Tsukuba Geoenvironmental Sciences, Vol. 10, pp. 33-41, Dec. 26, 2014

Measuring mobility and urban structure: A case study of four capital cities in Asia

Konstantin GREGER[*] and Yuji MURAYAMA[**]

Abstract

We analyzed four movement data sets provided by the Center for Spatial Information Science (CSIS) at the University of Tokyo which contain information about the movements of sample populations in Dhaka, Bangladesh; Hanoi, Vietnam; Jakarta, Indonesia; and Manila, Philippines. We showed a number of similarities and dissimilarities between trip activities, transportation modes, and mobility ranges of the four cities' sample individuals. A more detailed look at the commuting processes involved revealed differences between the four study areas regarding the number of subtrips per commuting trip as well as differences between morning and evening commutes and differences between trips for men and women within the cities. It also became obvious that commuting trips take place at different times during the day in each of the four study areas, hinting that a two-shift system is in place both at schools and at workplaces in some of the cities. The mobility range, an important component to describe the mobility behavior of the population of a study area, also showed differences between the cities. This can be attributed to the spatio-temporal distribution characteristics we were able to extract by looking at the spatial distribution of the stationary population at night and during the day.

Key words: Asia, behavior, mobility, person trip, urban spatial structure

1. Introduction

Mobility behavior and its interaction with urban structures have not been analyzed in a concise and quantitative way for South and Southeast Asian cities in the past. This study comprises the first intermediary results of an ongoing research project, "Mobility and Urban Structure" at the Division of Spatial Information Science (SIS), University of Tsukuba, Japan, in which we analyze data from four Asian capital cities: Dhaka, Bangladesh; Hanoi, Vietnam; Jakarta, Indonesia; and Manila, Philippines. The purpose of this study is to analyze urban spatial structures of these four cities in terms of population movements over 24 hours and clarify the causes of the differences in the spatial flow patterns.

Not only is this the first study to perform detailed quantitative analyses of the mobility data available for these four cities, but it is also unique in its comparative approach to the four study areas. In Section 2, we give an overview of the overall structure of the data sets used and the technical challenge of working with four massive datasets of detailed movement data for four study areas. We then provide insights into the mobility behavior of the four cities' populations in Section 3. This comprises both spatial and temporal patterns and also the questions of why, i.e. trip purpose, and how, i.e. mode of transportation and people movement, where we put a specific focus on commuting behavior. Lastly, we also analyze the correlation between the mobility data and the four study areas' spatial urban structures, and how the latter can be inferred from the former. Section 4 provides a summary of our findings.

2. Data structure and technical challenge

This study uses movement data sets provided by the Center for Spatial Information Science (CSIS) at the University of Tokyo, which contain information about the movements of sample populations. These include the origins and destinations of all trips, both regarding time and location, as well as the purpose of each trip and the modes of transportation used per subtrip. A trip is here defined as a movement from one location to another with a certain purpose. Each trip consists of one or more subtrips, which are characterized by a change in the mode of transportation. For the sake of these data sets, times of stationarity, i.e. when no locational change occurs, such as while a person is at home or at the workplace, are also captured as trips, but are labeled as stationarity events by a bespoke mode of transportation. In addition, the data sets also contain several socio-demographic attributes about the sample individuals such as sex, age, and occupation.

Each data set in itself contains an immense depth of information about individual movements and resulting movement patterns and their connection with the urban spatial structure of the respective city. The comparison of all four data sets allows us to reveal interesting insights into the commonalities and differences among these four Asian capital cities. It should be mentioned at this point that the four datasets were collected at different points in

[*] Graduate School of Life and Environmental Sciences, University of Tsukuba
[**] Faculty of Life and Environmental Sciences, University of Tsukuba

time: Dhaka in 2009, Hanoi in 2004, Jakarta in 2002, and Manila in 1996 (CSIS, 1996, 2002, 2004, 2009). While this represents a range of 13 years, we are confident that the status quo in terms of urban spatial structure, society, and the resulting mobility requirements of these four cities has not changed too dramatically, and that the data can, hence, be compared as planned.

One particular challenge is the fact that the four data sets had been originally collected as separate research projects and, hence, differ quite significantly in their structures. While the team at CSIS has put great effort into consolidating the data sets into a common data structure, the variable encodings still show significant differences. Especially attributes of trip purpose, mode of transportation, and occupation varied significantly over the four data sets. We decided to create more generalized groups of all three attributes to normalize the data and allow for comparisons among the four study areas. We also relabeled the trip purposes as "activities".

Another challenge is the amount of data that needs to be processed. Each data set contains not only the aforementioned items regarding of how and why people moved and their socio-demographic attributes, but also point locations in 1-minute intervals over the course of 24 hours of data collection. This results in at least 1,440 time-and geocoded data points per person and data for trips that were started on the day of data collection but ended after midnight. Given the large sample populations, this results in a massive number of data items that need to be analyzed as Table 1 shows.

