# ESTIMATION OF PUBIC TRANSPORT USERS FROM CALL DETAIL RECORDS IN YANGON, MYANMAR

**Nitchanan Prapaitrakul, Yoshihide Sekimoto, Takehiro Kashiyama, and Ko Ko Lwin**

## Background

Estimation of public transport users in an area is valuable for urban and transportation planning purposes, particularly to understand the movement of commuters between their origins and destinations for strategically locating public transport stations or stops, and to estimate the passenger demand in order to correspondingly set an appropriate supply of public transport. This paper proposes an alternative method that is relatively low cost and less complex than conventional demand estimation methods such as trip surveys.

## Call Details Records ; CDR

record of a telecommunications transaction such as a call or text message collected by the *Cell Towers* of the service provider.
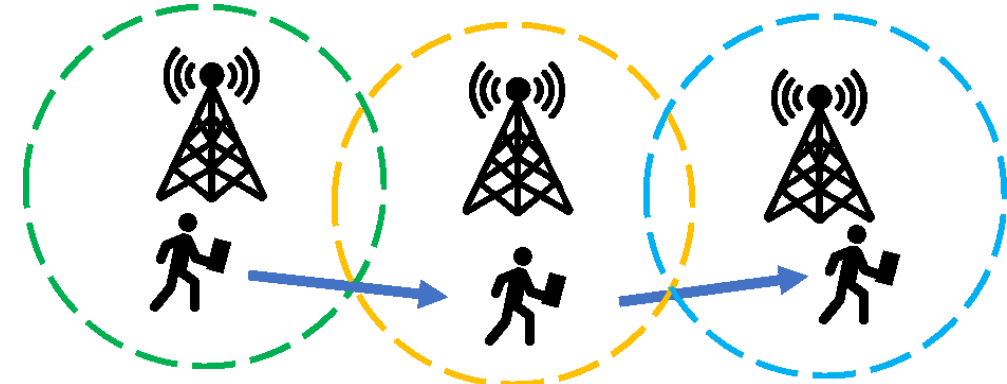
## Methodology
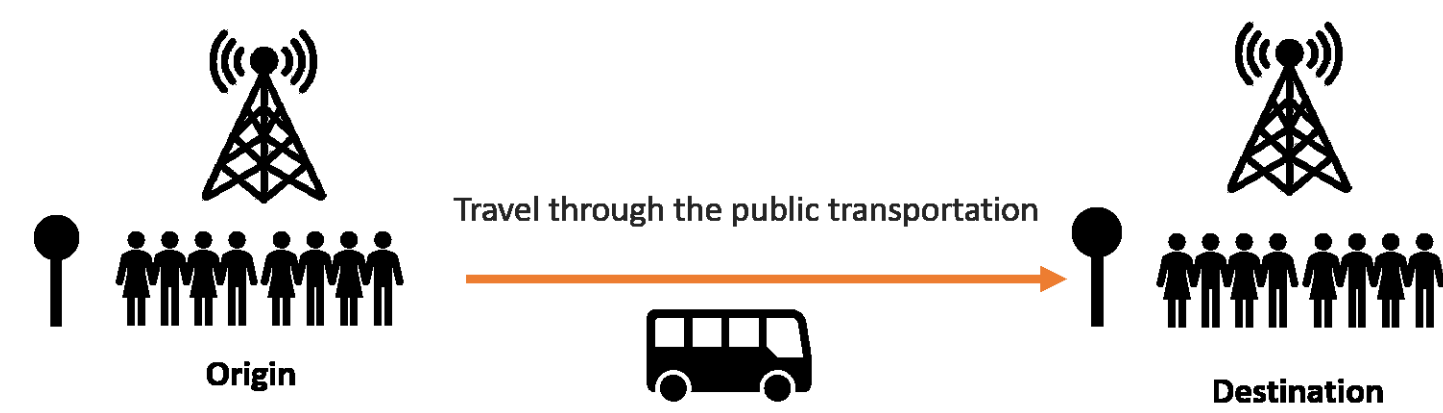
The methodology in this paper is based on the basic habits of people using the public transport in their daily life.

### DATASET DESCRIPTION

Around 1.9 million users per day from 1st to 7th of December 2015 in Yangon, Myanmar.

