Application of Big Data for Transport Planning and Practices: Case Study on (i) GPS Log for Traffic Modeling and (ii) Bus Probe Data for Improved Traffic Control in Phnom Penh

Kiminari Takahashi

International Development Center of Japan

JICA Expert for PPTMTC

1. Conventional traffic survey data vs “Big Data”

1.1 Conventional traffic survey for transport planning and practices

1.2 Limitation of conventional traffic survey for transport planning and practices

- Small sample, assumed to represent ‘average day of neutral month’
- Expensive but quality not always good
- Increasing resistance to answer questionnaire, omission or simplification of journeys
- Variable data quality: errors, sampling bias
- Donor driven which makes difficult to maintain by local cities in recipient developing countries
- Rapid growth and innovation in developing countries requires speedy and timely baseline data for transport planning and implication

1.3 Potential “big data” applied for transport planning and practices

Mobile station data: Door-to-door trips, high sample rate, high representativity

Smart cards: Stop-to-stop trips mostly by public transport

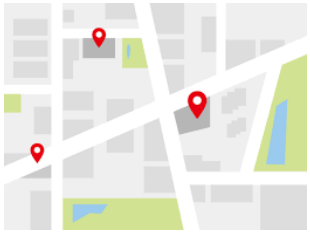
Mobile GPS log: Door-to-door trips, App dependent and bias → *Transport modeling in Phnom Penh*

Vehicle GPS probe data: Road network speed

→ *Improved traffic control in Phnom Penh*

Wifi and Bluetooth: Local OD pattern

1.4 Limitation of big data in case of mobile data



- They do not provide key data for transport planning
  - Lack of personal and family attribute
- App dependent and data bias
  - Unable to identify apps
  - Single GPS log and over 1000 logs in one ID
- Insufficient spatial resolution (network) and discontinuities in the data (apps)
  - Difficult in identifying transport mode

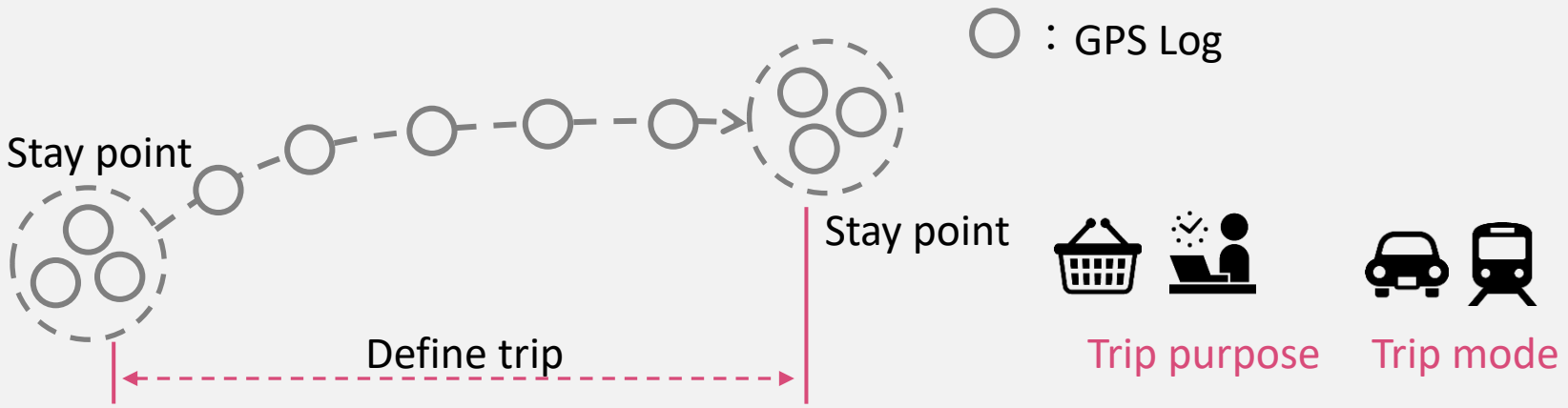
→ Algorithms for data cleaning, sample selection, analysis, data fusion and sample expansion are vital for transport planning and practices