We were able to overcome these difficulties by using a combination of on-disk storage of the data in a PostgreSQL/PostGIS database (PostgreSQL Global Development Group, 2014) and in-memory analysis and visualization using R (R Core Team, 2014). Preliminary spatial analyses and cartographic visualization were performed in QGIS (QGIS Development Team, 2014) since it provides the necessary interfaces for both R and PostgreSQL/PostGIS and is highly customizable using Python scripts (Python Software Foundation, 2014). All analysis scripts can be found in the research project's GitHub repository (http://github.com/kogreger/capital-cities).

3. Mobility behavior

3.1. Activity

In all four cities, the activity of "going home" accounted for the majority of trips. This stems from the fact that return trips of all other activities (e.g. work, education, shopping, etc.) are included in this category. There are quite significant differences in the remaining activities. This can partly be explained by the disproportionate distribution of occupation groups prevalent in the data sets.

Of the 105,121 non-stationary trips in Dhaka (Fig. 1a), 23% were to the workplace and 13% to an educational institution (e.g. school or university). At only 2%, running errands did not play a major role. Similarly, in Hanoi (Fig. 1b), 23% and 11% of the 122,530 non-stationary trips were for work and education, respectively. Errands accounted for 9% of trips. 20% and 14% of the 746,149 non-stationary trips in Jakarta (Fig. 1c) were for workplaces and educational institutions respectively, and 10% were for errands. No trips for social activities and recreation were registered, but, for more than 20,000 trips, the purpose was unknown. The data for Manila (Fig. 1d) shows a distribution very similar to that in Dhaka. Of the 324,312 non-stationary trips, 21% were to the workplace, 18% to educational institutions, and only 2% for errands. Of particular note here is the higher number of trips to schools and universities. This can be attributed to the very high number of students represented in the sample population.

3.2. Mode of transportation

One major component of mobility behavior is the choice of a mode of transportation. As a result of the aforementioned differences in the data collection processes (cf. Section 2), the raw data of the four study areas could not be compared. After grouping them into 12 more general modes of transportation we were able to use the four data sets for comparative analyses. These sets include one pseudo-mode of transportation to represent stationarity and one mode for unknown data. In none of the four study areas were any subtrips registered using an airplane, therefore, we excluded this mode from all analyses. The distribution of subtrips over the remaining nine modes of transportation did reveal some interesting facts.

We grouped all scheduled public road-bound transportation into the transportation mode "bus". This includes public buses, school, company, and tourist buses, as well as *patas* in Jakarta and *jeepneys* in Manila. All non-private unscheduled public modes of transportation were combined into the transportation mode "taxi". This includes regular taxis as well as *CNGs*, *minshuks*, and *rickshaws* in

Table 1. Data sets for the four study areas

	Dhaka	Hanoi	Jakarta	Manila
Data collection	1.-2. Oct. 2009	1.-2. Oct. 2004	1.-2. Oct. 2002	1.-2. Oct. 1996
Sample population	42,114	58,018	297,043	189,335
Number of data points	60,917,515	83,790,715	429,546,157	274,208,563

Dhaka, *cyclos* and *xe oms* in Hanoi, *bajaks*, *ojeks*, *becaks*, and *omprengans* in Jakarta, and *pedicabs* and *HOV taxis* in Manila.

In Dhaka, Jakarta, and Manila, walking made up the greatest part of the non-stationary subtrips at 39%, 45%, and 41%, respectively. The second- and third-most used modes of transportation in Dhaka (Fig. 2a) were taxis (36%) and buses (19%). Bicycles, motorcycles, and cars were not used widely at <1%, 1%, and 5% respectively. In Hanoi (Fig. 2b), almost half of the subtrips (45%) were covered by motorcycle, followed by walking (28%) and bicycle (21%). None of the other transportation modes played a bigger role there, buses accounted only for 3%. Jakarta (Fig. 2c) and Manila (Fig. 2d) showed very similar distributions of transportation modes with the majority of non-stationary subtrips being walking (45% and 41% respectively), followed by buses (31% and 34% respectively) and motorcycles (11% and 13% respectively). The bi-

