Chapter 1

Vehicles' Probe Data and Urban Connectivity Measurement

Souknilanh Keola[#]
Bangkok Research Center, IDE-JETRO

Masahito Ambashi
Economic Research Institute for ASEAN and East Asia

Abstract: This Chapter provides an exemplary review of connectivity measurement literature focusing on data required for quantitative analyses. It then describes vehicles' probe data and manual traffic count data collected for studies in later chapters of this report.

Keywords: Urban, Connectivity, Probe data
JEL Classification: O18, C55, C80

1. Introduction

Connectivity between different locations requires physical transport infrastructure. As such many studies have been attempting to measure connectivity through evaluation of physical transport infrastructure. For example, Fan and Chankang (2005) related road construction to development in China since the 1980s. Axhausen et al. (2008) constructed impressive long-term time scale maps of Switzerland, mainly from historical road data, to illustrate how the studied mountainous region had temporally shrunk between 1950 and 2000 as a result of increasing connectivity. Patarasuk and Bincord (2012) analyzed the impact of increased accessibility arising from road development on the land cover change in a province of Thailand between 1989 and 2006. Weiss et al. (2018) produced a global accessibility map, of the year of 2015, based on road data from OpenStreetMap (OSM) and Google.

Nevertheless, roads alone cannot fully capture connectivity in reality. This argument is rather apparent in sea or air transportation. For sea and air transport, connectivity is mostly about the frequency of traffic and rarely about whether there are water/air ways between locations. This is not to deny that sometimes canals need to be built to create new waterways. Furthermore, since most people cannot afford their own ships or airplanes, benefits from connectivity through water and air ways are mostly materialized from available traffic. Existing traffic, on the other hand, depends on various demand and

[#]Coressponding Author: Souknilanh Keola: Research Fellow, Bangkok Research Center/JETRO Bangkok, Thailand. E-mail: Souknilanh_keola@ide.go.jp

supply factors. Traffic between more populated regions, and/or larger industrial agglomeration, etc. would tend to be larger than among regions that are not. Rule and regulation that encourage factors mobility should also increase traffic, and thus the level of connectivity. In fact, most of the studies above attempted to consider traffic but had to compromise as a result of the lack of data. It is not easy to construct accurate road data even with modern technology. Quality of global road data set such the one available at Socioeconomic Data Applications Center (SEDAC) still varies significantly among countries. It is much more difficult to collect and compile high-quality traffic data. Vehicles are not detectable by most satellites with global coverage. Combinations of eyes above and the grounds are necessary. Uppoor et al. (2014) excellently summarized sources of traffic data. They categorized traffic data sources into perception and small-scaled measurements; road traffic imageries; roadside detectors; socio-demographic surveys; and real-world tracking. These sources naturally come with their own strengths and weaknesses. Small-scale surveys are cheap but very limited in spatial and temporal scopes. Traffic imageries by which car can be accurately counted are only available from aerial photographies or highly commercialized satellites, thus very costly to obtain. Roadside detectors cannot identify the origin and destination of traffic flows. Socio-demographic surveys cover extensive areas, frequently nation-wide, but at the expense of spatial and temporal resolution. This chapter explored temporally frequent GPS based probe data whose coverage is expanding but became available at relatively low costs as a result of widespread use of mobile devices with GPS signal receiver. In addition, it also looks at long term time series data of traffic counts with information of the type of vehicles.

2. Vehicles' Probe Data

For demonstration purpose, four sets of consecutive 48 hours starting at 17:00 pm local time in Thailand of March the 4th and September the 12th 2017, and March the 3rd and September the 11th of 2018 of commercial vehicles' probe data has been obtained from a location data vending arm of a major transnational car manufacturer in Thailand. The commercial vehicles include mostly trucks and taxis.

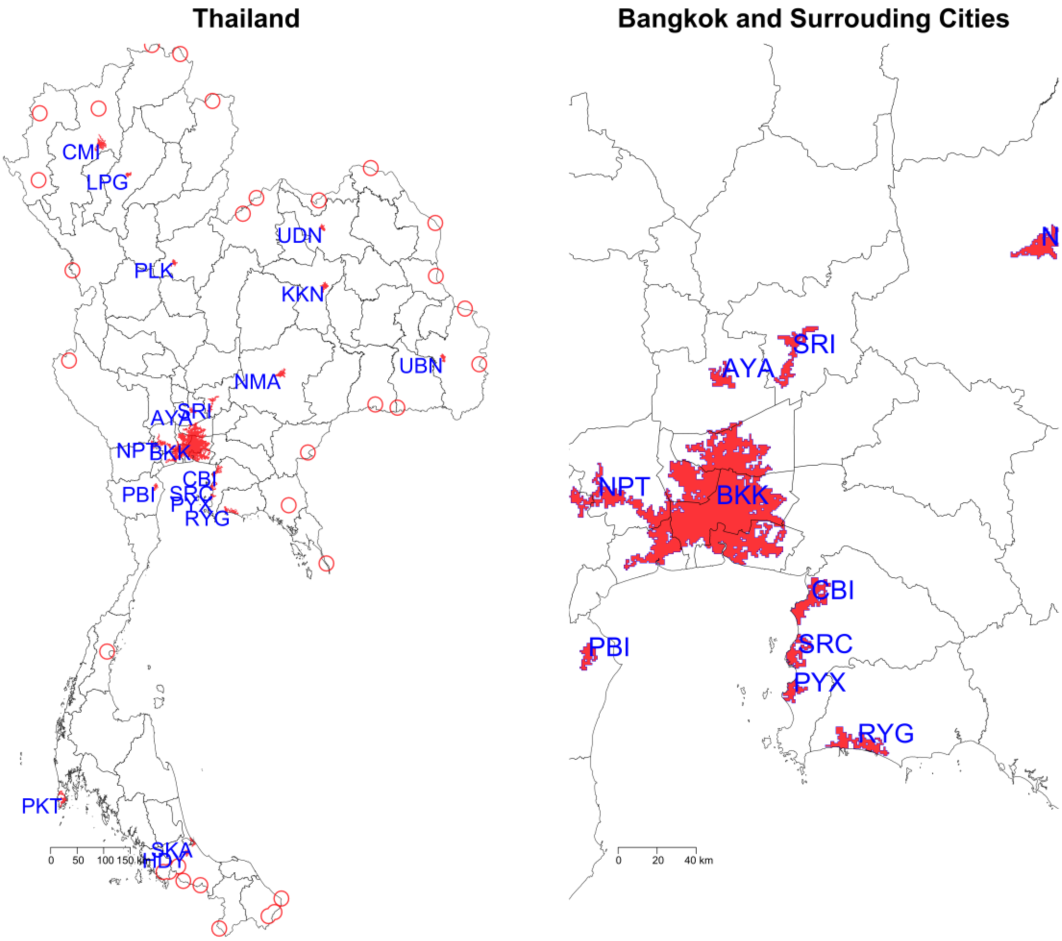
The probe data obtained for this study contains the following data.

