Journal of Environmental Monitoring

Cite this: *J. Environ. Monit.*, 2012, **14**, 2270

www.rsc.org/jem

TECHNICAL NOTE

A new analytical method for the classification of time–location data obtained from the global positioning system (GPS)

Taehyun Kim,[a] Kiyoung Lee,*[a] Wonho Yang[b] and Seung Do Yu[c]

Received 8th March 2012, Accepted 23rd May 2012
DOI: 10.1039/c2em30190c

Although the global positioning system (GPS) has been suggested as an alternative way to determine time–location patterns, its use has been limited. The purpose of this study was to evaluate a new analytical method of classifying time–location data obtained by GPS. A field technician carried a GPS device while simulating various scripted activities and recorded all movements by the second in an activity diary. The GPS device recorded geological data once every 15 s. The daily monitoring was repeated 18 times. The time–location data obtained by the GPS were compared with the activity diary to determine selection criteria for the classification of the GPS data. The GPS data were classified into four microenvironments (residential indoors, other indoors, transit, and walking outdoors); the selection criteria used were used number of satellites (used-NSAT), speed, and distance from residence. The GPS data were classified as indoors when the used-NSAT was below 9. Data classified as indoors were further classified as residential indoors when the distance from the residence was less than 40 m; otherwise, they were classified as other indoors. Data classified as outdoors were further classified as being in transit when the speed exceeded 2.5 m s^{-1}; otherwise, they were classified as walking outdoors. The average simple percentage agreement between the time–location classifications and the activity diary was 84.3 ± 12.4%, and the kappa coefficient was 0.71. The average differences between the time diary and the GPS results were 1.6 ± 2.3 h for the time spent in residential indoors, 0.9 ± 1.7 h for the time spent in other indoors, 0.4 ± 0.4 h for the time spent in transit, and 0.8 ± 0.5 h for the time spent walking outdoors. This method can be used to determine time–activity patterns in exposure-science studies.

Introduction

Exposure assessment is a critical part of determining the health impact of environmental contaminants. Since direct measurement of personal exposure can be time consuming and expensive, it can be estimated by microenvironmental concentration and duration of exposure. The accuracy in recording the time and location influences the accuracy of the exposure calculation. There are several methods used for acquiring time–location information, including time–activity diaries, questionnaires, and observation.[1] However, these methods are heavily affected by recall abilities and the voluntary participation of subjects. The data quality can be questionable due to differences among individuals' motivation to participate and their recall of events.

The shortcomings of the traditional time–location measures may be potentially addressed by the global positioning system (GPS).[2] GPS technology provides the convenience of recording time–location information in the form of coordinates (latitude

[a]Seoul National University Environmental Health, 1 Gwanak-ro Gwanak-gu, Seoul 151-742, Republic of Korea. E-mail: cleanair@snu.ac.kr
[b]Catholic University of Daegu, Department of Occupational Health, Kyongsan, Kyongsanbukdo 712-702, Republic of Korea
[c]National Institute of Environmental Research, Incheon, Republic of Korea

Environmental impact

The use of a new geological technology can improve the quality of an exposure assessment study. The global positioning system (GPS) has been suggested as an alternative way to determine time–location patterns in exposure assessment. This study provided a new analytical method of classifying time–location data obtained by the GPS. The proposed analytical method can determine residential time in various microenvironments accurately without recall bias and active participation of subjects. With the collection of accurate time–location data, recently developed time-resolution monitoring devices can have mutual benefit for detailed exposure profiles.

and longitude) with minimal human intervention, thus eliminating error from human sources. Because of the high received signal strength of GPS signals outdoors, GPS is commonly favored for outdoor activities.[3] The received signal strength can be significantly attenuated in indoor environments by the loss of line-of-sight, reflection from and absorption by structural materials, and blockage by conductive materials through the Faraday cage effect.[4] As a result, the indoor signal strength is generally below the sensitivity of the receiver, resulting in low rate of reception and reduced accuracy.[5]

