보건학석사 학위논문

New Analytical Method for Classification of Time–Location Data Obtained from the Global Positioning System (GPS)

GPS 로부터 얻은 자료의 시간–장소 정보에 대한 새로운 분석 방법의 개발

2013년 2월

서울대학교 보건대학원

환경보건학과 환경보건 전공

김 태 현

Abstract

New Analytical Method for Classification of Time–Location Data Obtained from the Global Positioning System (GPS)

Taehyun Kim

Department of Environmental Health

Graduate School of Public Health

Seoul National University

Personal exposure studies commonly use time–activity diaries and/or questionnaires to collect data on the study subjects' activities and locations during monitoring period. However, these methods are heavily affected by the recall abilities and participation of subjects. 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 minute 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 GPS were compared with the activity diary to determine selection criteria for 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 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; 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 GPS results were 1.6 ± 2.3 h for time spent in residential indoors, 0.9 ± 1.7 h for time spent in other indoors, 0.4 ± 0.4 h for time spent in transit, and 0.8 ± 0.5 h for time spent walking outdoors. This method can be used to determine time-activity patterns in exposure-science studies.

Keywords: Time-Location, GPS, microenvironmental classification, NMEA code

Student number: 2010-22039

Contents

List of Tables

List of Figures

Ⅰ. Introduction

Exposure assessment is a critical part of determining the health impact of environmental contaminants. Personal exposure is determined by contact over time and space between a person and one or more biological, chemical, or physical agents. Exposure science identifies and defines the exposures that occur, or are anticipated to occur, in human populations. Although personal exposure can be directly measured, it is often time consuming and expensive. Because exposure is directly proportional to the duration of exposure, the accuracy in recording the time and location influences the accuracy of the exposure calculation. Although the time resolution of direct-reading monitors allows identification of short-term or peak exposures, new methods for measuring time and location have not been as forthcoming.

There are several methods used for acquiring time–location information, including time–activity diaries, questionnaires, and observation (Freeman and de Tejada 2002). Environmental exposure sampling studies, such as the Total Exposure Assessment Methodology Study and the National Human Exposure Assessment Survey, have traditionally relied on time–activity diaries and surveys to provide information about study subjects' activities and locations during the exposure-monitoring period. 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. Additionally, it is highly impractical to ask subjects to record their exact locations for a period of time. Some researchers have suggested direct observation of subjects to record their time–location data. This can be done in studies with a small number of subjects over a limited period of time, but only at the expense of the subjects' privacy.

The shortcomings of the traditional time–location measures may be potentially addressed by the global positioning system (GPS) (Larsson, Burlin et al. 2002). GPS determines the position of the receiver by receiving radio signals from 24 satellites and three back-up satellites that revolve around the earth twice a day in predetermined orbits. These satellites are operated by the Navigational System with Timing and Ranging (NAVSTAR) and maintained by the US Department of Defense. The GPS technology provides the convenience of recording time–location information in the form of coordinates (latitude and longitude) with minimal human intervention, thus eliminating error from human sources. Because of the high reception rate of GPS outdoors, GPS is commonly used for outdoor activities (Schutz and Chambaz 1997). The reception rate can be significantly reduced by the Faraday cage effect (blocked by conductive materials such as earth and metal) (Stopher, FitzGerald et al. 2008). As a result, the indoor signal reception rate is generally nonexistent or low, and the accuracy is poor (Elgethun, Fenske et al. 2003).

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 GPGGA (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 (Langley 1995).

The GPS is used to identify location based on longitude, latitude, and height. A GPS receiver needs at least three satellite signals to calculate a location. However, the NSAT is generally greater when there is no Faraday cage effect. The speed of the GPS receiver can be determined by using the carrier phase-derived Doppler measurements or the receiver-generated Doppler measurements (Kaplan and Hegarty 2005). Therefore, the speed of the GPS receiver can be accurately determined (Witte and Wilson 2004) 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.

Ⅱ. Methods

2.1 Equipment and measurement methods

A field technician carried a GPS monitor (GPS 741, Ascen, Korea) while simulating scripted activities. Simulating the scripted activities and recording time diary were conducted by a professionally trained technician. The daily activity pattern was based on activity patterns of Seoul population. When the 2358 subjects were grouped by having similar activity patterns, there were classified into 10 different groups. These 10 groups consist of three 20~40-year-old office workers (I, II and III), teenager students, university students, senior citizen, house wife, 20~30-year-old the jobless, female service worker and male service worker (Korea NIER 2010).