imei: international mobile equipment identity
lat: latitude
lng: longitude
speed: kilometers per hour
direction: from 0 to 360 degree
error: from 0 to 1000
acc: engine status
meter: meter status for taxi
ts: timestamp in second from January 1st, 1970
datasource: code of data provider

Only probe data of vehicles appearing in two or more pre-defined city boundaries, or at least once in border areas as illustrated in Figure 1, were included in the data set. In order to avoid Modifiable Areal Unit Problem (MAUP), arising from the subjective and ambiguous spatial unit of analysis. The boundaries of cities in Thailand are generated following Keola (2018). The boundaries of top 19 cities in Thailand are defined by continuing area with consistent density thresholds of population and infrastructure density represented by two remote sensing data sets, i.e., LandScan and DMSP-OLS (Defence Meteorological Satellite Program Operational Linescan System) nighttime light. The thresholds of population and infrastructure density for urbanness is 200 or more people, and 40 or more nighttime light intensity per square kilometers. The differences between these remotely-sensed cities and administrative boundaries in Thailand are as follows. Even though Bangkok remains the largest cities by area in Thailand, the remotely-defined boundary is larger than the administratively defined area. The remotely-sensed Bangkok extends beyond adjacent provinces in the North, East, and South but left out some Northeastern area of its own. On the contrary, the boundary of Chiang Mai, the second largest city is much smaller than the boundary of its first-level (province) administrative boundary. This statement holds for the boundaries of all other cities derived according to the thresholds mentioned above. In other words, for most cities or urban areas administratively defined in Thailand, the physical aspects remotely-sensed by satellite is quite different from Bangkok. Furthermore, the area of Bangkok is vast when compared to the rest of the cities in Thailand. The area of Bangkok is about 3,411 km^2, while it is 310 km^2 for Chiang Mai, or 243 km^2 Nakhon Pathom. However, I argue that this reflects rightly the polycentric urbanization in Thailand. Unless otherwise stated, the area names in this chapter refer to the boundaries generated by the density of remotely-sensed population and Nighttime light.

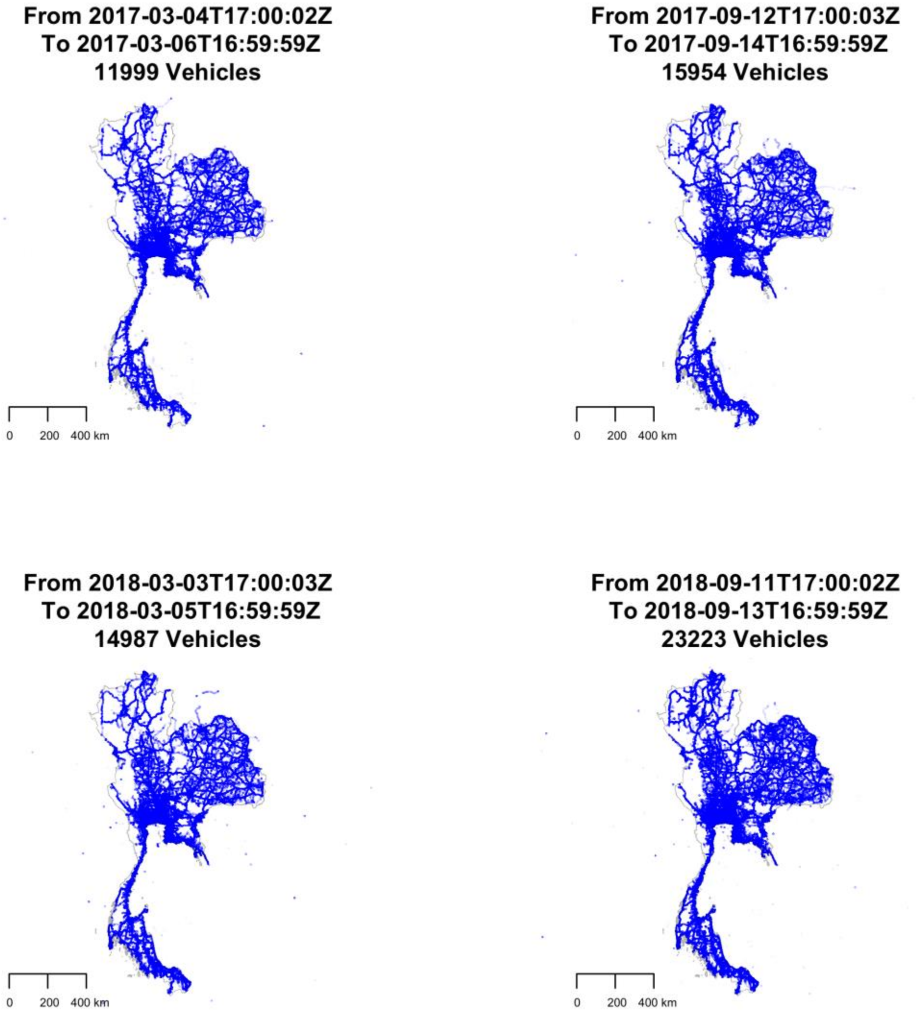
The following cleaning is first applied. Speed is computed for each data points based on positions and time stamp of previous data points. This is especially important for GPS-based probe data because the accuracy of reported location vary due to many factors affecting the strength of receding satellite signal. In fact, many locations in other continents are observed, which resulted in unrealistic enormous speed. Data points with speed higher than 120 km/h are then excluded. This threshold is selected based on the fact that it is quite difficult for most trucks to travel beyond this speed in Thailand. It may be easier for taxis travel faster, but that would violate a traffic law, and thus sensible to exclude them when evaluating intra- or inter-urban connectivity. Many of data points with unrealistic location information are removed through the above cleaning process.