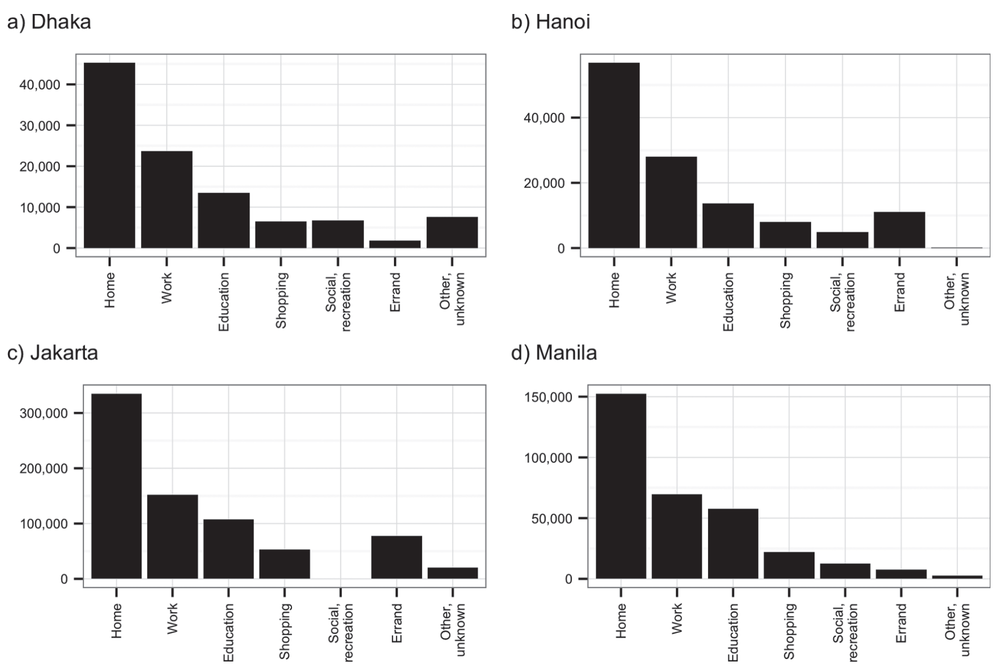

Fig. 1 Number of trips per activity

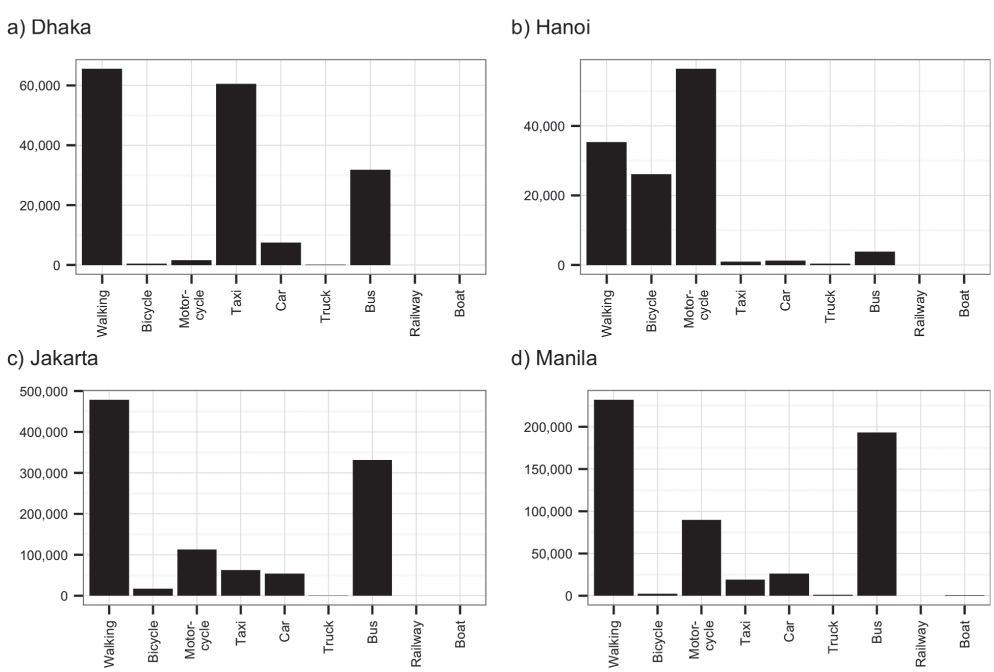

Fig. 2 Number of subtrips per mode of transportation

cycle does not play a major role in either city; similarly, taxis and private cars account for only 3–6% in both study areas.

3.3. Mobility range

We understand mobility range as another important component to describe the mobility behavior of the population of a certain study area. We operationalized a person's mobility range as his or her total distance traveled over the course of the sampling period (i.e. 24 hours plus the end of all trips started before midnight). The average total distance traveled was 12.2 km, 16.5 km, 38.3 km, and 13.4 km in Dhaka, Hanoi, Jakarta, and Manila respectively.

A more detailed analysis of the average total distance traveled showed that those under 20 years old and those over 80 years old in Dhaka covered rather short overall distances, on average under 10 km. The total traveled distance of the medium age groups was between 12.9 km and 15.2 km with the highest values at the age groups of 50 to under 55 years old and 60 to under 65 years old, both with an average of over 15 km (Fig. 3a). In terms of their occupation, students covered short trips, 8.4 km on average, which probably only lead them to a local school and back home. In contrast, employees need to cover longer distances over the course of the day. The average total distance for blue-collar workers was 20.5 km, almost 4 km longer than that for white-collar workers at 16.6 km (Fig. 3b). This could be explained by the living quarters of the former being located farther from their workplaces.

The data from Hanoi shows that, regarding the ages of the sample individuals, only people older than 65 years traveled on average less than 10 km over the course of the day. In contrast, people in the age groups between 15 and under 40 years old covered an average distance of over 20 km (Fig. 3c). In contrast to the data from Dhaka, students in Hanoi seemed to have to travel further to their educational institutions; their average total travel distance amounted to 18.8 km. Only housewives/househusbands and unemployed and retired people traveled less than 10 km in total on average. Unlike in the Dhaka data set, blue-collar workers in Hanoi traveled on average 3.8 km less than white-collar workers, 19.0 km and 22.8 km respectively (Fig. 3d).