2. GPS Log data for transport planning in Phnom Penh

Research by Kai Kusunose and Prof. Kuniaki Sasaki

2.1 Objectives of research

1 Use mobile GPS log data to extract trips, estimate trip purpose and trip mode.



2 Examine whether estimation using mobile GPS log data be a substitute for conventional person-trip surveys.

Source: Kusunose and Sasaki (2023)

2.2 Features of Mobile GPS log data

Mobile GPS log data purchased from private log data suppliers (i.e., Lifesight and Quadrant)

One month data (January 2020) acquired and data volume summarized below

Cambodia

No of logs: 117,123,391
No of ID: 87,876

Phnom Penh

No of logs: 74,771,575
No of ID: 45,176

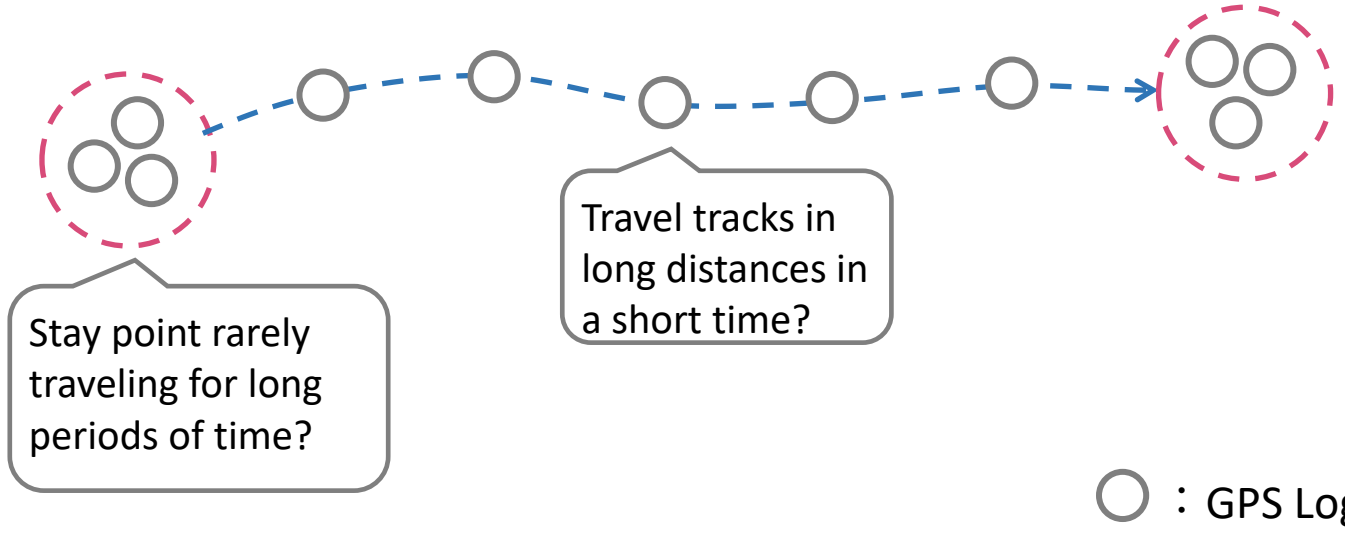
Source: Kusunose and Sasaki (2023)