Figure 1. Top 19 Remotely-Sensed Cities and Border Areas in Thailand



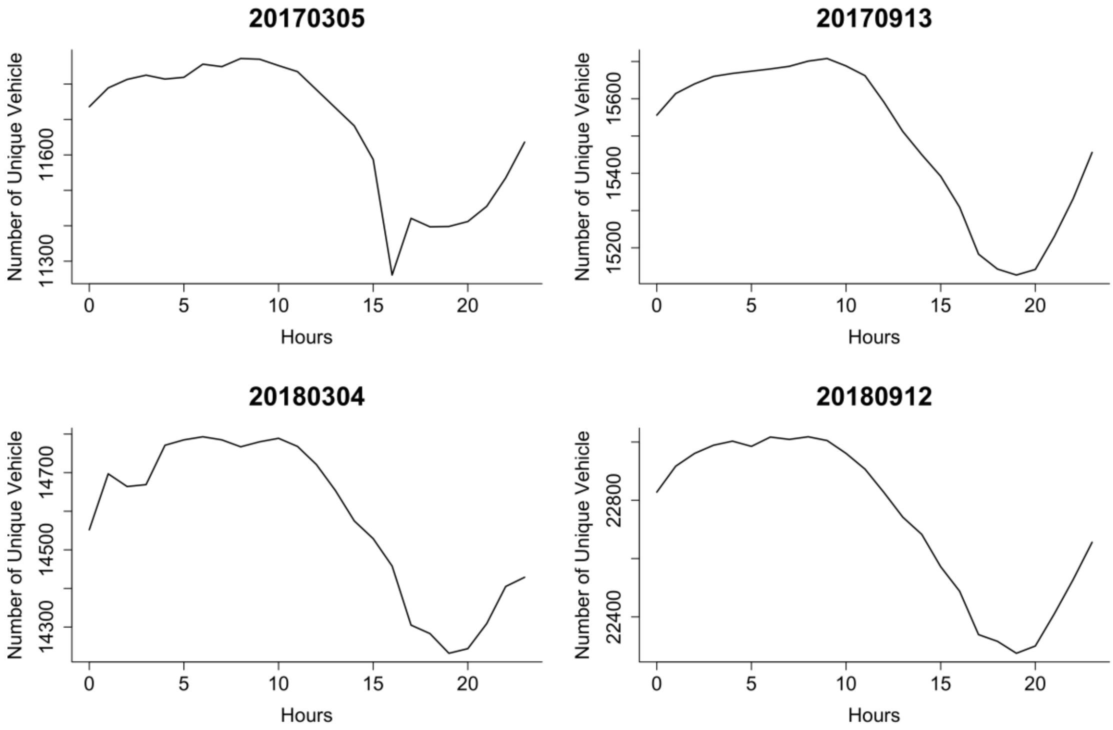
Source: City boundaries (red) is based on Keola (2018). Border area boundaries (blue) is a 10 kilometers buffer from the location of nearest custom houses. The administrative boundary is based on GADM.
Note: AYA: Ayutthaya, BKK: Bangkok, CBI: Chonburi, CMI: Chiang Mai, HDY: Hat Yai, KKN: Khon Kaen, LPG: Lampang, NMA: Nakhon Ratchasima, NPT: Nakhon Pathom, PBI: Phetchaburi, PKT: Phuket, PLK: Phitsanulok, PYX: Pattaya, RYG: Rayong, SKA: Songkhla, SRC: Sriracha, SRI: Saraburi, UBN: Ubon Ratchathani, UDN: Udon Thani.

Figure 2. Location of All Obtained Probe Data Points



Source: Author based on probe data.

Figure 3. Hourly Number of Vehicles



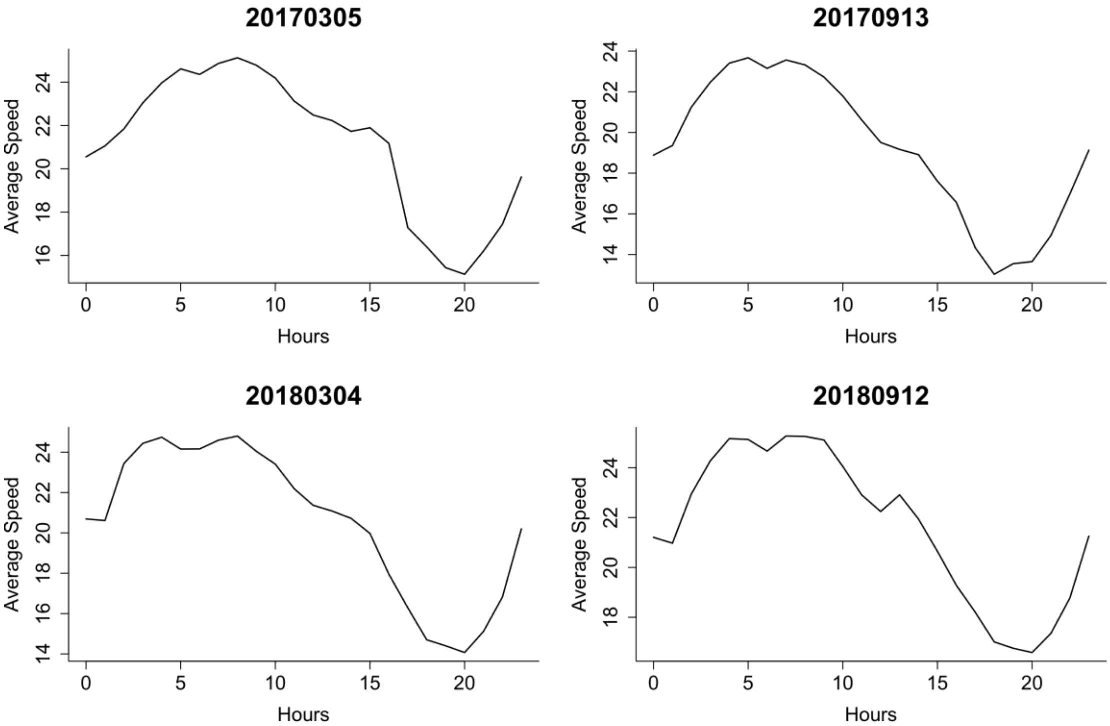
Source: Author based on probe data.

