)
6	Workday	14.07[b]	0.07	8.88[b] (8.85)[c]	0.43 (0.20)[d]	0.00	0.00	0.65 (0.02)[d]	0.00	24.10 (9.07)
7	Workday	11.47[b]	0.08	7.72[b] (7.69)[c]	0.16 (0.05)[d]	0.00	0.00	2.06 (0.04)[d]	2.43[f,h,i] (1.60)[c,d]	23.91 (9.39)
8	Non-Workday	16.55[b]	3.46	0.00	0.00	0.00	0.00	1.77	2.24[f,j] (0.76)[c,d]	24.02 (0.76)
9	Non-Workday	16.07	0.00	0.00	0.00	0.00	0.00	1.87	6.10[e,h,,k,l] (0.13)[c]	24.04 (0.13)

[a]No parentheses indicates no missing GPS data, [b]Occurrence of GPS samples with poor signal quality (number of satellites used ≤ 4 or position dilution of precision > 3.0), [c]Missing GPS data due to entering building that obstructed satellite signal, [d]Missing GPS data due to time for GPS to reacquire satellite signal after leaving building that obstructed signal, [e]Time spent at restaurant, [f]Time spent at store, [g]Time spent at park, [h]Time spent at friend's house, [i]Time spent at movie theater, [j]Time spent at doctor's office, [k]Time spent at library, [l]Time spent at fitness club , [m]Individual microenvironment times may not sum to total due to rounding