GPS signals are based on a code system developed by the National Marine Electronics Association (NMEA). The NMEA code system includes various sentences such as GGA (GPS fix data), GPGSV (GPS satellites in view), and GPRMC (GPS recommended minimum data) sentences. Each NMEA sentence has a different composition. The GPGGA sentence shows essential fix data, which provide data on 3D location and accuracy. The GPGSV sentence indicates overall satellite data; it contains satellite information codes such as the number of satellites (NSAT) in view, NSAT used, satellite identification (SID), elevation, azimuth, and signal-to-noise ratio (SNR). The GPPMC sentence shows recommended minimum data.[6]

The GPS is used to identify a location based on longitude, latitude, and elevation. A GPS receiver needs at least three satellite signals to calculate a location. However, the NSAT is generally greater when there is direct line-of-sight to the satellites. The speed of the GPS receiver can be determined by using the carrier phase-derived Doppler measurements or the receiver-generated Doppler measurements.[7] Therefore, the speed of the GPS receiver can be accurately determined[8] even when the positioning accuracy of the GPS receiver is low. Some commercial GPS receivers can receive more NMEA codes, such as used-NSAT and speed.

The purpose of this study was to develop an analytical method of using GPS data to classify time–location information. Time–location data obtained by GPS were compared with an activity diary while a field technician simulated various activities, and selection criteria were identified for classification of the time–location data. In addition to the geocoordinates and time, additional data from the NMEA codes were used.

Methods

A field technician carried a GPS monitor (GPS 741, Ascen, Korea) while simulating scripted activities. In this study, activity patterns of four groups were simulated. The four groups were teenage students (3 days), elderly persons (6 days), housewives (3 days), and male service workers (6 days). The actual time of each activity was also recorded using a pre-printed activity diary. The experiments were conducted in Seoul, Korea.

Time–location data were collected by the GPS monitor. The data storage capacity of this GPS device is approximately 125 000 points of geospatial coordinate data. The GPS monitor was set to record geological data once every 15 s. The size of the GPS device was 7.2 cm long, 4.7 cm wide and 2.0 cm high, and it weighed 64 g. It was carried in a cross-bag during the sampling time. The GPS monitor could be operated for approximately 25 hours before recharging.

The GPS monitor was configured to collect the NMEA code before the field application. The collected GPS data were downloaded onto a computer and analyzed. The raw speed data were collected in units of km h^{-1}, and the NSAT data included both used-NSAT and viewed-NSAT. The speed data were converted to units of m s^{-1} and the used-NSAT data were extracted from NSAT (used/viewed).

The GPS data were classified into four microenvironments based on used-NSAT, distance from residence, and speed. Indoors and outdoors were determined by used-NSAT. When the used-NSAT value was below a certain number, the point was classified as indoors. When the used-NSAT was equal to or greater than a certain number, the point was classified as outdoors. Indoor status was confirmed only when the used-NSAT was maintained for at least 3 min. The appropriate threshold of used-NSAT to accurately classify indoor *versus* outdoor status was determined from the time–activity pattern manually recorded in the diary. The diary record was then used to verify the accuracy of the GPS data classification. To quantify the accuracy of the GPS data analysis, the percentage agreement with the diary record and the kappa coefficient were calculated.

The distance between a GPS data point and the residential geocode was used to further classify indoors into residential indoors and other indoors. The geocode of the technician's residence was confirmed using Google Earth. The distance between a GPS data point and the residence was divided into 10 m intervals. The appropriate threshold distance from the residential geocode to classify residential indoors *versus* other indoors was determined by comparison with the activities recorded in the diary. The data for 3 days were excluded from this analysis because the technician intentionally stayed outside the residence for a certain amount of time while following the pre-scripted activities.

Speed was utilized to further classify outdoor points into transit or walking. When the speed was above a certain threshold, the point was classified as transit; otherwise, it was classified as walking. Walking status was finalized only when the speed was maintained for at least 3 min. The appropriate threshold speed to accurately classify transit *versus* walking status was determined from the time–activity pattern recorded in the diary.