The analysis of the data for Jakarta resulted in figures that left us questioning the overall quality and cleanliness of the data set. Across the two aforementioned dimensions of age and occupation group, we did not find any meaningful distributions. The age groups' average distances ranged from 36.0 km to 43.1 km, with the age groups from 60 years old to 75 years old traveling the farthest distances, over 40 km on average (Fig. 3e). Due to the largely missing occupation data for the sample population of the Jakarta data set, the resulting statistics do not reveal any relevant trends. The average total travel distance of 39.7 km for white-collar workers was slightly farther than that of the 39.1 km for blue-collar workers and the 37.5 km for the majority of the population with unknown occupations (Fig. 3f). It is worth noting, though, that all these distances are farther than the distances in any of the other study areas and more than twice as far overall.

The data from Manila shows a distribution similar to that from Hanoi in regard to the average total distance traveled per age group. The age groups under 15 years old and over 65 years old traveled on average less than 10 km; the overall farthest distances were covered by people in the age group from 20 to under 25 years old, at 17.4 km on average (Fig. 3g). The average total distance covered by white-collar workers at 20.3 km was 1.5 km farther than that of the 18.8 km for blue-collar workers. This was the case in Hanoi and Jakarta but was contrary to the observations from Dhaka (Fig. 3h).

3.4. Commuting behavior

As the analysis of activities in Section 3.1 has shown, most of the non-homeward trips lead to workplaces or educational institutions, 62.0% in Dhaka, 63.3% in Hanoi, 63.2% in Jakarta, and 74.2% in Manila. We accommodated for this fact by taking a closer look at commuting behavior in the four study areas. We did this separately for the morning and evening commuting periods and also for men and women (Table 2). As commuting times differ per occupation type and cultural surrounding, we decided to formulate the two time periods very generally as "before noon" and "after noon", but, for reasons of simplicity, we will refer to them as "morning" and "evening" commuters in the following.

We determined the commuting data by looking only at trips that were marked either with the purpose of going to work or going to an educational institution and that led home from either of these two locations (i.e. workplace or school). It should be mentioned here that this excludes all commuting trips with intermediary stops, such as shopping on the way home, as these appear as two separate trips in the movement data sets: one from the workplace or school to the market (activity: shopping) and one from the market to the home location (activity: home).

The data for Dhaka shows that the evening commute takes, on average, 6–7 minutes longer than the morning commute. In Hanoi and Jakarta, the differences amounts to only 2–4 minutes. Furthermore, Hanoi had not only the overall shortest commuting time of 19–21 minutes per direction but, at 1.0, also the lowest average number of subtrips per commuting trip. This can be explained by the aforementioned prevalence of motorcycles there, which allows individuals to travel from origin to destination without changing the mode of transportation (cf. Section 3.2). In the remaining cities, the commuting times ranged between 32 and 67 minutes and the average number of subtrips ranged between 1.3 and 2.0. Also, women in Dhaka appeared to change their mode of transportation slightly less often than did men, although the reasons for this behavior are unclear. The same held true for the fact that women in both Dhaka and Jakarta appeared to change their mode of transportation slightly more often during the eve-