Some summary statics are provided here to demonstrate what may be done on this data. Figure 3 depicts the number of vehicles based on unique imei by the hour between 00 am to 11 pm. It should be noted that for each period, data starts from 5 pm of the previous day, and last for 48 hours. The peak of the number of vehicles, whose location data is reported, increases to the highest level between 5 to 10 am before declining to the bottom around 8 pm throughout all four periods of 48 hours. Figure 4 shows the change in average speed by time of the day. The curves in Figure 4 resemble that of curves in Figure 3 suggesting unintuitively that speed increases when more vehicles are on the streets. The fact that some of the vehicles report location data when they are standing still, or even when the engine is not turning on may partly account for this result. The other may be that the average decreases even when there are less vehicles on the street because demand for mobility during that period is also low.

The location information, i.e., longitude and latitude, of probe data make it very easy to generate similar summary statistics on a geographical subset of the original data. For example, we generate an hourly number of vehicles by cities as in Figure 5, after geotagging each data points with boundaries of the top 19 cities in Thailand. More or less than about 4,000 vehicles in obtained probe data are in Bangkok. This is more than five times the levels in the cities with second (Chonburi) and third (Sriracha) the largest number of vehicles. Chiang Mai, the second largest city by area, does not show the second largest number of vehicles, but this can be a result of the fact that this probe data set represents only a part of the whole vehicles in Thailand. Nonetheless, the peak of the number of samples in Chiang Mai is merely about 150, much fewer than Chonburi and

Sriracha.

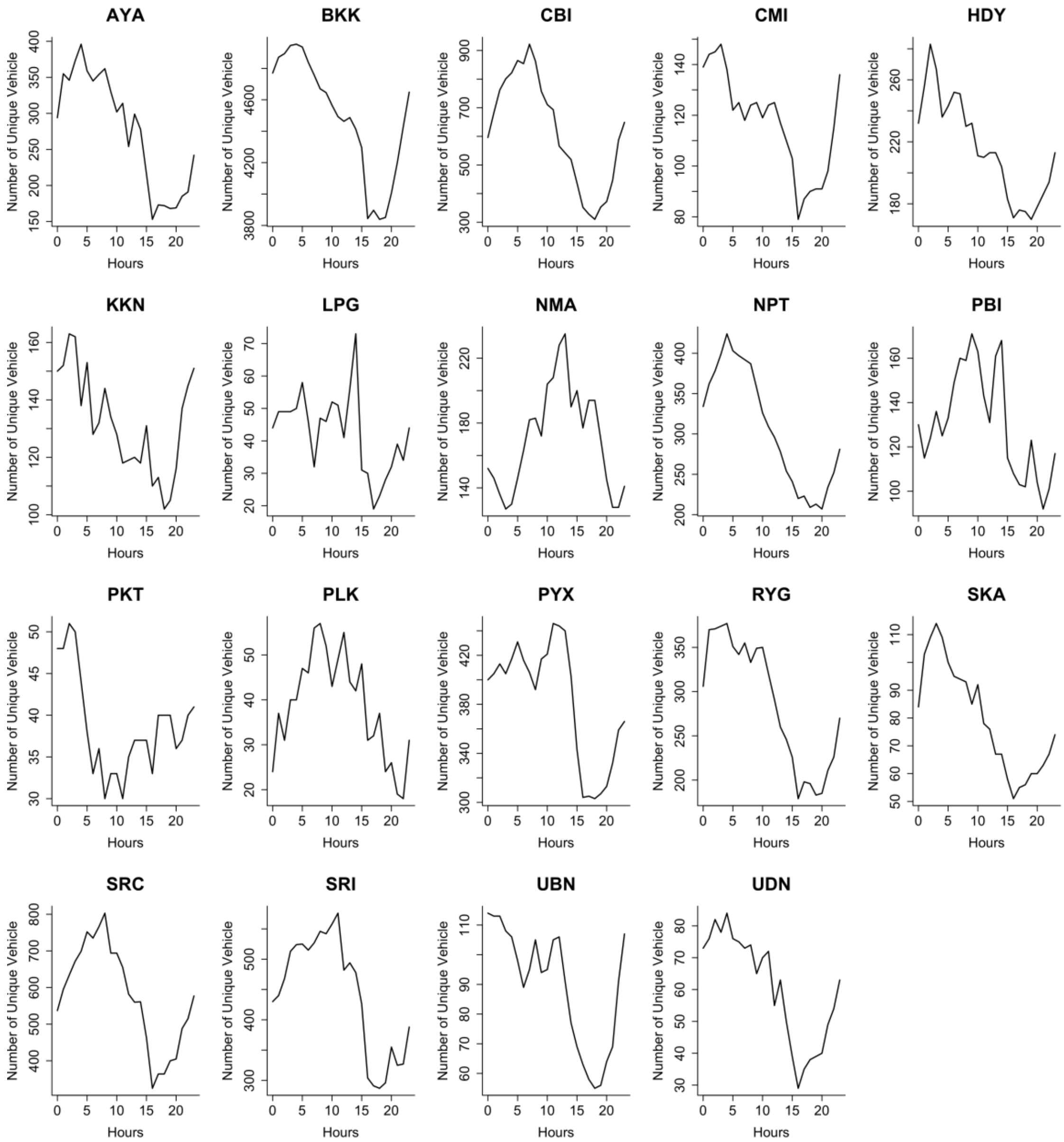
Figure 4. Hourly Average Speed (km/h)



Source: Author based on probe data.