The average used-NSAT and speed in each microenvironment were compared by Student's *t*-test using SAS 9.1 software package (SAS Institute Inc., Cary, NC). A *P*-value less than 0.05 was considered statistically significant. Simple percentage agreement (the proportion of cases that were accurately predicted) and kappa coefficient were calculated to compare the accuracy of the GPS data analysis for each microenvironment, as determined from the time–activity diary. The kappa coefficient is a more robust measure than simple percentage agreement because it takes into account the agreement occurring by chance.[9]

Results

A total of 18 daily time–activity simulations were conducted. The average daily number of GPS data points collected was 5086 $\pm$ 2913. The average amount of time spent in the residential indoors, other indoors, transit, and walking outdoors microenvironments were 14.9 $\pm$ 3.7, 6.7 $\pm$ 3.3, 0.9 $\pm$ 0.6, and 1.6 $\pm$ 0.4 h, respectively. The average total difference between the scripted and actual times spent in activities was less than 30 min per day. The differences between the scripted and actual times spent in the residential indoors, other indoors, and outdoor microenvironments were 8.6 ± 12.7, 8.6 ± 10.9, and 11.3 ± 12.8 min per day, respectively.

The used-NSAT value was used to classify indoors *versus* outdoors. Although used-NSAT ranged from 1 to 12, for 95.5% of the entire dataset, the used-NSAT ranged from 4 to 10. The average used-NSAT when the technician was outdoors, as determined from the activity diary, was 8.4 ± 1.3, whereas the average used-NSAT indoors was 5.7 ± 2.1. The used-NSAT outdoors was significantly greater than that of indoors ($P < 0.0001$). Fig. 1 shows the average accuracy of the indoor and outdoor classification determined from used-NSAT. The accuracy of the indoors *versus* outdoors classification was greatest when a used-NSAT threshold of 9 was used; with this classification criterion, the simple percentage agreement was 90.9 ± 4.6%, and the kappa coefficient was 0.50.

When the technician stayed in the residential indoor microenvironment, the total number of recorded GPS data points was 46 512. Of these 46 512 data points, 98.7% were within 100 m of the residence. Table 1 shows the number of data points at intervals of 10 m from the residence and the proportion of points correctly classified as residential indoors. When a distance less than 40 m from the residence was classified as residential indoors, the simple percentage agreement was 97.7%. With a distance of 50 m from the residence as the criterion threshold, the simple percentage agreement was reduced to 95.1%. Therefore, a distance of 40 m from the residence was selected as the criterion to classify residential indoors.

When the technician stayed outdoors, the data points were classified as transit or walking. The average speed of transit and walking were 4.4 ± 5.4 m s^{-1} and 0.7 ± 2.0 m s^{-1}, respectively. The average speed of transit was significantly greater than that of walking ($P < 0.0001$). Fig. 2 shows the average accuracy of transit *versus* walking classification according to speeds. When the criterion threshold for transit *versus* walking was a speed of 2.5 m s^{-1}, the simple percentage agreement was 90.4 ± 7.1%, and the kappa coefficient was 0.79.

The entire dataset was classified into the four microenvironments using a used-NSAT value of 9, a distance of 40 m from the residence, and an outdoor speed of 2.5 m s^{-1}. The simple percentage agreement of the classified data with the activity diary

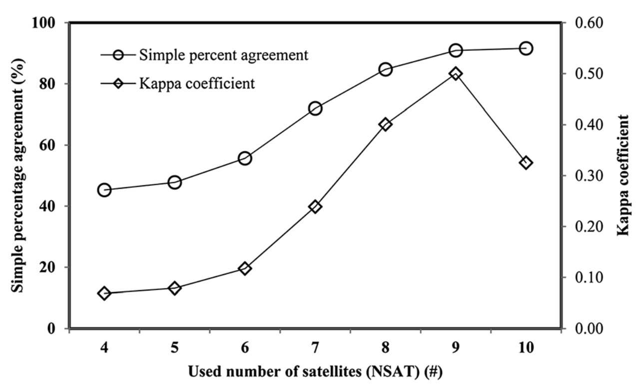

Fig. 1 Accuracy of indoor and outdoor classification by used number of satellites (NSAT).