2.3 Definition of ‘Stay Point’

【Features of GPS log data】

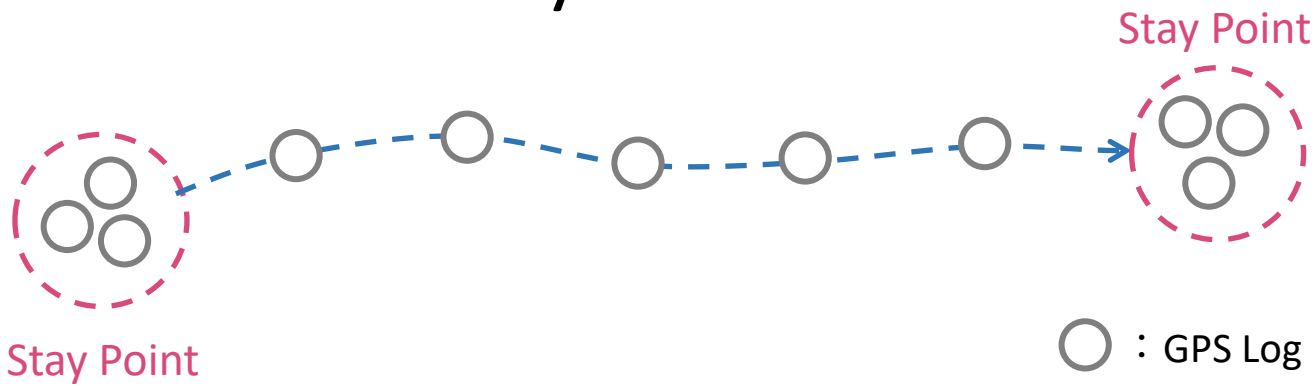
A sequence of points with time and location information without trip attribute information.

▶ **Necessary to define trip from the relationship between the acquisition time difference and distance/location between plots.**



Source: Kusunose and Sasaki (2023)

2.3 Definition of ‘Stay Point’



Stay Point (SP) defined as group of plots that remain within a certain range for more than a certain amount of time

① Stay more than 5 minutes
Within 50 meter range

② Stay more than 15 minutes
Within 100 meter range

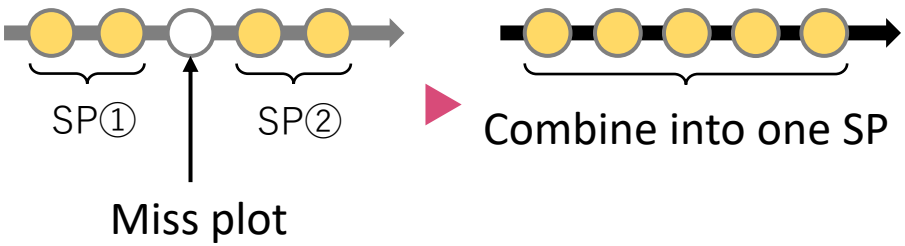
Source: Kusunose and Sasaki (2023)

2.4 Data cleaning and trip identification

Following cleaning procedures performed on the extracted SPs.

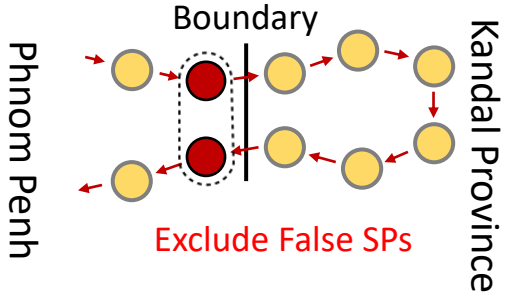
① Exclude the plot where SPs are split by miss plot

Combine split SPs where a single miss plot which does not meet conditions causes SP to break off



② Exclude SPs that were falsely detected near the target zone boundary

Data outside the Phnom Penh area is removed, so SPs are incorrectly detected near the boundary.



As a result, a total of 1506,670 trips were extracted from 45,176 IDs (Average of 33.35 trips per ID)

Source: Kusunose and Sasaki (2023)

2.5 Trip and data analysis classification

	Trip classification	Data analysis			Trip	
		OD	Trip Purpose	Trip mode	Nos	%
A	Both ends with SPs with clear departure and arrival time and detailed plot intervals	○	○	○	105,785	7.0
B	Both ends with SPs with clear departure and arrival time but scattered plot intervals	○	○	△	30,264	2.0
C	Both ends with SPs without clear departure and arrival time with detailed plot intervals	○	○	△	432,772	28.6
D	Both ends with SPs without clear departure and arrival time and with scattered plot intervals	○	○	△	118,105	7.8
E	Both ends with SPs without any plot interval within 2 hours trip range	○	○	×	364,115	24.0
F	Both ends with SPs without any plot interval over 2 hours trip range	○	○	×	32,002	2.1
G	Trips crossing boundary with detailed plot intervals	○	×	×	17,891	1.2
H	Trips crossing boundary without detailed plot intervals	○	×	×	76,765	5.1
I	Unclear SPs with detailed plot intervals	△	×	×	74,781	4.9
J	Unclear SPs with scattered plot intervals	△	×	×	254,190	16.8
K	A group of plots which cannot be defined as trip	×	×	×	7.666	0.5