Table 1

Figure 2

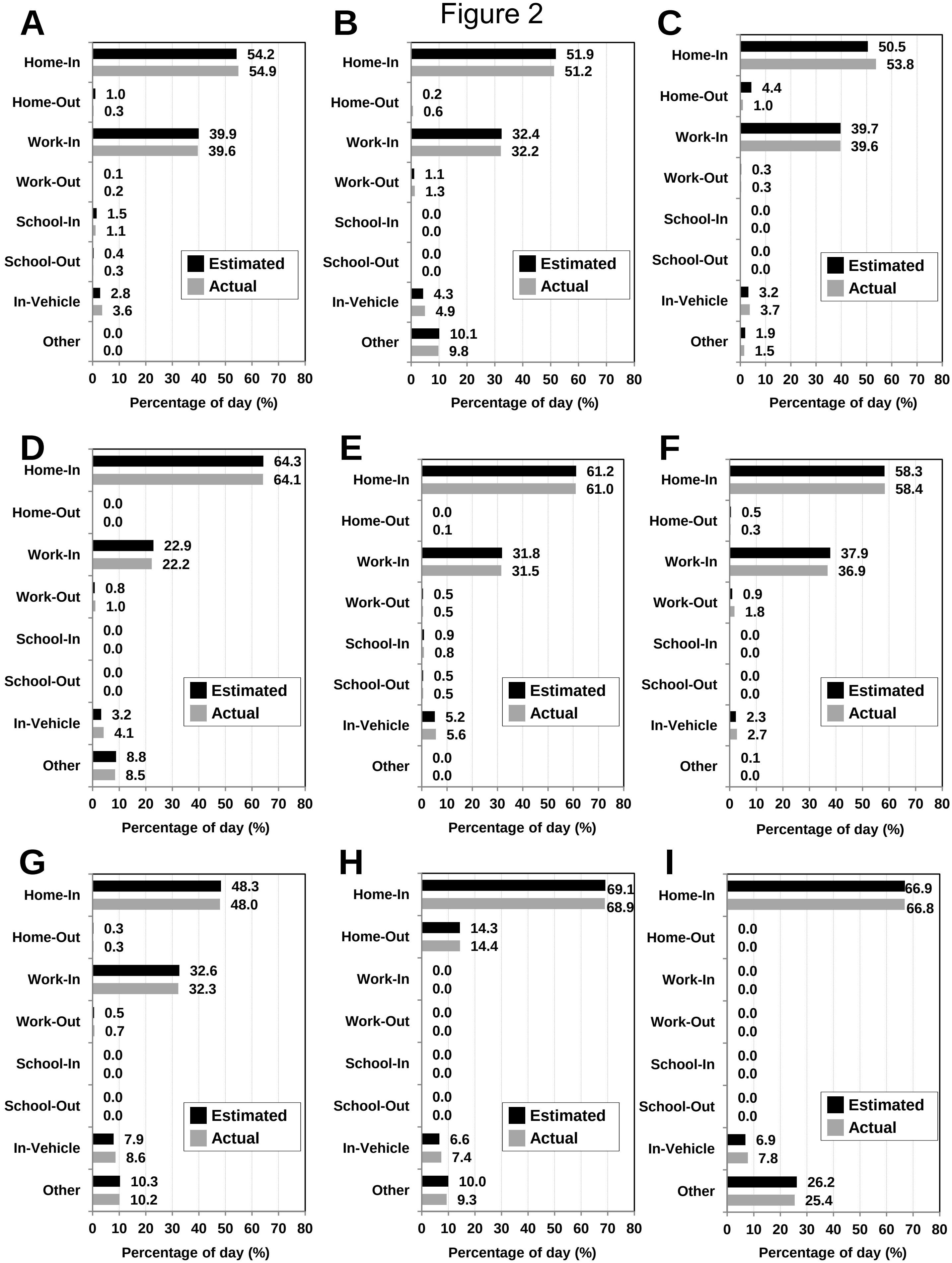

A
Home-In 54.2 54.9
Home-Out 1.0 0.3
Work-In 39.9 39.6
Work-Out 0.1 0.2
School-In 1.5 1.1
School-Out 0.4 0.3
In-Vehicle 2.8 3.6
Other 0.0 0.0
Estimated
Actual
Percentage of day (%)
B
Home-In 51.9 51.2
Home-Out 0.2 0.6
Work-In 32.4 32.2
Work-Out 1.1 1.3
School-In 0.0 0.0
School-Out 0.0 0.0
In-Vehicle 4.3 4.9
Other 10.1 9.8
Estimated
Actual
Percentage of day (%)
C
Home-In 50.5 53.8
Home-Out 4.4 1.0
Work-In 39.7 39.6
Work-Out 0.3 0.3
School-In 0.0 0.0
School-Out 0.0 0.0
In-Vehicle 3.2 3.7
Other 1.9 1.5
Estimated
Actual
Percentage of day (%)
D
Home-In 64.3 64.1
Home-Out 0.0 0.0
Work-In 22.9 22.2
Work-Out 0.8 1.0
School-In 0.0 0.0
School-Out 0.0 0.0
In-Vehicle 3.2 4.1
Other 8.8 8.5
Estimated
Actual
Percentage of day (%)
E
Home-In 61.2 61.0
Home-Out 0.0 0.1
Work-In 31.8 31.5
Work-Out 0.5 0.5
School-In 0.9 0.8
School-Out 0.5 0.5
In-Vehicle 5.2 5.6
Other 0.0 0.0
Estimated
Actual
Percentage of day (%)
F
Home-In 58.3 58.4
Home-Out 0.5 0.3
Work-In 37.9 36.9
Work-Out 0.9 1.8
School-In 0.0 0.0
School-Out 0.0 0.0
In-Vehicle 2.3 2.7
Other 0.1 0.0
Estimated
Actual
Percentage of day (%)
G
Home-In 48.3 48.0
Home-Out 0.3 0.3
Work-In 32.6 32.3
Work-Out 0.5 0.7
School-In 0.0 0.0
School-Out 0.0 0.0
In-Vehicle 7.9 8.6
Other 10.3 10.2
Estimated
Actual
Percentage of day (%)
H
Home-In 69.1 68.9
Home-Out 14.3 14.4
Work-In 0.0 0.0
Work-Out 0.0 0.0
School-In 0.0 0.0
School-Out 0.0 0.0
In-Vehicle 6.6 7.4
Other 10.0 9.3
Estimated
Actual
Percentage of day (%)
I
Home-In 66.9 66.8
Home-Out 0.0 0.0
Work-In 0.0 0.0
Work-Out 0.0 0.0
School-In 0.0 0.0
School-Out 0.0 0.0
In-Vehicle 6.9 7.8
Other 26.2 25.4
Estimated
Actual
Percentage of day (%)

Table 2. Statistics for model evaluation across all participants

Actual ME	Estimated ME (number of samples)								Actual duration (number of samples)	Estimated duration (number of samples)	FN[a] (number of samples)	FP[b] (number of samples)	Sens[c] (%)	Spec[d] (%)	Acc[e] (%)	PPV[f] (%)
	Home-In	Home-Out	Work-In	Work-Out	School-In	School-Out	In-Vehicle	Other								
Home-In	90057	1110	0	0	0	0	0	0	91167	90732	1110	675	98.8	99.0	98.9	99.3
Home-Out	548	2402	0	0	0	0	1	0	2951	3580	549	1178	81.4	99.2	98.9	67.1
Work-In	0	0	40478	34	0	0	0	0	40512	41027	34	549	99.9	99.5	99.6	98.7
Work-Out	0	0	395	603	0	0	0	0	998	747	395	144	60.4	99.9	99.7	80.7
School-In	0	0	0	0	308	23	0	0	331	414	23	106	93.1	99.9	99.9	74.4
School-Out	0	0	0	0	34	97	1	0	132	154	35	57	73.5	100.0	99.9	63.0
In-Vehicle	127	68	154	110	72	34	7332	471	8368	7334	1036	2	87.6	100.0	99.3	100.0
Other	0	0	0	0	0	0	0	11179	11179	11650	0	471	100.0	99.7	99.7	96.0