In this study, activity patterns of four groups were simulated. The four groups were teenage students (3 days-persons), elderly persons (6 days-persons), housewives (3 days-persons), and male service workers (6 days-persons). Time–activity patterns of the four groups are shown in Figure 1. The actual time of each activity was also recorded using a pre-printed activity diary. The experiments were conducted in Seoul, Korea.

Group description	Activity pattern
Teenage student	
Senior citizen	
House wife	
Male service worker	

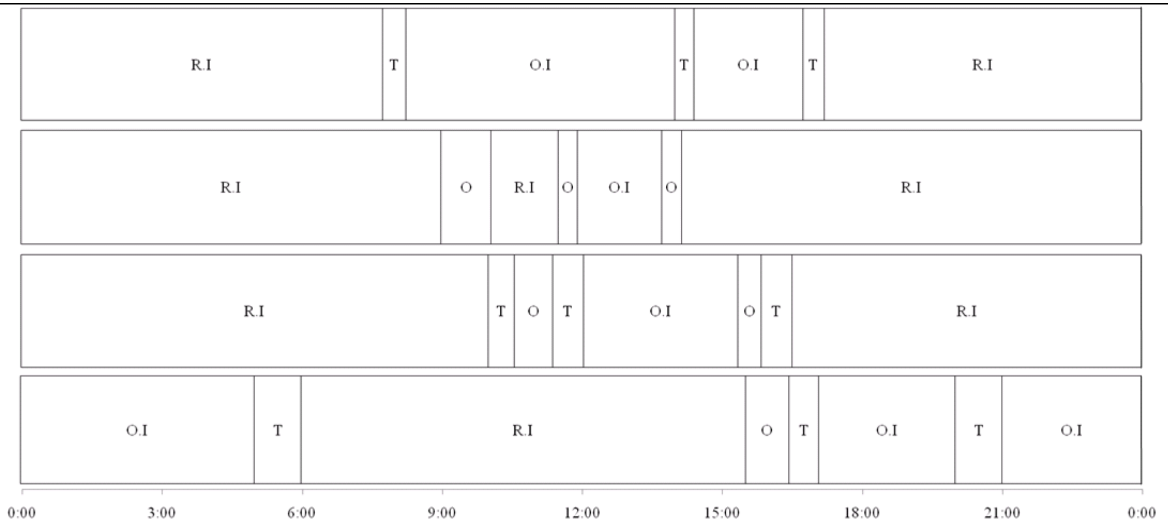

Figure 1. Simulated activity patterns by groups.

R.I: Residential indoor, O.I: Other indoors, T: Transit and O: Walking outdoors

In this study, a house was used as residential indoor for activity script. Figure 2 shows the external appearance of the residential building. The building, located in Seoul, Korea, was a three story brick building. The building floor space was about 100 m^3. The simulation was conducted on first floor. As shown in Figure 2, there were large windows in living room. The neighbor buildings were three to four story buildings.

Figure 2. External appearance of a building used as residence in in activity simulation.

Date of GPS measurement, activity groups and weather of each day are shown in Table 1. Each activity group has different number of measurement times. The GPS measurement was conducted as a part of 'Research for personal exposure assessment by time activity patterns on a nation' (Korea NIER, 2010).

Table 1. Date of GPS measurement, activity groups and weather condition

Date	Activity group	Weather condition
2010-05-07	A male service worker	Sunny
2010-05-08	A male service worker	A halo
2010-05-09	A male service worker	Smoke and fog
2010-06-23	An old person	Sunny
2010-06-24	An old person	Smoke and fog
2010-06-25	An old person	Smoke and fog / a halo
2010-07-17	An old person	Rain / shower / smoke and fog / lightning and thunder
2010-07-18	An old person	Rain
2010-07-19	An old person	Rain
2010-07-21	A teenage student	Rain
2010-07-22	A teenage student	Rain
2010-07-23	A teenage student	Rain / shower / smoke and fog / lightning and thunder
2010-07-24	A house hold	Rain / smoke and fog
2010-07-25	A house hold	Shower / smoke and fog
2010-07-26	A house hold	Rain / shower / smoke and fog
2010-07-27	A male service worker	Smoke and fog
2010-07-28	A male service worker	Sunny
2010-07-29	A male service worker	Rain

Time–location data were collected by the GPS monitor. The data storage capacity of this GPS device is approximately 125,000 points of geological 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 sampling time. The GPS monitor could be operated for approximately 25 hours before recharging. Table 2 shows specific description of the GPS device.