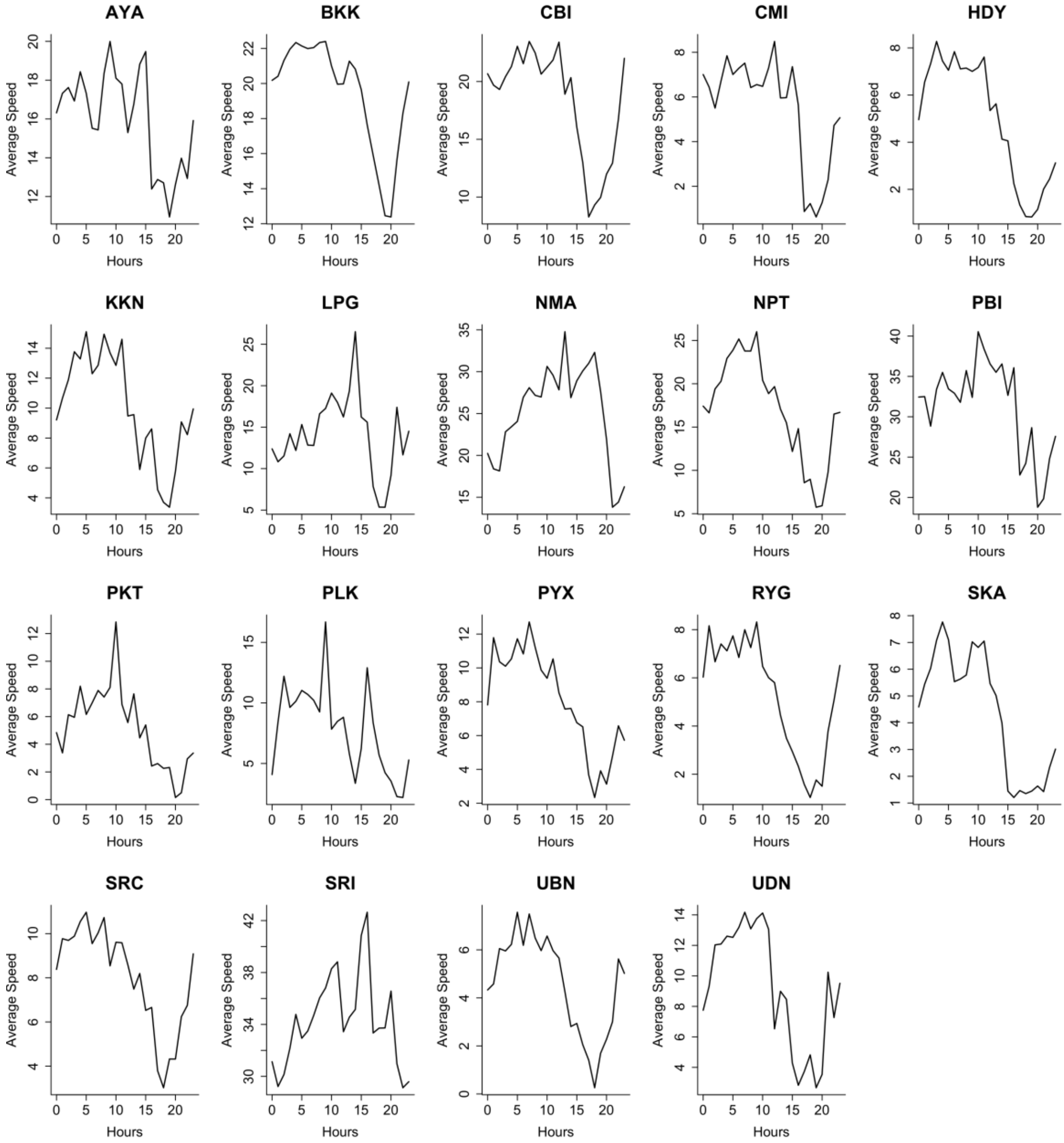
Figure 6 shows average speed by cities and by the time of the day. Figure 6 looks counter-intuitive at first glance because the seemingly most congested Bangkok does not indicate the smallest average speed. On the contrary, the average speed in remote cities such as Ubon Ratchathani and Hat Yai fluctuate around 2 and 8 km per hour. Average speed in Chiang Mai is also similarly very low. On the other hand, the average speed in Bangkok fluctuates between 12 and 22 kilometers per hour. One possible explanation may be that more cars can move in Bangkok, as there are much more customers and goods to carry than smaller cities. The smaller demand in smaller cities can also be observed in Figure 7. Average distance traveled is the longest in Bangkok about 14 kilometers between 5 am and 3 pm followed by nearby cities such as Nakhon Ratchasima, Nakhon Pathom, and Phetchaburi.