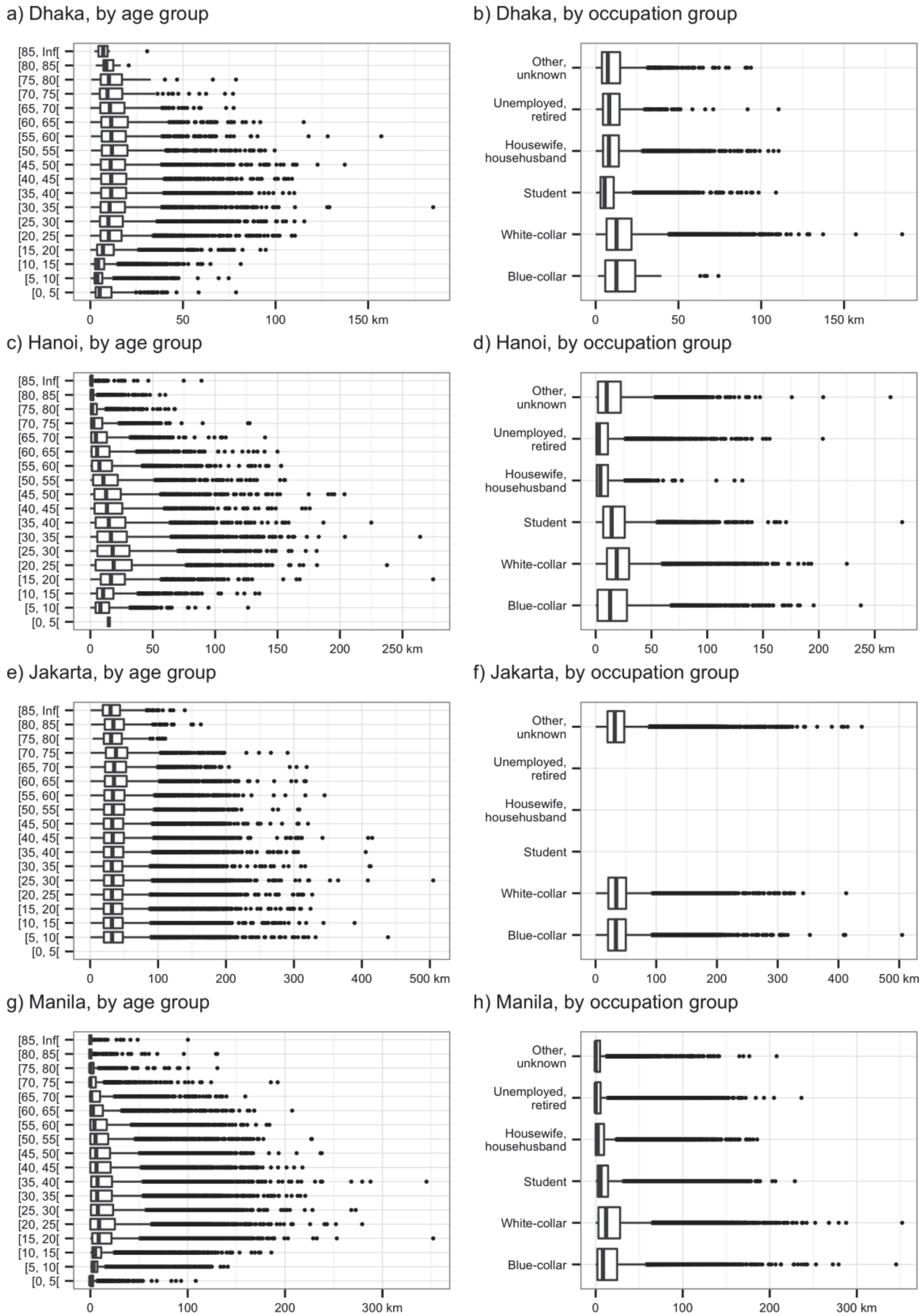

Fig. 3 Frequency distribution of average total distance traveled by age group and occupation group

ning commute as compared to the morning commute. Manila shows four patterns that differ from the other three study areas. First is the aforementioned significantly longer commuting times, which lasted between 40 and 67 minutes – the shortest trips being on average as long as the longest trips in Dhaka and Jakarta and twice as long as those in Hanoi. Second is that the commuting trips of men took between 13 and 20 minutes longer than those of women. And third is that the evening commute for women was 6 minutes longer than the morning commute, while the overall average as well as the data for men showed that the morning commute took 22 and 27 minutes longer respectively. Lastly, women took, on average, 0.4 more subtrips during their commuting trips, which means that they tended to change the mode of transportation more often.

We were interested to see how the temporal commuting patterns differ between the four study areas. To find out more about this, we established the number of commuting trips (i.e. to work, to an educational institution, or to home from either of these two) that were started within each hour of the day. The results are shown in Fig. 4 and reveal a number of interesting facts.

The data for Dhaka (Fig. 4a) show a rather regular commuting pattern, similar to that identified by Greger (2014) in a similar study analyzing commuting behavior for Tokyo, Japan. That is, one distinct peak of people commuting to work or school in the morning, where students appear to begin travel one hour earlier than workers, and a converse trend of homeward trips over the course of the afternoon, with peaks at 12:00, 14:00, and 17:00, which can be under-

Table 2. Morning and evening commuting data

Dhaka		Mean duration	Mean subtrips
All	Morning	34 min.	1.6
	Evening	40 min.	1.6
Men	Morning	36 min.	1.6
	Evening	42 min.	1.6
Women	Morning	30 min.	1.4
	Evening	37 min.	1.5

Hanoi		Mean duration	Mean subtrips
All	Morning	19 min.	1.0
	Evening	20 min.	1.0
Men	Morning	20 min.	1.0
	Evening	21 min.	1.0
Women	Morning	19 min.	1.0
	Evening	20 min.	1.0

Jakarta		Mean duration	Mean subtrips
All	Morning	34 min.	1.5
	Evening	37 min.	1.6
Men	Morning	35 min.	1.5
	Evening	38 min.	1.5
Women	Morning	32 min.	1.5
	Evening	36 min.	1.6

Manila		Mean duration	Mean subtrips
All	Morning	62 min.	1.7
	Evening	42 min.	1.4
Men	Morning	67 min.	1.6
	Evening	40 min.	1.3
Women	Morning	47 min.	2.0
	Evening	53 min.	1.7