Table 2. Specific description of GPS device (GPS741, Ascen Korea)

Characteristics	Description
Size (cm^3)	4.7 x 7.2 x 2.0
Weight (g) with battery	64
Number of channel (#)	66
Data logging points capability (#)	125,000
Protocol	NMEA-0183 (V3.01)

"Logger tool" software from the Ascen website (http://www.ascen.co.kr) was used to configure the GPS data. Figure 3 shows interface of "Logger tool" software. With this program, download of NMEA codes was available. The GPS monitor was configured to collect the NMEA code before the field application. The collected GPS data were downloaded onto a computer using the "Logger tool" software and analyzed in Microsoft Excel 2007. The raw speed data were collected in units of km/h, and the NSAT data included both used-NSAT and viewed-NSAT. In Excel, the speed data were converted to units of m/s, and the used-NSAT data were extracted from NSAT (used/viewed).

Figure 3. 'Logger Tool' software for setting configuration and download NMEA codes of GPS.

2.2 Data analysis methods

The GPS data were classified into four microenvironments based on used-NSAT, distance from residence, and speed. The analysis flow is shown in Figure 4.

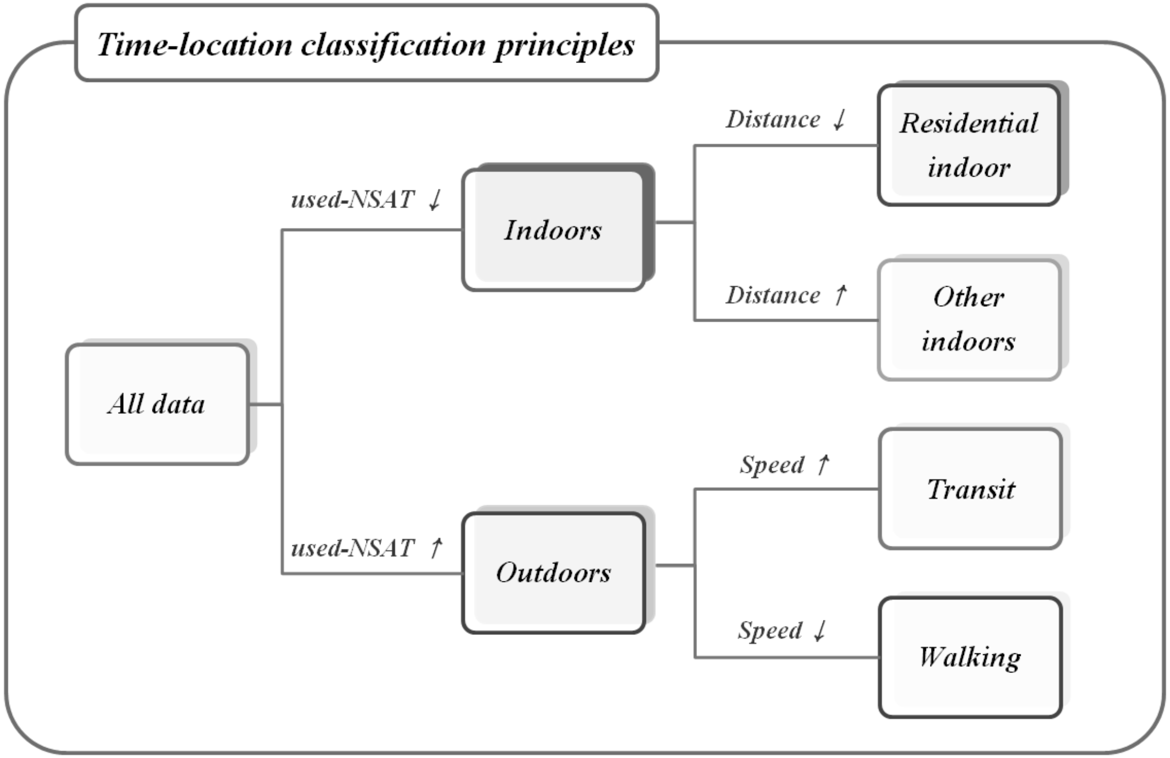

Figure 4. Flow chart of data-analysis process.

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. Because used-NSAT could be temporarily bounced when indoor, 3 min threshold was set discretionally. 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 exact center of the building was utilized as a residence indoor coordinate. 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 near 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 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.

2.3 Statistical analysis

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 (Paul E. Glynn 2011).