Figure 5. Hourly Average Number of Vehicle by Cities



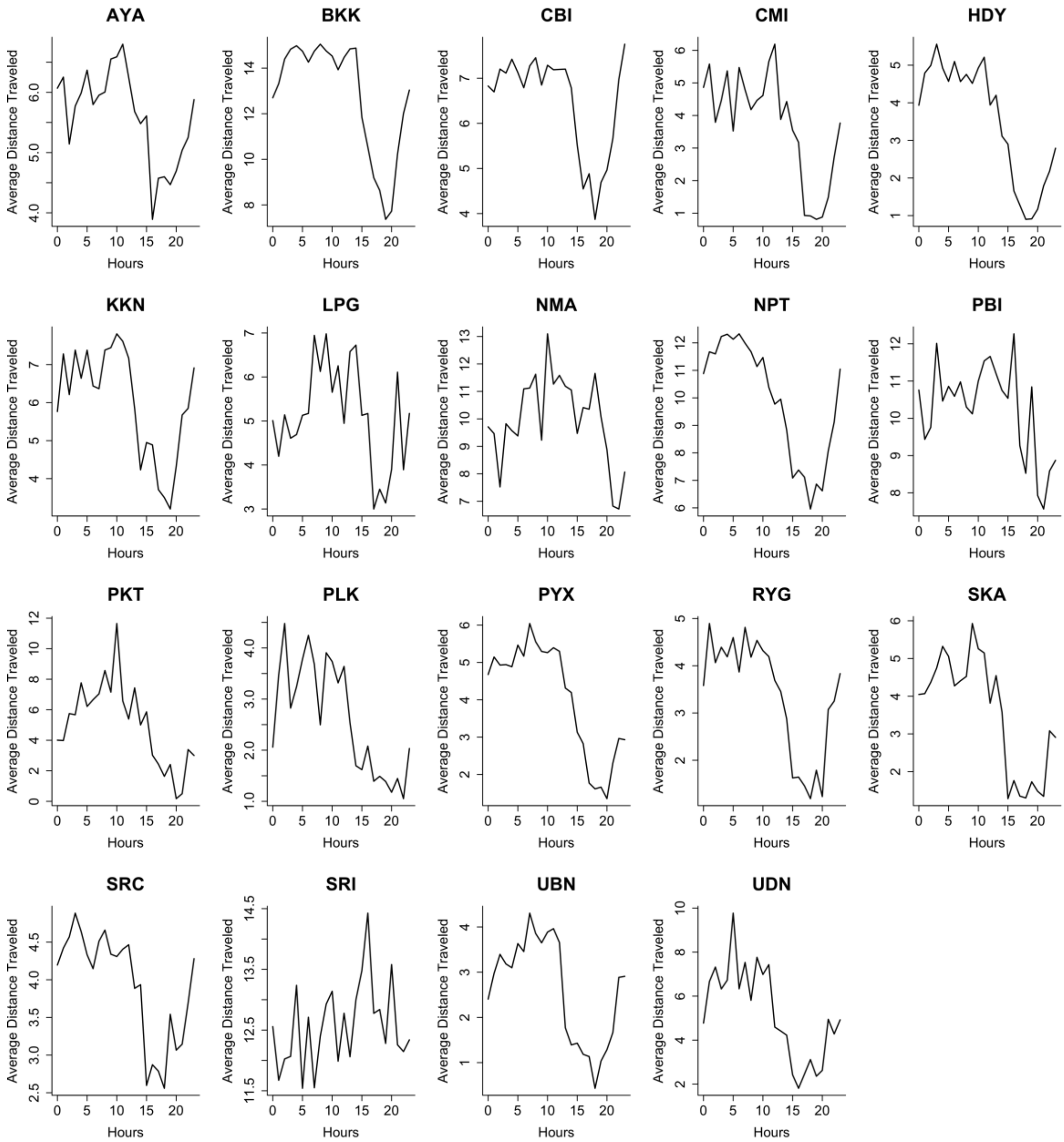
Source: Author based on probe data.
Note: AYA: Ayutthaya, BKK: Bangkok, CBI: Chonburi, CMI: Chiang Mai, HDY: Hat Yai, KKN: Khon Kaen, LPG: Lampang, NMA: Nakhon Ratchasima, NPT: Nakhon Pathom, PBI: Phetchaburi, PKT: Phuket, PLK: Phitsanulok, PYX: Pattaya, RYG: Rayong, SKA: Songkhla, SRC: Sriracha, SRI: Saraburi, UBN: Ubon Ratchathani, UDN: Udon Thani.

Figure 6. Hourly Average Speed of Vehicle by Cities (km/h)



Source: Author based on probe data.
Note: AYA: Ayutthaya, BKK: Bangkok, CBI: Chonburi, CMI: Chiang Mai, HDY: Hat Yai, KKN: Khon Kaen, LPG: Lampang, NMA: Nakhon Ratchasima, NPT: Nakhon Pathom, PBI: Phetchaburi, PKT: Phuket, PLK: Phitsanulok, PYX: Pattaya, RYG: Rayong, SKA: Songkhla, SRC: Sriracha, SRI: Saraburi, UBN: Ubon Ratchathani, UDN: Udon Thani.

Figure 7. Hourly Average Traveled Distance of Vehicle by Cities (km)



Source: Author based on probe data.

Note: AYA: Ayutthaya, BKK: Bangkok, CBI: Chonburi, CMI: Chiang Mai, HDY: Hat Yai, KKN: Khon Kaen, LPG: Lampang, NMA: Nakhon Ratchasima, NPT: Nakhon Pathom, PBI: Phetchaburi, PKT: Phuket, PLK: Phitsanulok, PYX: Pattaya, RYG: Rayong, SKA: Songkhla, SRC: Sriracha, SRI: Saraburi, UBN: Ubon Ratchathani, UDN: Udon Thani.

3. Thailand’s Annual Average Daily Traffic (AADT)

This section describes Thailand's Annual Average Daily Traffic obtained to compliment vehicles' probe data. The data represents the Annual Average Daily Traffic (AADT) during 2007-2017, collected by the Department of Highways, Ministry of Transport. Particularly, the data indicates:

(1) AADT of nationwide major highways and
(2) AADT of main streets in Bangkok's metropolitan area.