Table 1 Characteristics of residential indoor GPS data by distance from residence

Distance range (m)	Number of residential indoor points (#)	Number of total points (#)	Proportion of residential indoor points (%)
0–10	9064	9281	97.7
0–20	25 764	26 329	97.9
0–30	34 545	35 265	98.0
0–40	38 539	39 435	97.7
0–50	40 528	42 629	95.1

was 84.3 ± 12.4%, and the kappa coefficient was 0.71. The accuracies of the classification into each microenvironment are shown in Fig. 3. The simple percentage agreements for the data classified as residential indoors, other indoors, transit, and walking outdoors were 89.3 ± 23.6%, 86.4 ± 30.3%, 45.3 ± 28.4%, and 48.9 ± 25.0%, respectively. The accuracies for residential indoors and other indoors were relatively greater than those for transit and walking. The average differences between the time spent in each microenvironment according to the time diary *versus* the GPS results were 1.6 ± 2.3 h for residential indoors, 0.9 ± 1.7 h for other indoors, 0.4 ± 0.4 h for transit, and 0.8 ± 0.5 h for walking.

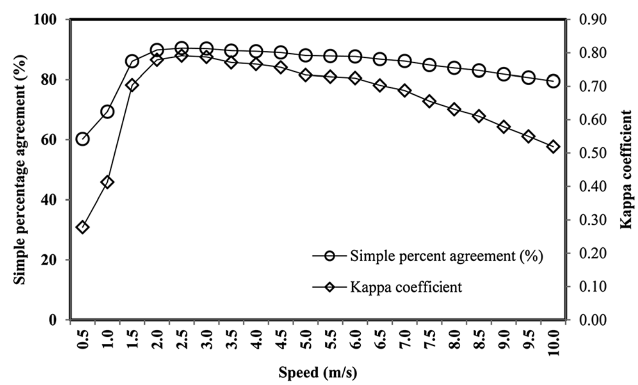

Fig. 2 Accuracy of the classification as transit or walking based on the speed used as the criterion threshold.

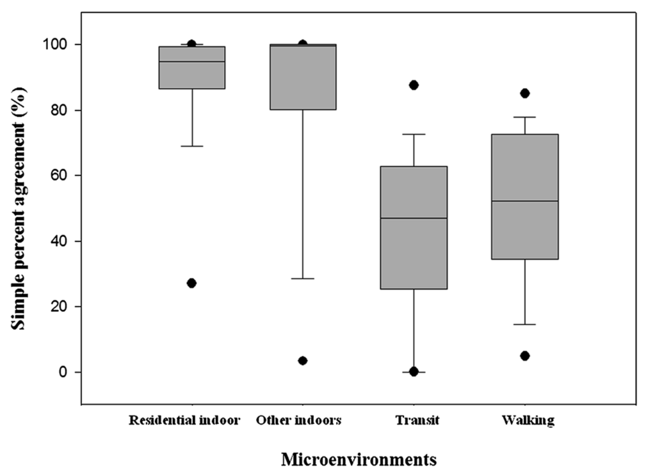

Fig. 3 Distribution of the simple percentage agreement between the classification of data into the four microenvironments and the time–activity diary.

Discussion

The time–activity patterns of four population groups were simulated to evaluate the accuracy of GPS as an analytical method of determining time activity. The scripted activities were obtained from the Time Use Survey from Statistics Korea. Simulation by a technician was deemed to be the best approach to collect time–activity data with high accuracy because asking the general public to manually record all their time–activity information would be extremely burdensome. The technician accurately recorded time–activity information, and the actual time–activity pattern was close to the script. Although four population groups were simulated in this study, the times spent in each of the four microenvironments (residential indoors, other indoors, transit, and walking outdoors) were similar to patterns in the Korean population.[10]

This is the first study to apply used-NSAT as a criterion to classify GPS location data as indoors. The NSAT is reduced when the GPS device is indoors. When various values of used-NSAT were used as the criterion threshold for classifying GPS location data as indoors, it was found that a used-NSAT value of 9 gave the best accuracy. Although the simple percentage agreement was slightly higher with a used-NSAT of 10, the kappa coefficient was decreased. Therefore, a used-NSAT below 9 indicated that the time activity should be classified as indoors.