### PERSON TRIP OD

The person trip origin-destination (PTOD) extracted and converted from the CDR dataset. The PTOD data attributes extracted include the SIM card ID, origin Voronoi ID, destination Voronoi ID, origin longitude, origin latitude, destination longitude, destination latitude, departure time, and arrival time.

### DATA PRE-PROCESSING AND FILTERING

Weekday Data, Speed between 15.3 km/h – 32.9 km/h +-5
*average travel speed range in Yangon's roads [Train speed :  24 km/h]
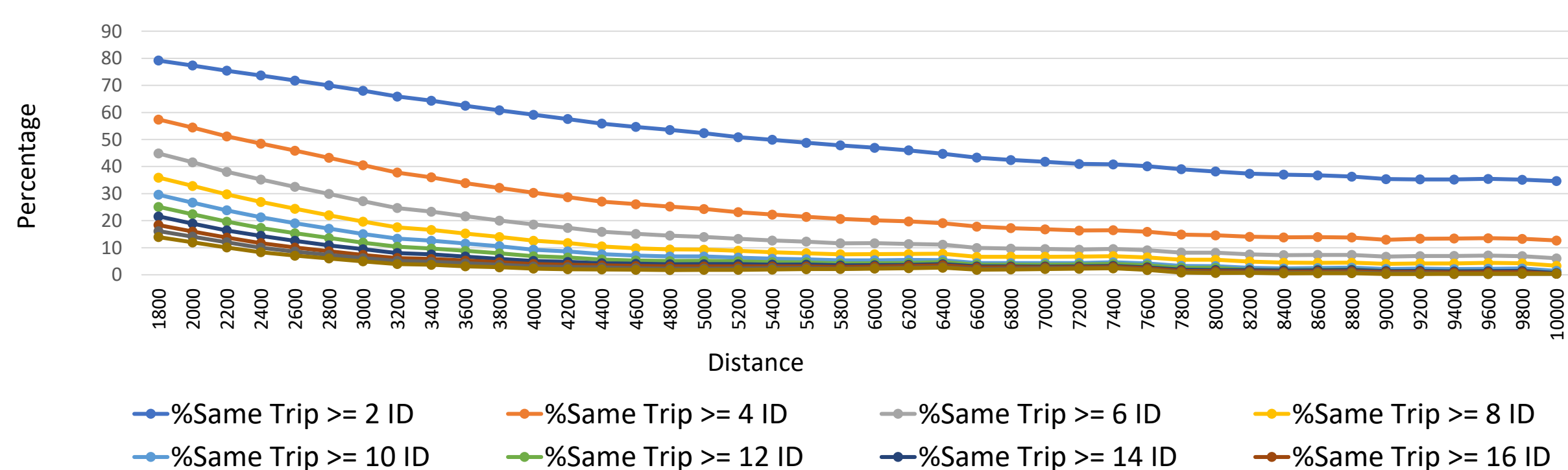
### OPTIMAL VARIABLES

#### Variable 1 : Minimum Distance Feature

There are some PTOD records that show **"jumping"** between their stay location and the imminent cell tower even if they are not actually moving or just walking around the same place.