This dataset includes the volume of each specific classification by vehicle. The Department of Highways officially utilizes such data to evaluate congestion and the requirement for future extension of the highways’ capabilities. The data was manually collected by officers of the regional and provincial divisions of the Department of Highways. Following the administrative process, the director of each regional and provincial division verified the accuracy of data. Then the data was submitted to the central server via https://tims.doh.go.th/. Finally, statisticians at the headquarters of the Department of Highways compile the data and upload it to the website for public access. All data has been gathered and transformed into the GIS format. The attribute table file (.dbf) of GIS format contains the AADT of major highways during 2007-2017. Full details of the data structure are listed in Appendix A. The classifications of vehicles are listed below.

1) Passenger car (less than 7 persons)
2) Passenger car (more than 7 persons)
3) Small bus
4) Medium bus
5) Large bus
6) Small truck (4 wheels)
7) Medium truck (6 wheels)
8) Large truck (10 wheels)
9) Full trailer
10) Semi-trailer
11) Motorcycles

To extend the insight on this classification, example pictures of each vehicle classification are illustrated below.

Figure 8. Various Types of Vehicles in Thailand

(1) Passenger car (less than 7 persons)

(2) Passenger car (more than 7 persons)

(3) Small bus

(4) Medium bus

(5) Large bus

(6) Small truck (4 wheels)

(7) Medium truck (6 wheels)

(8) Large truck (10 wheels)