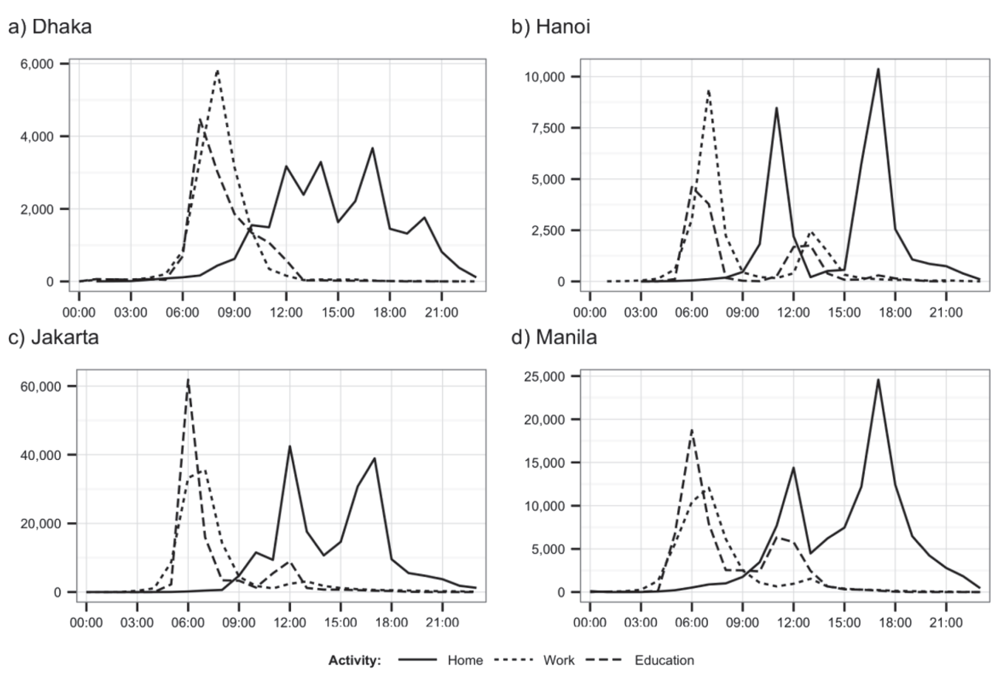

Fig. 4 Number of commuting trips started per activity and hour

stood to correspond to ending times for schools, part-time jobs, and full-time jobs respectively.

In contrast, the data for Hanoi (Fig. 4b) show two distinct peaks for commuting trips to work at 7:00 and 13:00 and another two peaks for trips to educational institutions at 6:00 and 12:00–13:00. Two very distinct peaks of homeward trips also emerge at 11:00 and 17:00. These observations make us believe that a two-shift system is in place there, both at schools and at workplaces, which causes people to return home around noon after working or studying since the morning and to return to work or school afterwards, only to go back home in the late afternoon. In further studies, it would be worthwhile to analyze in more detail if the same persons are commuting twice per day or if this is a case of a changing shift where some people work or study before noon and others after noon.

The data from Jakarta (Fig. 4c) show a situation similar to that from Hanoi (Fig. 4b). While the morning peak for students appears sharply at 6:00, the morning peak for people going to work is more stretched out, from 6:00–7:00, which can be seen as an indicator of more flexible working arrangements. A large number of people went back home from their workplace or school at 12:00, at which time another commuting peak can be noticed, although it is a lot less pronounced and is almost exclusively for students. After this, a second distinct homeward peak is obvious at 16:00–17:00, again slightly more stretched out than in the other cities.

The data from Manila (Fig. 4d) are almost identical to that from Jakarta regarding the temporal distribution of commuting trips. The data show one distinct peak for trips to educational institutions at 6:00, a slightly more spread out peak of trips to workplaces from 6:00–7:00, and secondary, less pronounced commuting peaks around noon, 11:00–12:00 for students, and, to a lesser degree, at 13:00 for workers. The data also reveal two distinct peaks of homeward trips at 12:00 and 17:00, the later significantly more pronounced.

3.5. Spatio-temporal patterns

Lastly, we were also interested in the spatio-temporal patterns defined by the movements of people in the four study areas. We believe that these are caused partly by urban spatial structures. In order to study this, we decided to look at the distances from the city center for each of the four study areas and created buffer rings at distances between 5 km and 50 km. The maps in Fig. 5 show these buffer rings for all four cities together with the spatial distribution of the 1-minute point positions of all sample individuals over the complete sampling period. Some of these point positions are located outside of the 50 km buffer ring (to the north in Dhaka and to the south in Jakarta and Manila), but we considered these outliers to be negligible since this study focuses mainly on the highly urbanized central areas of the four capital cities. The maps also reveal significant water bodies in the areas covered by these analyses for the cases of Jakarta and Manila. This plays a role in the comparison of the results for the four cities.

In Dhaka (Fig. 6a), the data show a rise in the number of people being stationary at the 5 km buffer, which represents the downtown area, and a decline in the stationary population between the 5 and 15 km buffers can be observed. This can be understood as a movement of people from these latter areas towards the city center during the day, mostly between 6:00 and 21:00. The stationary populations in the outer areas, more than 15 km from the city center, stay mostly the same over the course of the day.

The data from Hanoi (Fig. 6b) show a similar pattern, albeit much less pronounced. The stationary population in the 5 km buffer ring increases between 7:00 and 20:00 and decreases in the 15 km buffer ring. In addition, a strong peak can be observed in the three innermost buffer rings (i.e. up to 15 km) at 7:00, which coincides with the main commuting period for the working population. This can be explained by the common practice in Hanoi, as in other Vietnamese cities, to eat breakfast in small soup kitchens that offer food on the roadside. The stationary populations in the outer buffer rings, more than 15 km from the city center, show no significant fluctuations.