Jumping Error

— % Amount of data after noise cutting   — % Same ID,Trip   — % Same ID,Trip,Date
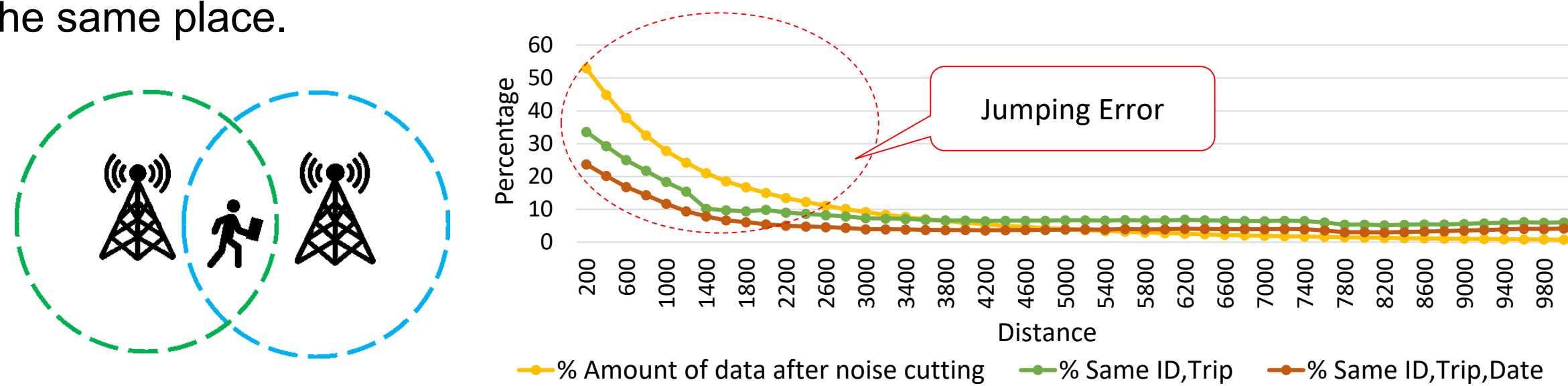
#### Variable 2 : Minimum Number of People Using the Same Trip

% amount of data and the number of  people using the same origin and destination trips



—♦— %Same Trip >= 2 ID    —♦— %Same Trip >= 4 ID    —♦— %Same Trip >= 6 ID    —♦— %Same Trip >= 8 ID
—♦— %Same Trip >= 10 ID   —♦— %Same Trip >= 12 ID   —♦— %Same Trip >= 14 ID   —♦— %Same Trip >= 16 ID

☐ extract PTOD files from the following conditions:

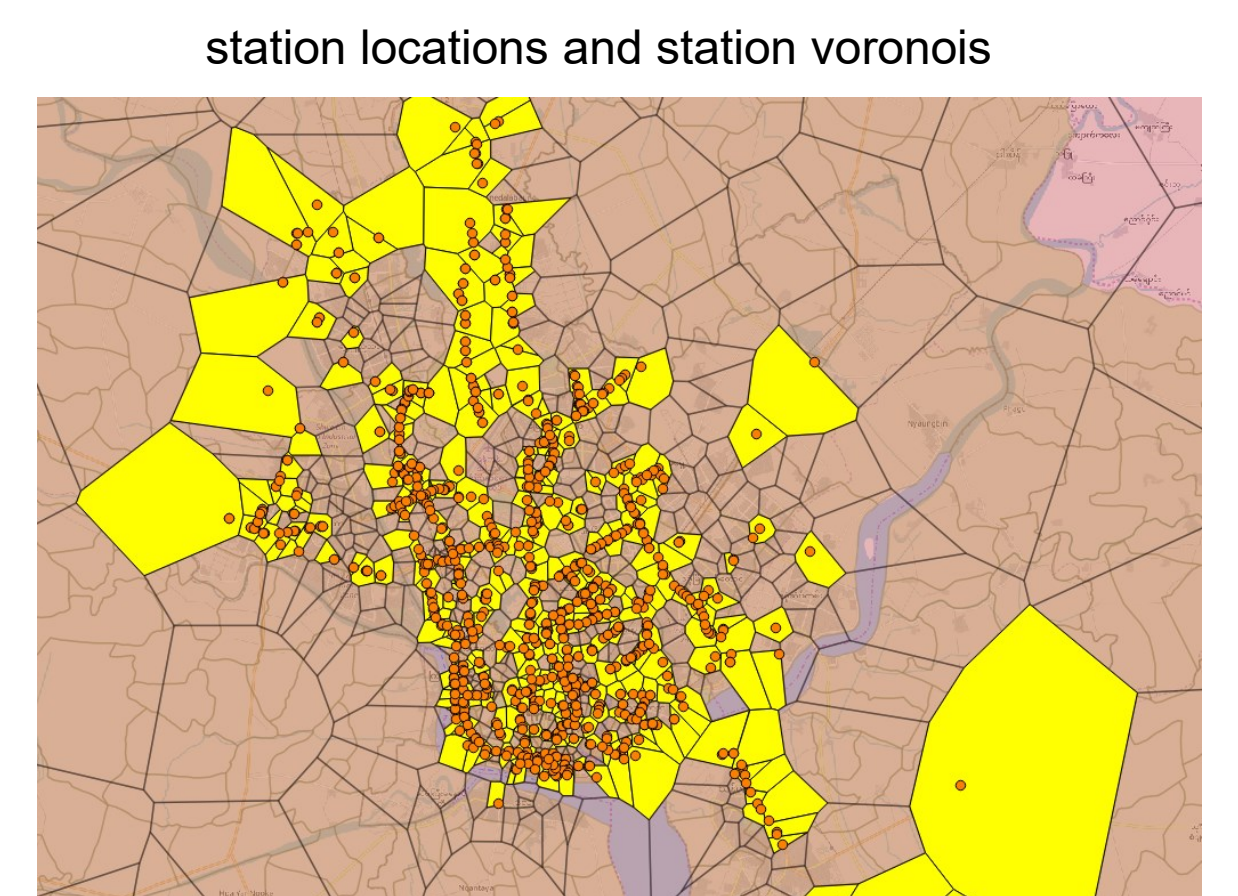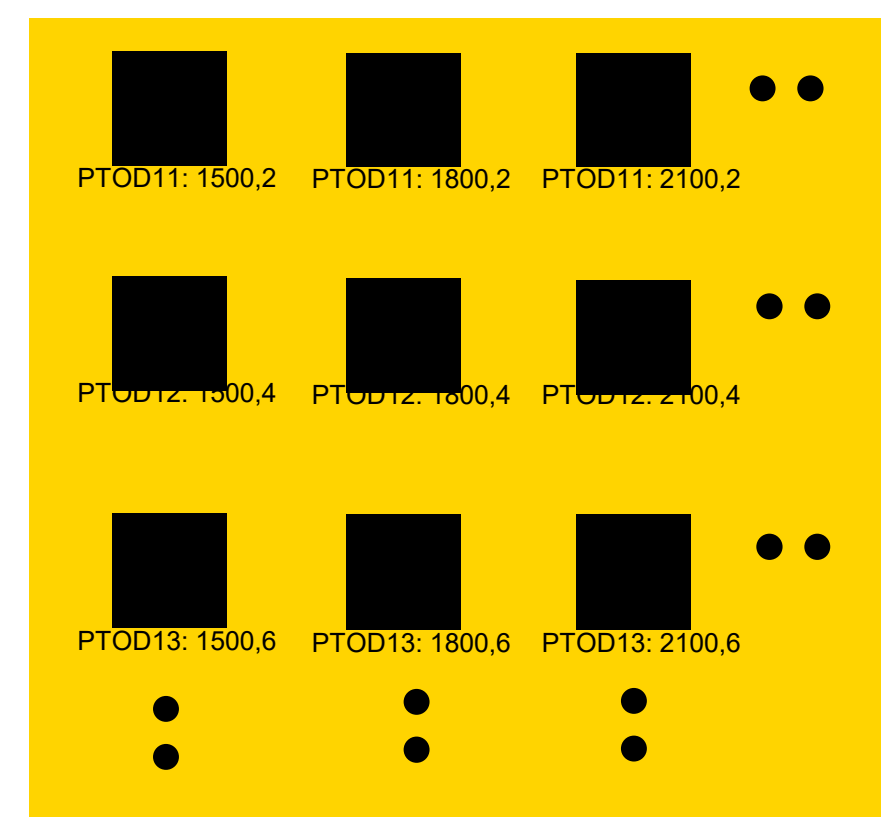$$PTOD_{ij} = \{a_k \mid a_{k,d} \geq D_i\} \,\forall\, \{a_k \mid a_{k,p} \geq P_j\}; \; a \in PTOD_{ij}$$

Where $D$ = { 1500,1800,2100,2400,2700 };  the set of distance value (meters)

$P$ = { 2, 4, 6, …, 20 }; number of People using the same trip
$k$  is the row index
$d$  is the trip distance
$p$  is the number of people using the same trip

station locations and station voronois



☐ calculate the accuracy of each extracted file by counting the number of similar Voronoi IDs from the constructed files and the station locations.