Ⅲ. Results

3.1 Time–activity simulations

A total of 18 daily time–activity simulations were conducted. The average daily number of GPS data points collected was 5086 ± 2913. The average amount of time spent in the residential indoors, other indoors, transit, and walking outdoors microenvironments were 14.9 ± 3.7, 6.7 ± 3.3, 0.8 ± 0.6, and 1.6 ± 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.

3.2 Indoor and outdoor classification by used-NSAT

Indoors and outdoors were determined by used-NSAT. Distributions of used-NSAT for indoor and outdoor status by time diary are shown in Figure 2. The average used-NSAT of outdoor status are significantly higher than that of indoor data (2.7244; $P<0.0001$, Student's t test). Section of number of used-NSAT in indoor and outdoor could not be completely divided. When respectively checking the numbers of used-NSAT which counts over 10%, used-NSAT of indoor data is concentrated from 4 to 8, they count for 84.9% of total indoor data. On the other hand, used-NSAT of outdoor data is concentrated from 7 to 10, they count for 88.8% of total outdoor data.

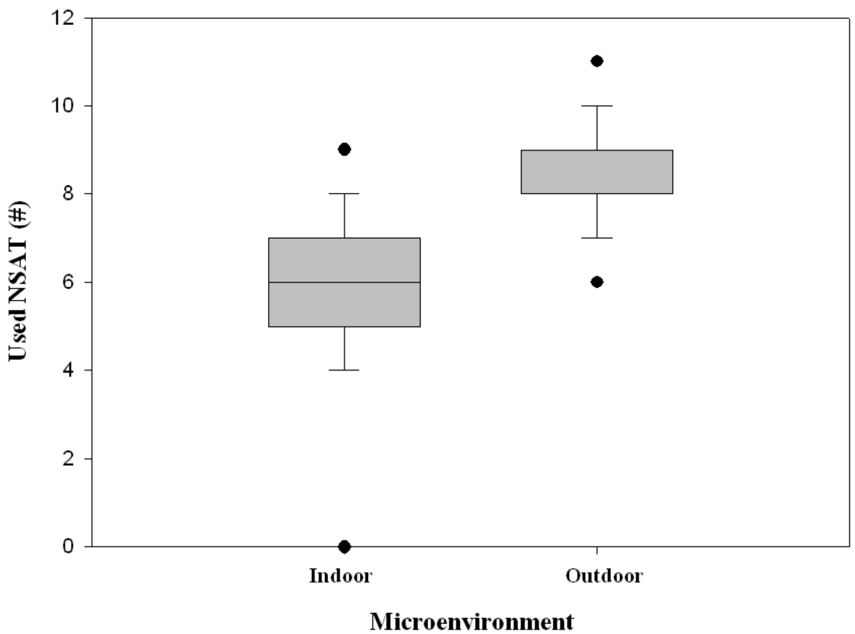

Figure 5. Used-NSAT distributions of indoor and outdoor status by time diary.

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 indoors ($P < 0.0001$). Figure 3 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.

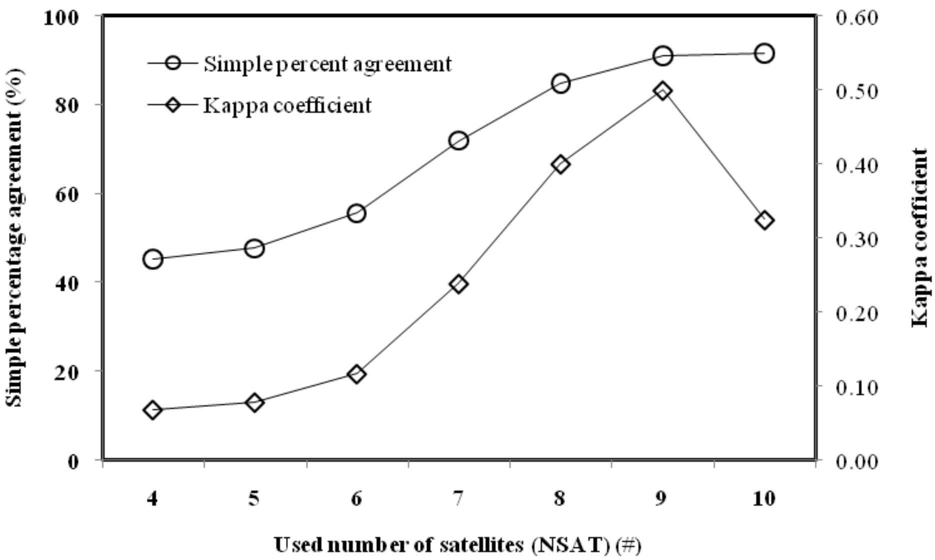

Figure 6. Accuracy of indoor and outdoor classification by used number of satellites (NSAT).