The data from Jakarta (Fig. 6c) once again appears to be erroneous. The spatio-temporal distribution of the stationary population is characterized by a very large increase of the population present in the 15 km buffer ring from 8:00 until the end of the sampling period, which, to the best of our knowledge, is not explained by any real-world phenomena. Since the data for all other buffer rings except the 10 km buffer ring decreases at the same time, it seems as if almost the entire population from all other areas is moving to the city center, leaving the outer parts of the city completely deserted. The total stationary population inside the 15 km buffer ring increases from just under 70,000 people at 8:00 to almost 225,000 at 23:00, more than three times as many people. This peculiarity aside, the data also show a partly similar pattern to that of the aforementioned data sets, namely an increase in the stationary population figures in the city center, here in the two innermost buffer rings up to 10 km from the city center. We decided to exclude the Jakarta data set from the analyses in this case, since deficiencies in the data quality are very obvious, which would most likely render any results questionable.

The data from Manila (Fig. 6d) show the pattern already established in the other three study areas, an increase in the stationary populations in the city center with a concurrent decrease in the stationary populations in the buffer rings

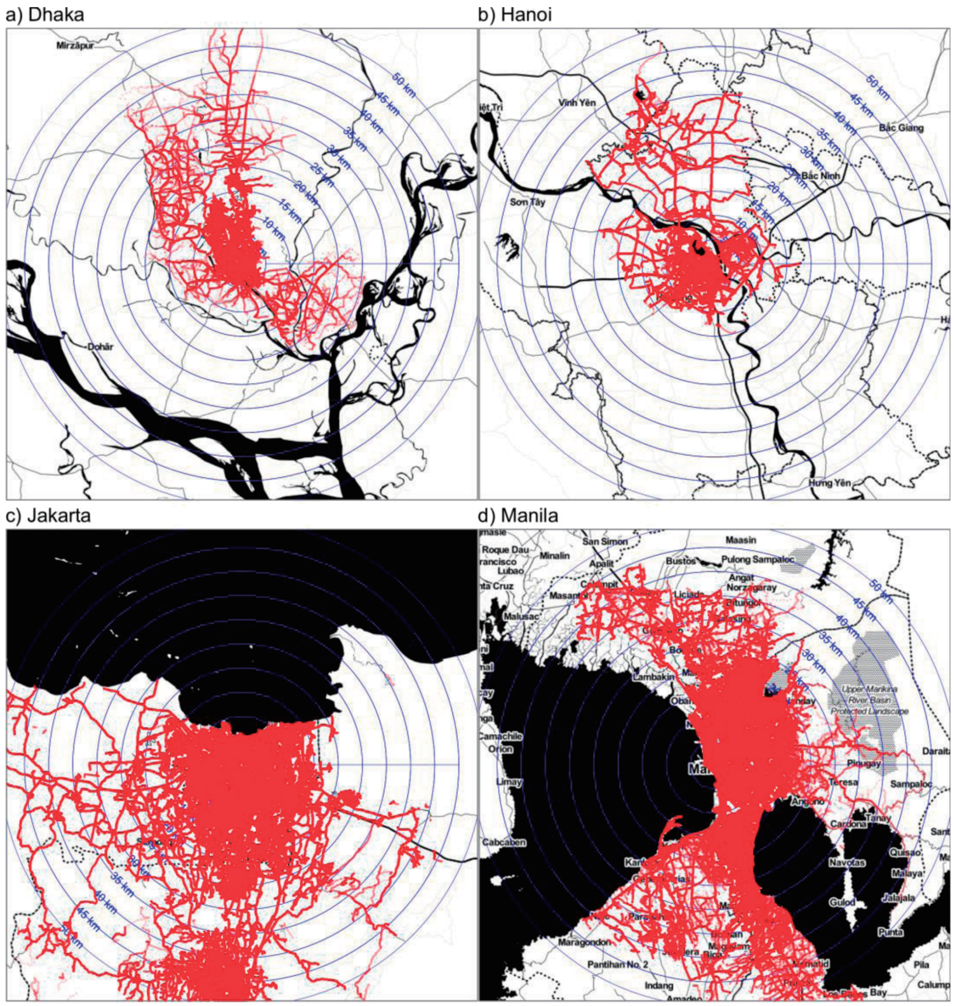

Fig. 5 Spatial distribution of the 1-minute point positions of all sample individuals over the sampling period