When the GPS device was located indoors, geocodes tended to be scattered. When the GPS data indicated a location within 40 m of the residence, 97.7% of the time, the GPS device was actually in the residential indoor microenvironment. Based on these data, geocodes within 40 m of the residence were classified as residential indoors. As only one house was used in this monitoring, further investigation may be needed to confirm this classification criterion. For example, the distance criterion might be affected by urban or rural status, characteristics of the building, the type of construction materials (*e.g.*, wood, concrete, *etc.*), and locations of windows.

Using distance from residence as a criterion, data classified as indoors (based on used-NSAT) were further classified as residential indoors or other indoors. Although distance from residence should be evaluated under various conditions, the concept of classification by the distance from the residence seems reasonable. This approach requires a residential geocode, which can be obtained from an accurate residential address. This approach can just as easily be applied to other buildings or to the work place. If a detailed classification of other indoor activity patterns is needed, the distance to the other indoor location can be supplied. Knowing the geocodes of other indoor environments, daily GPS data can be used to determine times spent in many other indoor microenvironments.

Transit time is often important for exposure assessment. In this study, a speed of 2.5 m s^{-1} was applied to further classify GPS data already classified as outdoors (on the basis of used-NSAT) as transit or walking. We applied a 3 min rule because transit could be temporarily stopped; if a speed less than 2.5 m s^{-1} lasted more than 3 minutes, it was classified as walking; otherwise, it was classified as transit. The transit and walking classifications were about 90% accurate. A previous time–location study classified speeds over 5 m s^{-1} as transit without any real justification.[11] Air pollution levels in transit are generally high, and these can contribute substantially to daily personal exposure.[12] Exposure during transit can be affected by travel modes.[13] Accurate methods to identify travel modes are important to understand exposure on transit. Thus, much research is focused on the automatic identification by GPS of travel modes, such as walking, bicycle, passenger car, and transit.[14–17]

Time–location classification with four microenvironments was substantially accurate when our criteria were applied to GPS data. The simple percentage agreement was approximately 84.3%, and the kappa coefficient was 0.71. This indicates substantial agreement.[18] The accuracy of data that were classified as outdoors was relatively lower (55.4 $\pm$ 26.0%) than that of data classified as indoors (94.9 $\pm$ 4.8%). As the average time apportioned to outdoors (2.4 h) was much less than that apportioned to indoors (21.6 h), the lower outdoor classification accuracies may not have a significant impact on the overall analyses of time–activity data. The low accuracy for data classified as outdoors may be due to a delay in reception when the GPS device is restarted to receive satellite signals. If GPS reception stopped for less than 1 h, restarting of data recording usually commenced within 15–40 s. However, if GPS reception stopped for more than 1 h, it took 2–3 min or longer to restart data recording, and it was found to take as long as 15 min if the receiver was moving fairly rapidly.[4]

Use of NMEA codes improved the classification of GPS data into microenvironments. Geocodes by themselves were not sufficient to classify GPS location data as indoors or outdoors. Given that not all GPS devices can record NMEA codes, selection of a GPS device with the capability for NMEA code reception is necessary. GPS devices can be survey, mapping, or consumer grade by performance and can vary in price.[19] The GPS device used in this study was of consumer grade, and the price was less than US $100. However, for an exposure-science study, a commercial grade device would be appropriate because large sample sizes are usually needed for such studies.

The new analytical GPS method presented here was conducted under limited conditions. If a more accurate data-analysis method is needed, various conditions such as urban *versus* rural, variations in characteristics of buildings, existence of windows, and weather conditions, and the type of GPS device should be evaluated. This study was conducted in Seoul, the biggest city in Korea, where there are many high-rise buildings. Thus, blocking of the GPS satellite signals may have been more frequent than in a rural area. Building structures may result in different satellite signal reception rates and positional errors. Buildings with concrete structures more frequently block GPS satellite signals than do buildings with wood frames. Satellite signals may be unable to penetrate into indoor spaces that have no windows. Although the GPS satellite signal reception rates were high, positioning accuracy was reduced indoors, even with windows.[5,20,21] During the sampling period, wet weather conditions, including thick clouds, rain, lightning, and thunder, were observed for 14 days, and these conditions may have had an effect. Finally, the GPS devices of various manufacturers have varying performance levels.[22]