3.3 Residential indoor and other indoor classification by distance

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. Figure 7 shows distribution of GPS data with residential indoor. Geocodes in residential indoor were scattered one sided. It could be explained that geocodes were mainly scattered through the large windows in living room.

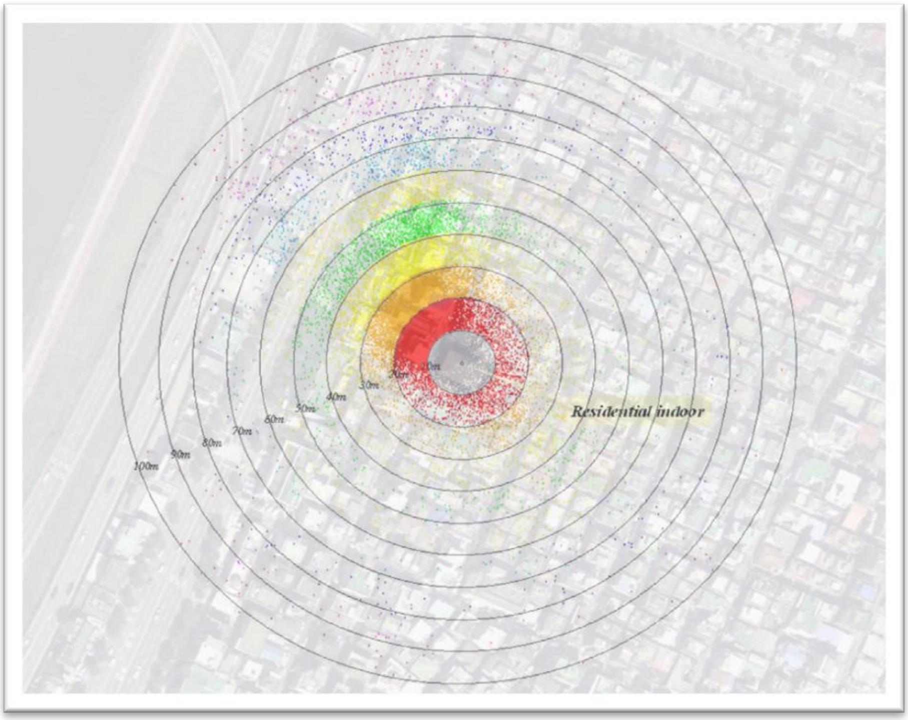

Figure 7. Distribution of GPS data reception in residential indoor at interval of 10 m.

Table 1 shows the number of data points at intervals of 10 m from center of 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%. Where distance is 40-50(m), residential indoor percentage was reduced from 95.2 % to 63.4 %. (This is not shown in Table) Therefore, a distance of 40 m from the residence was selected as the criterion to classify residential indoors.

Table 3. 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	9,064	9,281	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

3.4 Transit and walking classification by speed

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 and 0.7 ± 2.0 m/s, respectively. The average speed of transit was significantly greater than that of walking ($P < 0.0001$). Figure 4 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, the simple percentage agreement was 90.4 ± 7.1%, and the kappa coefficient was 0.79.