### DPERCENTAGE ACCURACY REGARDING TWO VARIABLES

| | | Number of people using the same trip | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
| Minimum distance (m.) | 1500 | 84.01 | 83.74 | 83.35 | 83.06 | 82.63 | 82.34 | 81.95 | 81.75 | 81.32 | 80.99 |
| | 1800 | 84.32 | 83.90 | 83.20 | 82.66 | 81.97 | 81.50 | 80.88 | 80.57 | 79.92 | 79.42 |
| | 2100 | 85.40 | 84.95 | 84.10 | 83.43 | 82.67 | 81.95 | 81.23 | 80.63 | 79.64 | 78.44 |
| | 2400 | 86.08 | 85.35 | 83.80 | 82.45 | 81.20 | 80.06 | 78.97 | 77.26 | 75.44 | 72.97 |
| | 2700 | 85.61 | 84.50 | 81.52 | 78.74 | 76.44 | 75.02 | 73.64 | 71.07 | 67.89 | 61.70 |

The maximum accuracy in Table 1 is **86.08%**, having a minimum distance value of **2,400 m**, and the minimum number of people who use the same trip is **2**

### RESULTS
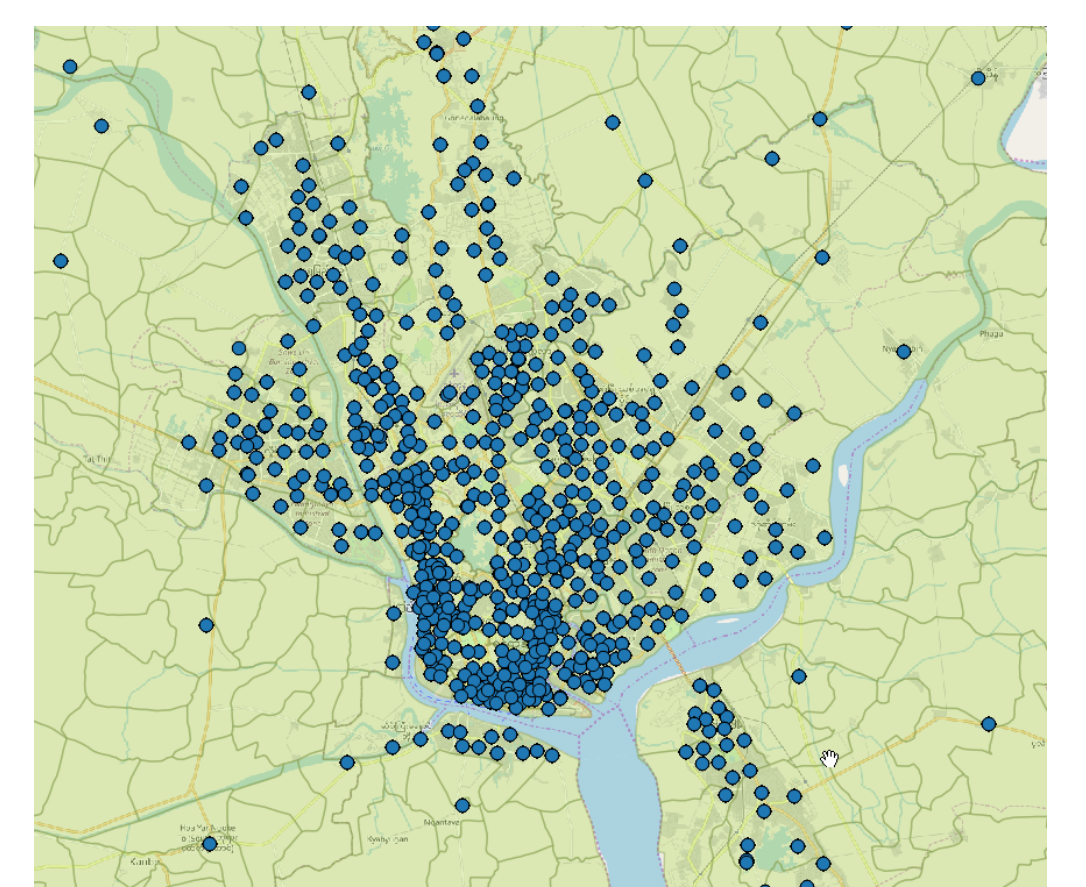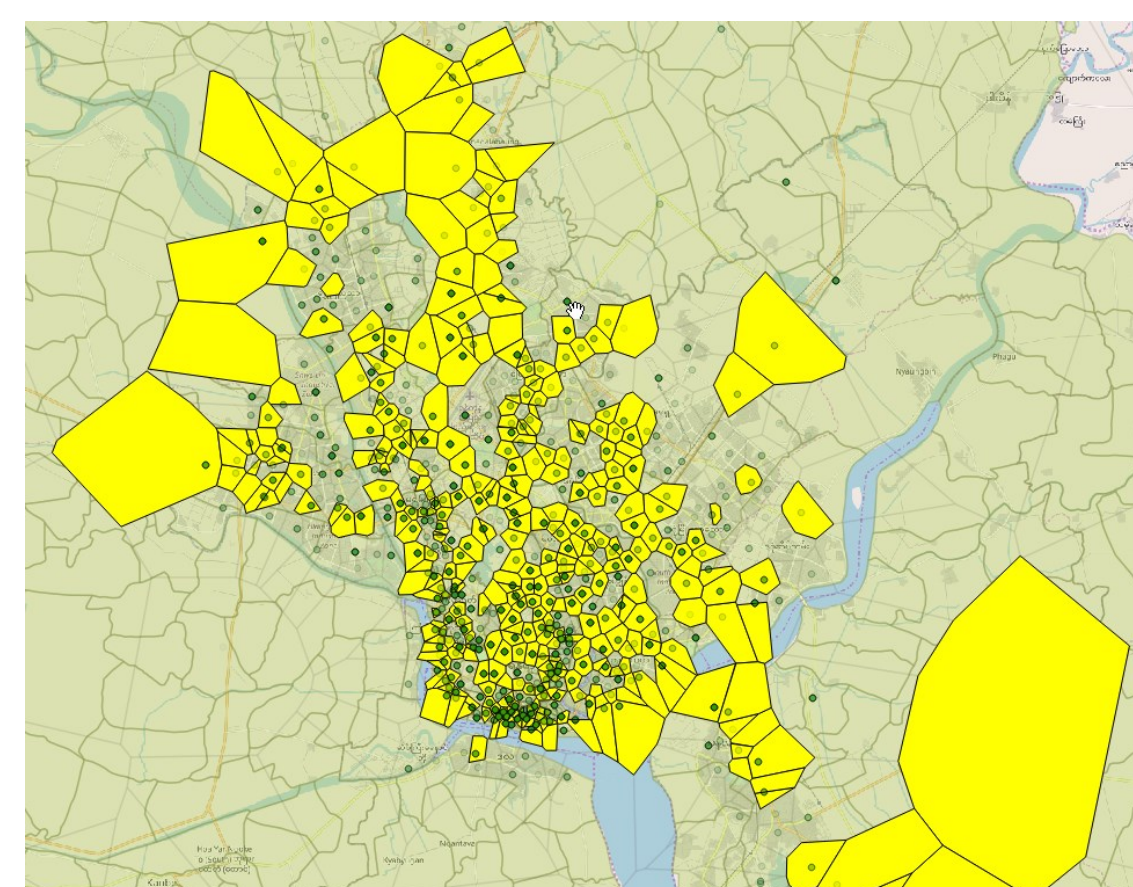


Left figure presents the best extracted PTOD data locations (green dots) with regards to two optimal variables as shown in Table 1 and are expected to be the public transport users. When comparing to the original PTOD data the PTOD data using this methodology is more related to the station locations than the original PTOD data locations on the right figure.

Moreover, the average speed obtained from this PTOD data is 32.62 km/h, which is similar to the average speed of the public transport in Yangon City's roads (32.9 km/h) as reported in the JICA survey report.

## Conclusions

This paper shows a new simple method how to extract the public transport users from CDR data in the Yangon area. Many researches have tried to extract the vehicle modes from some attributes in CDR records such as speed, stay points durations, etc.. The difficulty is how to separate the private vehicles and the public transport vehicles on the roads. We try to use another way to solve this problem in order to urban and transportation planning. This paper shows a new simple method to extract public transport user data. The future work, we will combine our method and others method to improve our model in order to extract the PTOD trip of the public transport users from the original PTOD data in Yangon then using it in term of the public transport management mobility considering the different time in a day.