[a]False negatives indicate underestimation, [b]False positives indicate overestimation, [c]Sensitivity indicates underestimation, [d]Specificity indicates overestimation, [e]Accuracy, [f]Positive predictive value

Table 2

GPS-based Microenvironment Tracker (MicroTrac) Model to Estimate Time-Location of Individuals for Air Pollution Exposure Assessments: Model Evaluation in Central North Carolina

Michael S. Breen,[1] Thomas C. Long,[2] Bradley D. Schultz,[1] James Crooks,[3] Miyuki Breen,[3] John E. Langstaff,[4] Kristin K. Isaacs,[1] Yu-Mei Tan,[1] Ronald W. Williams,[1] Ye Cao,[2] Andrew M. Geller,[5] Robert B. Devlin,[3] Stuart A. Batterman,[6] Timothy J. Buckley[1]

[1]National Exposure Research Laboratory, US EPA, RTP, NC, USA

[2]National Center for Environmental Assessment, US EPA, RTP, NC, USA

[3]National Health and Environmental Effects Research Laboratory, US EPA, RTP, NC, USA

[4]Office of Air Quality Planning and Standards, US EPA, RTP, NC, USA

[5]Immediate Office of the Assistant Administrator, US EPA, RTP, NC, USA

[6]Environmental Health Sciences, University of Michigan, Ann Arbor, MI, USA

The supplementary information consists of 4 pages (S1-S4).

Supplementary Information

Classification of GPS Samples and Building Boundaries

The details of the classification model are shown in the decision tree (Figure 1A). There are three paths to classify a GPS sample as Home-In. For the first path, we determine whether the GPS position is within the home building boundary. To account for the spatial inaccuracy of the GPS position (2.5 m root mean square; Qstarz International, Taipei, Taiwan) and since people tend to spend more time indoors than outdoors, we include a 5 m spatial buffer for the home building boundary. In addition, transient GPS positional errors larger than 5 m can occur, which can displace actual home-indoor GPS receiver locations beyond the 5 m buffer of the home building boundary. To account for these transient spatial inaccuracies in the GPS positions, we include a +/- 15 s temporal buffer (+/- 3 GPS samples), and determine whether any of the seven GPS positions (3 previous, 1 current, and 3 subsequent GPS samples) are within the home building boundary or within 5 m of the boundary ($POS_{home\_bld,15s}$). If condition $POS_{home\_bld,15s}$ is true, the GPS sample is classified as Home-In.

For the second path, a GPS sample is classified as Home-In when the GPS position is near the home and the GPS sample has poor signal quality, which can occur while indoors. To examine this condition, we determine whether the GPS position is within 1 km of the centroid of the home building boundary. To account for large transient spatial errors in the GPS position from multipath errors that can occur soon before entering a steel/concrete building (which is then often followed by complete loss of GPS signal), we include a +/- 15 s temporal buffer (+/- 3 GPS samples), and determine whether any of the seven GPS positions (3 previous, 1 current, and 3 subsequent GPS samples) are within 1 km of the home radius ($POS_{home\_1km,15s}$). Next, we determine whether the GPS sample has poor signal quality. To account for GPS signal attenuation that can occur briefly before complete loss of GPS signal when entering a steel/concrete building, we include a +/- 15 s temporal buffer (+/- 3 GPS samples), and determine whether any of the seven GPS samples (3 previous, 1 current, and 3 subsequent GPS samples) have poor signal quality (PSQ_{15s}). If conditions $POS_{home\_1km,15s}$ and PSQ_{15s} are true, the GPS sample is classified as Home-In.

For the third path, a GPS sample is classified as Home-In when the PSQ_{15s} is not true, current GPS position is within 1 km of the home but not within the home building boundary (POS_{home_1km}), filtered GPS speed (FSPD) is below 18 km/h, and GPS sampling time is when there is no natural light outdoors (DARKNESS; period between astronomical dusk and dawn). The DARKNESS condition accounts for any GPS spatial errors that may occur when the GPS receiver is not moving for extended periods of time (e.g., sleeping). The FSPD condition accounts for time spent In-Vehicle while near the home (within 1 km). To account for multipath errors that can produce both large transient spatial errors and large positive speed spikes, the FSPD condition is examined after the temporal-buffered GPS position and PSQ conditions. When the DARKNESS condition is not true in the third path, a GPS sample is classified as Home-Out.

Temporal Filtering of GPS Speed Samples

The GPS speed is filtered across the entire time-course of GPS samples with a temporal morphological filter. The morphological filter consists of a circular structuring element with a temporal radius of 1 min. The structuring element was used to perform a morphological closing operation to remove negative transient speed spikes. This filter design allowed for the removal of negative speed spikes with a duration of approximately 2 min or less.