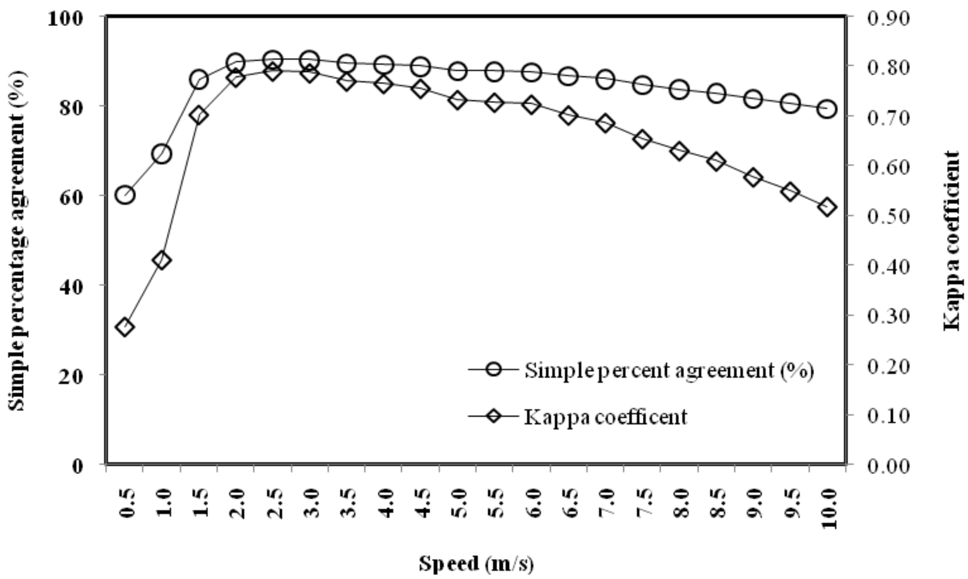

Figure 8. Accuracy of the classification as transit or walking based on the speed used as the criterion threshold.

3.5 Four microenvironments classification

The entire data set 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. The simple percentage agreement of the classified data with the activity diary was 84.3 ± 12.4%, and the kappa coefficient was 0.71. The accuracies of the classification into each microenvironment are shown in Figure 5. The simple percentage agreements for the data classified as residential indoors, other indoors, transit, and walking 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 diary and GPS results were 1.6 ± 2.3 h for time spent in residential indoors, 0.9 ± 1.7 h for time spent in other indoors, 0.4 ± 0.4 h for time spent in transit, and 0.8 ± 0.5 h for time spent walking outdoors.

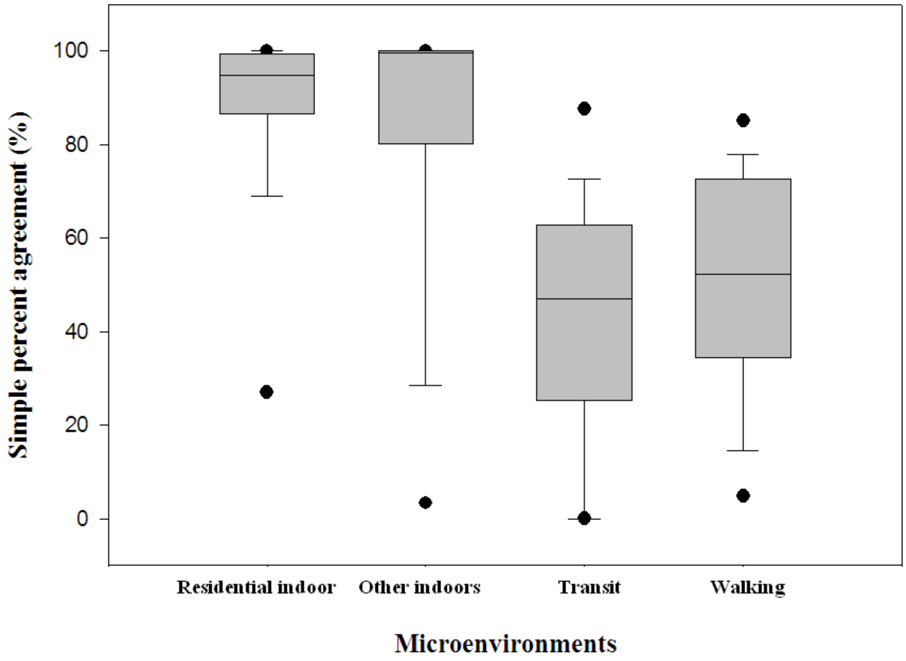

Figure 9. Distribution of the simple percentage agreement between the classification of data into the four microenvironments and the time–activity diary.

Ⅳ. Discussion

4.1 Scripted activity and patterns in Korean population

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. Because asking the general public to manually record all their time–activity information would be extremely burdensome, simulation by a technician was deemed to be the best approach to collect time–activity data with high accuracy. 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 four microenvironments (residential indoors, other indoors, transit, and walking outdoors) were similar to patterns in the Korean population (Yang, Lee et al. 2010).

4.2 Trends of used-NSAT at indoors and outdoors

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.

4.3 Residential indoor and other indoors classification by distance

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 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.

4.4 Transit and walking classification by speed

Transit time is often important for exposure assessment. In this study, a speed of 2.5 m/s was applied to further classify GPS data already classified as outdoors (on the basis of used-NSAT) as transit or walking. Because transit could be temporality stopped, we applied a 3-min rule; if a speed less than 2.5 m/s 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 as transit without any real justification (Elgethun, Yost et al. 2006). Air pollution levels in transit are generally high, and these can contribute substantially to daily personal exposure (Kaur, Nieuwenhuijsen et al. 2007). Exposure during transit can be affected by travel modes (Chan, Lau et al. 2002). 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 (Chung and Shalaby 2005; Du and Aultman-Hall 2007; Bohte and Maat 2009; Gonzalez, Weinstein et al. 2010).