(9) Full trailer

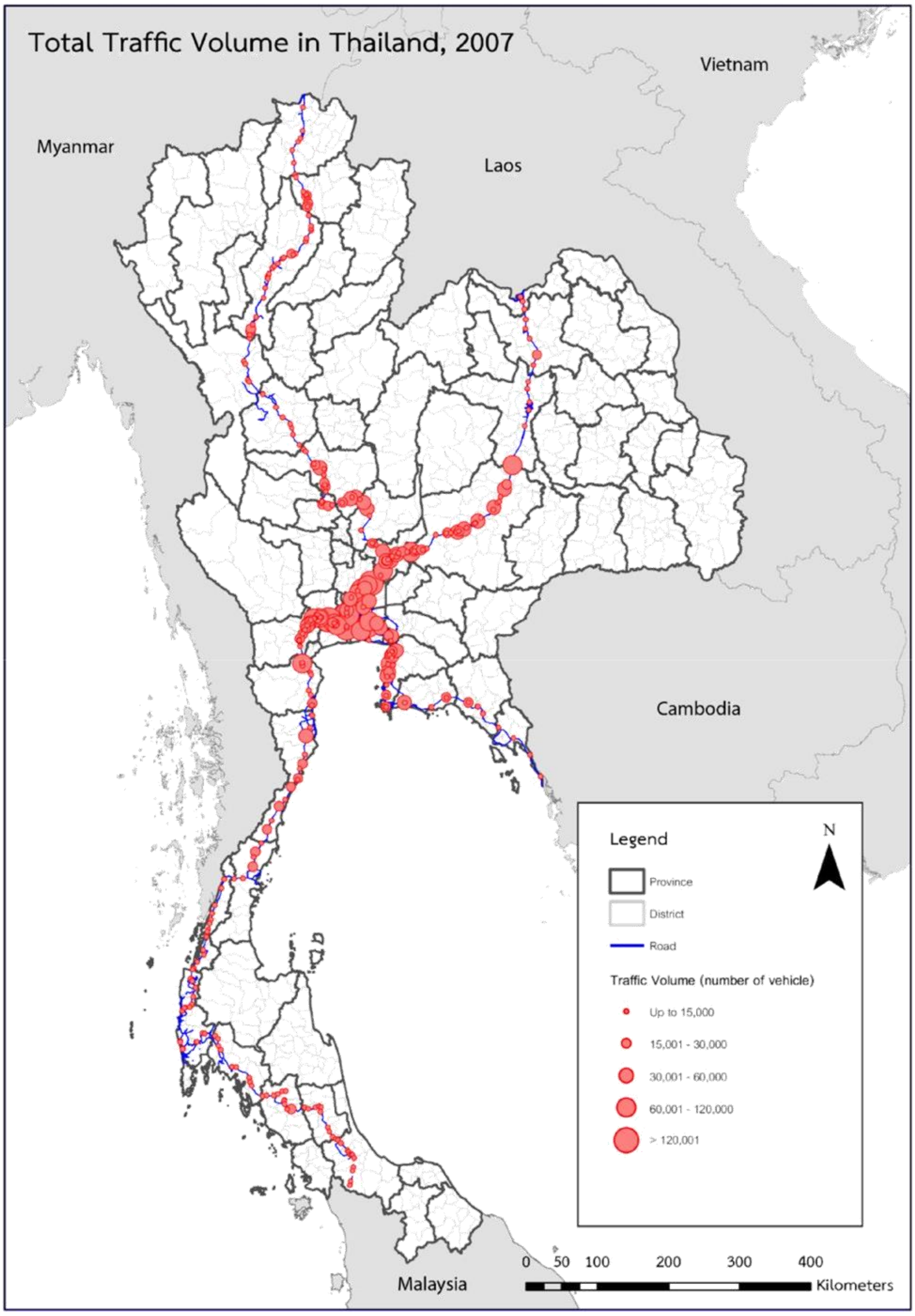
(10) Semi-trailer

(11) Motorcycles

Source: Department of Highways, Ministry of Transport

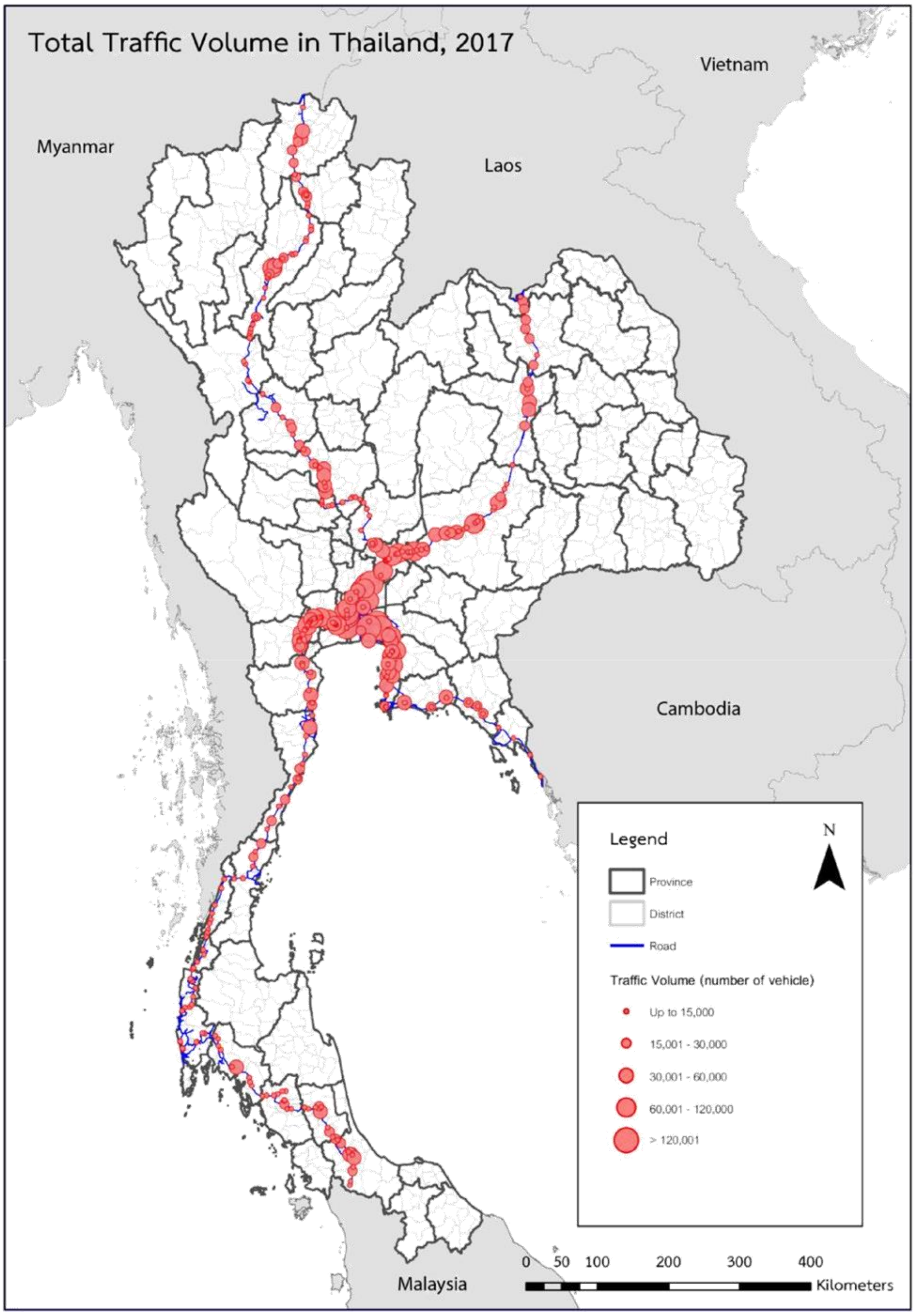
This dataset is very resourceful for various analyses related to the spatial distribution of population and economic activity. Appendix C shows the collection of maps exhibiting the density of traffic volume during 2007-2017. Notably, the traffic volumes are significantly proportionate to the distance from Bangkok, and they also relate to the size of local economic activity. Figure 9 and 10 shows traffic volume in each location in 2007 and 2017. Traffic in Bangkok metropolitan area increases substantially although this is difficult to judge visually. The increase of traffic towards major border gates with Malaysia in the South, Laos in the Northeaster, and Laos and Myanmar in the North can be confirmed visually.

Figure 9. Traffic Volume in Thailand in 2007



Source: IDE-JETRO and Thammasat University.

Figure 10. Traffic Volume in Thailand in 2017



Source: IDE-JETRO and Thammasat University.

Summary

In this chapter, we made an exemplary review of connectivity measurement literature focusing on data required for quantitative analyses. Measuring connectivity without flow data risk confusing of a necessity with the sufficient conditions. GPS-based vehicles’ probe is undisputedly one of the best sources of currently available data to measure connectivity with consideration of traffic. The cost of generating and obtaining vehicles’ probe data has become comparatively cheap, with wide-spreading usages of GPS-enable mobile devices. Actual spatial extension, intensity, and speed of connectivity can be quickly done with a sufficiently large amount of probe data. This chapter has also shown that much can be done even with 48 hours of probe data.

References

Axhausen, K. (2008). Accessibility: long-term perspectives. *Journal of Transport and Land Use*, 1(2), 5-22.

Fan, S., & Chan-Kang, C. (2005). *Road development, economic growth, and poverty reduction in China* (Vol. 12). Intl Food Policy Res Inst.

Patarasuk, R., & Binford, M. W. (2012). Longitudinal analysis of the road network development and land-cover change in Lop Buri province, Thailand, 1989–2006. *Applied Geography*, 32(2), 228-239.

Schüssler, N., & Axhausen, K. W. (2008). Identifying trips and activities and their characteristics from GPS raw data without further information. *Arbeitsbericht Verkehrs-und Raumplanung*, 502.

Taylor, M. A. (2012). Remoteness and accessibility in the vulnerability analysis of regional road networks. *Transportation research part A: policy and practice*, 46(5), 761-771.

Uppoor, S., Trullols-Cruces, O., Fiore, M., & Barcelo-Ordinas, J. M. (2014). Generation and analysis of a large-scale urban vehicular mobility dataset. *IEEE Transactions on Mobile Computing*, 13(5), 1061-1075.

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., ... & Mappin, B. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688), 333.