Conclusion

A method was developed to adequately analyze GPS data by comparing the time–activity patterns of four simulated population groups with GPS data. Time–location information was accurately classified into four microenvironments. Used-NSAT from the NMEA codes was used to classify time–location data as indoors or outdoors. Data classified as indoors were further classified as residential indoors or other indoors by using the distance range of scattered geocodes as a discrimination criterion. Data classified as outdoors were further classified as transit or walking by using speed as the discrimination criterion. This GPS analytical method can be easily and accurately applied to surveys of time–activity patterns in air pollution-exposure and epidemiological studies.

Acknowledgements

This study was partially funded by the Dr and Mrs Michael M. Park Scholarship of Seoul National University (2011-2-247) and was partially supported by the National Institute of Environmental Research, Korea.

References

1 N. C. G. Freeman and S. S. de Tejada, *Chemosphere*, 2002, **49**, 979–992.
2 P. Larsson, L. Burlin, E. Jakobsson and K. Henriksson-Larsen, *J. Sports Sci.*, 2002, **20**, 529–535.
3 Y. Schutz and A. Chambaz, *Eur. J. Clin. Nutr.*, 1997, **51**, 338–339.
4 P. Stopher, C. FitzGerald and J. Zhang, *Transport. Res. C Emerg. Tech.*, 2008, **16**, 350–369.
5 K. Elgethun, R. A. Fenske, M. G. Yost and G. J. Palcisko, *Environ. Health Perspect.*, 2003, **111**, 115–122.
6 R. B. Langley, *GPS World*, 1995, **6**, 54–57.
7 E. D. Kaplan and C. J. Hegarty, *Understanding GPS: Principles and Applications*, Artech House, Norwood, MA, 2005.
8 T. H. Witte and A. M. Wilson, *J. Biomech.*, 2004, **37**, 1891–1898.
9 P. C. Weisbach and P. E. Glynn, *Clinical Prediction Rules: A Physical Therapy Reference Manual*, Jones & Bartlett Publishers, 2011.
10 W. Yang, K. Lee, C. Yoon, S. Yu, K. Park and W. Choi, *J. Exposure Sci. Environ. Epidemiol.*, 2010, **21**, 310–314.
11 K. Elgethun, M. G. Yost, C. T. E. Fitzpatrick, T. L. Nyerges and R. A. Fenske, *J. Exposure Sci. Environ. Epidemiol.*, 2006, **17**, 196–206.
12 S. Kaur, M. J. Nieuwenhuijsen and R. N. Colvile, *Atmos. Environ.*, 2007, **41**, 4781–4810.
13 L. Y. Chan, W. L. Lau, S. C. Zou, Z. X. Cao and S. C. Lai, *Atmos. Environ.*, 2002, **36**, 5831–5840.
14 E.-H. Chung and A. Shalaby, *Transport. Plann. Tech.*, 2005, **28**, 381–401.
15 W. Bohte and K. Maat, *Transport. Res. C Emerg. Tech.*, 2009, **17**, 285–297.
16 J. Du and L. Aultman-Hall, *Transport. Res. Pol. Pract.*, 2007, **41**, 220–232.
17 P. A. Gonzalez, J. S. Weinstein, S. J. Barbeau, M. A. Labrador, P. L. Winters, N. L. Georggi and R. Perez, *IET Intell. Transp. Syst.*, 2010, **4**, 37–49.
18 J. R. Landis and G. G. Koch, *Biometrics*, 1977, **33**, 159–174.
19 C. Rizos, *Manual of Geospatial Science and Technology*, 2002.
20 J. Wu, C. Jiang, Z. Liu, D. Houston, G. Jaimes and R. McConnell, *Environ. Health Insights*, 2010, **4**, 93–108.
21 C. Adams, P. Riggs and J. Volckens, *J. Environ. Monit.*, 2009, **11**, 1331–1339.
22 K. Lee, J. Y. Kim, K. Putti, D. H. Bennett, D. Cassady and I. Hertz-Picciotto, *J. Toxicol. Environ. Health Sci.*, 2009, **35**, 447–453.