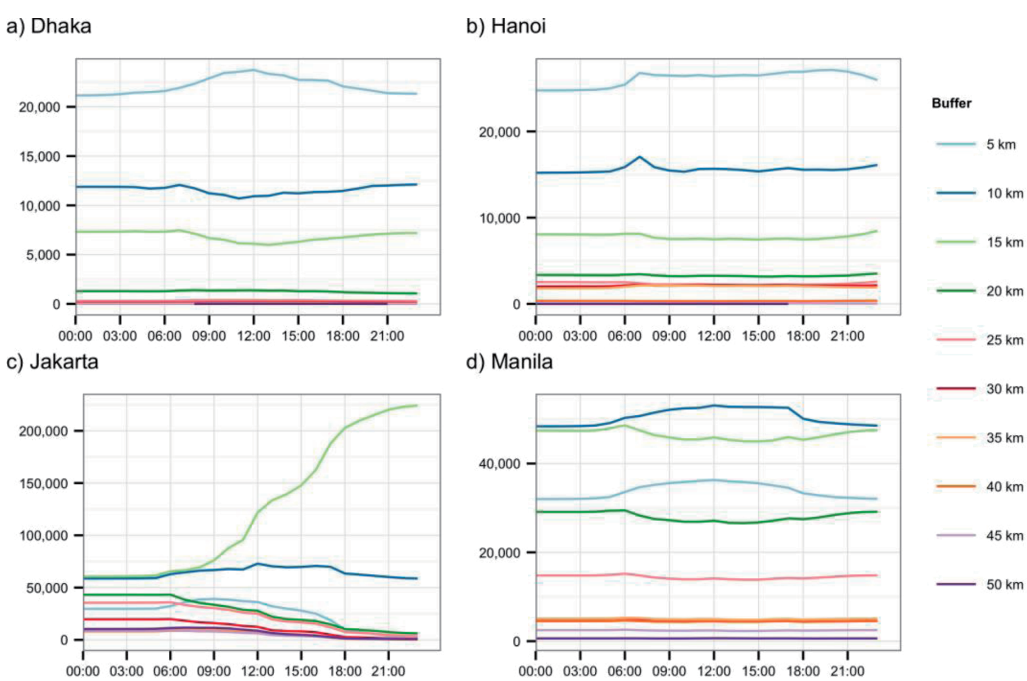

Fig. 6 Number of stationary people per buffer ring and hour

further away from the city center. In the case of this data, an increase can be observed in the two innermost buffer rings, up to 10 km from the city center, and a corresponding decrease can be observed in the three adjacent buffer rings between 15 km and up to 25 km. It is also worth noting that, in contrast to the data from Dhaka and Hanoi, the most populated buffer ring was not the innermost 5 km zone but instead the adjacent 10 km buffer ring. Here, even the 15 km buffer ring had a greater stationary population than the city center. This can be attributed to the characteristic shape of Manila on the isthmus between Manila Bay in the west and Lake Laguna in the east.

4. Concluding remarks

Overall, we believe that the preliminary findings we presented in this study show not only the potential of the four data sets used, but also a general confirmation of our assumption that the analysis of people's movement can allow for conclusions regarding the urban spatial structure of the respective study areas. Over the course of the research project "Mobility and Urban Structure" at the Division of Spatial Information Science (SIS), University of Tsukuba, Japan, we will continue to analyze the four data sets introduced in this study and also make comparisons with similar data from Tokyo. We hope that this will allow us to draw conclusions regarding the correlation between mobility behavior in capital cities of developing countries and mobility behavior in the capital city of a highly industrialized country.

Acknowledgements

The Center for Spatial Information Science at the University of Tokyo provided us with data sets within the scope of joint research project No. 536. We would like to extend our gratitude to Prof. Ryosuke Shibasaki, Prof. Yoshihide Sekimoto, and Dr. Hiroshi Kanasugi at University of Tokyo for their assistance in the progress of this joint research project. This study was supported by a Grant-in-Aid for Scientific Research A (No.24241053, 2012-15), "Reconstruction of urban geography toward the formation of sustainable urban space" (Representative: Masateru Hino).

References

CSIS (1996): *People flow 1996 Manila Metropolitan Area* [Data File]. https://joras.csis.u-tokyo.ac.jp/dataset/show/id/3019199600.

CSIS (2002): *People flow 2002 Jakarta Metropolitan Area* [Data File]. https://joras.csis.u-tokyo.ac.jp/dataset/show/id/3026200200.

CSIS (2004): *People flow 2004 Hanoi Metropolitan Area* [Data File]. https://joras.csis.u-tokyo.ac.jp/dataset/show/id/3027200400.

CSIS (2009): *People flow 2009 Dhaka Metropolitan Area* [Data File]. https://joras.csis.u-tokyo.ac.jp/dataset/show/id/3028200900.

Greger, K. (2014): Spatio-temporal building population estimation for highly urbanized areas using GIS. *Transactions in GIS* (in press, DOI: 10.1111/tgis.12086).

PostgreSQL Global Development Group (2014): *PostgreSQL* (version 9.2.4). http://www.postgresql.org/.

Python Software Foundation (2014): *Python Language Reference* (version 2.7). http://www.python.org.

QGIS Development Team (2014): *QGIS Geographic Information System. Open Source Geospatial Foundation Project (version 2.4)*. http://qgis.osgeo.org/.

R Core Team (2014): *R: A language and environment for statistical computing (version 3.1.0)*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Received 5 September 2014
Accepted 4 November 2014