4.5 Strength of this study

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 (Landis and Koch 1977). The accuracy of data that were classified as outdoors was relatively lower (55.4 ± 26.0%) than that of data classified as indoors (94.9 ± 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 (Stopher, FitzGerald et al. 2008).

Use of NMEA codes improved the classification of GPS data to 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 (Rizos 2002). The GPS device used in this study was 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.

4.6 Previous time-location studies

As previous time–location studies have experienced difficulties classifying GPS location data, other classification methods have been suggested, such as combination with other applications, including the use of temperature to distinguish between indoors and outdoors, use of a distance buffer region with GIS software, and use of Wi-Fi. However, these applications may be applicable only under specific conditions or may have other weaknesses. For example, using different temperatures to distinguish between indoors and outdoors may be possible only during a cold season. The method involving using a distance buffer region for all GPS points within a certain distance of a location of interest and all visited locations requires that information be known in advance.[21] The method using Wi-Fi had difficulty distinguishing between indoors and outdoors because Wi-Fi radiation could penetrate walls. Furthermore, the assignments were not compared with a diary, so the accuracy of the Wi-Fi method was not evaluated.

4.7 Limitations and future study

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 (Elgethun, Fenske et al. 2003; Adams, Riggs et al. 2009; Wu, Jiang et al. 2010). 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 (Lee, Kim et al. 2009).

This study developed a new analytical method for classification of time–location data obtained from GPS. To enhance accuracy of the analytical method, additional study should be conducted. GPS signal reception rate could be affected by many factors. Therefore GPS signal reception rate by diverse characteristics of residential building should be evaluated. Building materials such as concrete, wood and bricks, building area and height, the presence or absence of windows can also affect to GPS signal reception rate. A cloudy or rainy weather may interfere GPS signal reception rate. Regional characteristics such as characteristic of building and building densities may affect GPS signal reception rate. Therefore, further study may be needed to determine effect of various weather conditions, regional characteristics and various GPS devices.

Ⅴ. 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 in 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.

Ⅵ. 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

Adams, C., P. Riggs, et al. (2009). "Development of a method for personal, spatiotemporal exposure assessment." Journal of Environmental Montoring **11**(7): 1331-1339.

Bohte, W. and K. Maat (2009). "Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands." Transportation Research Part C: Emerging Technologies **17**(3): 285-297.

Chan, L. Y., W. L. Lau, et al. (2002). "Exposure level of carbon monoxide and respirable suspended particulate in public transportation modes while commuting in urban area of Guangzhou, China." Atmospheric Environment **36**(38): 5831-5840.

Chung, E.-H. and A. Shalaby (2005). "A Trip Reconstruction Tool for GPS-based Personal Travel Surveys." Transportation Planning and Technology **28**(5): 381-401.

Du, J. and L. Aultman-Hall (2007). "Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues." Transportation research part A-policy and practice **41**(3): 220-232.

Elgethun, K., R. A. Fenske, et al. (2003). "Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument." Environmental Health Perspectives **111**(1): 115-122.

Elgethun, K., M. G. Yost, et al. (2006). "Comparison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time-location patterns." Journal of exposure science & environmental epidemiology **17**(2): 196-206.

Freeman, N. C. G. and S. S. de Tejada (2002). "Methods for collecting time/activity pattern information related to exposure to combustion products." Chemosphere **49**(9): 979-992.

Gonzalez, P. A., J. S. Weinstein, et al. (2010). "Automating mode detection for travel behaviour analysis by using global positioning systemsenabled mobile phones and neural networks." IET Intelligent Transport Systems **4**(1): 37-49.

Kaplan, E. D. and C. J. Hegarty (2005). Understanding GPS: principles and applications. Norwood, MA, Artech House.

Kaur, S., M. J. Nieuwenhuijsen, et al. (2007). "Fine particulate matter and carbon monoxide exposure concentrations in urban street transport microenvironments." Atmospheric Environment **41**(23): 4781-4810.

Korea NIER (2010). "Research for personal exposure assessment by time activity patterns on a nation." Korea National Institute of Environmental Research.

Landis, J. R. and G. G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data." Biometrics **33**(1): 159-174.

Langley, R. B. (1995). "NMEA 0183: A GPS Receiver Interface Standard." GPS World **6**: 54-57.

Larsson, P., L. Burlin, et al. (2002). "Analysis of performance in orienteering with treadmill tests and physiological field tests using a differential global positioning system." Journal of Sports Sciences **20**(7): 529-535.

Lee, K., J. Y. Kim, et al. (2009). "Use of Portable Global Positioning System (GPS) Devices in Exposure Analysis for Time-location Measurement." Journal of Environmental Health Sciences **35**(6): 447-453.

Paul E. Glynn, P. C. W. (2011). Clinical Prediction Rules: A Physical Therapy Reference Manual, Jones & Bartlett Publishers.

Rizos, C. (2002). "Introducing the global positioning system." Manual of geospatial science and technology.

Schutz, Y. and A. Chambaz (1997). "Could a satellite-based navigation system (GPS) be used to assess the physical activity of individuals on earth?" European Journal of Clinical Nutrition **51**(5): 338-339.

Stopher, P., C. FitzGerald, et al. (2008). "Search for a global positioning system device to measure person travel." Transportation Research Part C-Emerging Technologies **16**(3): 350-369.

Witte, T. H. and A. M. Wilson (2004). "Accuracy of non-differential GPS for the determination of speed over ground." Journal of Biomechanics **37**(12): 1891-1898.

Wu, J., C. Jiang, et al. (2010). "Performances of different global positioning system devices for time-location tracking in air pollution epidemiological studies." Environmental health insights **4**: 93-108.

Yang, W., K. Lee, et al. (2010). "Determinants of residential indoor and transportation activity times in Korea." Journal of Exposure Science & Environmental Epidemiology **21**: 310-314.

국문초록

New Analytical Method for Classification of Time–Location Data Obtained from the Global Positioning System (GPS)

GPS 로부터 얻은 자료의 시간−장소 정보에 대한 새로운 분석 방법의 개발

김 태 현

환경보건학과

서울대학교 보건대학원

개인노출 연구에서는 연구 참여자의 활동 및 장소를 파악하기 위해서 주로 일지와 설문을 이용한다. 하지만 이러한 방법은 연구 참여자의 기억능력과 참여의지에 크게 영향을 받을 수 있다. 대안으로서 위성 위치 확인 시스템(GPS)을 이용하는 방법이 제시 되어 왔지만 이용하기에는 아직 제한이 있다.

이 연구의 목적은 GPS 로부터 얻은 자료를 시간–장소를 분류하는 새로운 분석방법을 평가하는 것이다. 한 명의 실험자가 GPS 장치를 소지한 체 다양한 활동일정을 시행하였고, 이를 분 단위로 정확히 활동일지에 기록하였다. GPS 장치는 위치 정보를 매 15 초 마다 자동으로 기록하도록 설정되었다. 총 18 일 동안 하루 단위의 활동일정을 시행하였다.

GPS 로부터 얻은 자료로 분류한 시간–장소 정보는 분류 기준을 정하기 설정하기 위하여 활동일지와 비교하였다. GPS 데이터는 4 개의 국소환경(주택실내, 기타실내, 교통수단, 실외도보)로 구분되었다. 분류 기준으로서 기록에 이용된 위성의 수(used–NSAT), 속도, 집에서의 거리가 이용되었다. Used–NSAT 가 9 미만일 경우 GPS data 는 실내로 구분하였다. 후에 집과의 거리가 40m 이내일 경우 주택실내로 분류하였고, 그렇지 않을 경우 기타실내로 분류되었다. 실외로 분류된 데이터는 속도가 2.5 m/s 이상일 경우 교통수단으로 분류되었고, 그렇지 않을 경우 실외도보로 분류되었다. 이렇게 분류된 시간–장소 정보와

활동일지와의 평균 단순 일치율은 84.3 ± 12.4%, kappa coefficient 는 0.71 로 나타났다. 시간-장소 정보와 활동일지와의 평균 시간 차이는 주택실내 1.6 ± 2.3 시간, 기타실내 0.9 ± 1.7 시간, 교통수단 0.4 ± 0.4 시간, 그리고 실외도보에서 0.8 ± 0.5 시간으로 나타났다. 이 방법은 노출평가 연구에서 시간-활동 정보(time-activity patterns)을 판단하기 위하여 이용될 수 있을 것이다.

주요어: 시간-장소, GPS, 국소환경 구분, NMEA code

학번: 2010-22039