2.6 Data fusion (with POI)

POI (Point of Interest) A specific point on a map; a GIS object treated as a POI, with its name and attributes added to the location information

POI downloaded from Open Street Map (OSM), available free of charge and used almost universally

Point data of Café in Phnom Penh

Polygon data of schools in Phnom Penh

Source: Kusunose and Sasaki

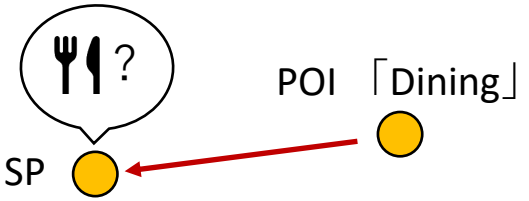
2.6 Data fusion (with POI)

An algorithm developed to estimate the location of SPs based on the distance to SPs.

Stay point at 3 AM

→「Home」

Tied to the nearest POI within 50m



some POIs excluded or set time constraints

11 POIs

Source: Kusunose and Sasaki (2023)

2.7 Data analysis (Trip purpose)

Comparison of trip purpose by Mobile GPS log and 2012 PT Survey

Purpose	Mobile GPS Log		2012 Person Trip Survey	
	No of trips	Share [%]	No of trips	Share [%]
To home	573,033	52.78	2,111,944	49.17
To work	171,996	15.84	850,387	19.80
To school	**28,764**	**2.65**	**620,414**	**14.45**
To business	**208,159**	**19.17**	**224,442**	**5.23**
To private	63,688	5.87	303,500	9.16
Other	30,970	3.68	94,024	2.19

Distortion observed between Mobile GPS Log and PT survey:

- **To school**: School kids prone not yet obtain Smart Phone or prohibited to bring Phone or switch off at school
- **To business**: PT survey hardly captures business trips due to omission and/or simplification of respondents journeys. Improved accuracy and substitute business trip survey by Mobile GPS log data

Source: Kusunose and Sasaki (2023)

3. Bus probe data for improved traffic control in Phnom Penh

Pilot bus priority signal under JICA PiBO

3.1 JICA Project for Improvement of Public Bus Operation in Phnom Penh (PiBO)

Overall Goal:

Ridership on the public transport in Phnom Penh is promoted: **70,000 pax per day by 2025 (6,000 pax per day in 2016)**

Project Purpose:

Bus operation management capacity in Phnom Penh is improved.

Output:

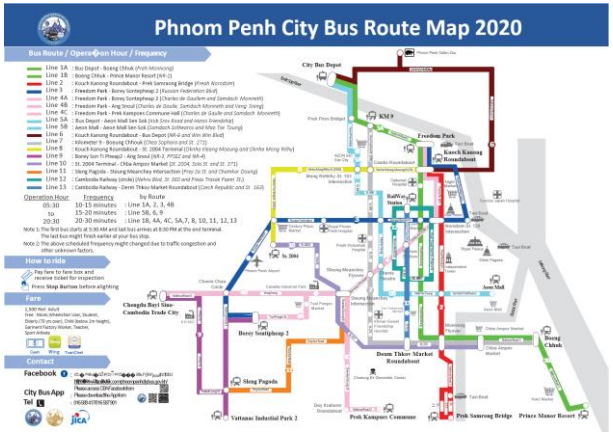
1. Improved **bus operation capacity**
2. Improved **maintenance and inspection capacity**
3. **Training of bus drivers**
4. Improved **business management capacity**
5. Improved policy planning for **public transport priority measures**

Project Duration:

January 2017 to August 2022 **including 20-month suspension (March 2020 to Oct 2021)**

Source: JICA PiBO (2022)

3.1 JICA Project for Improvement of Public Bus Operation in Phnom Penh (PiBO)



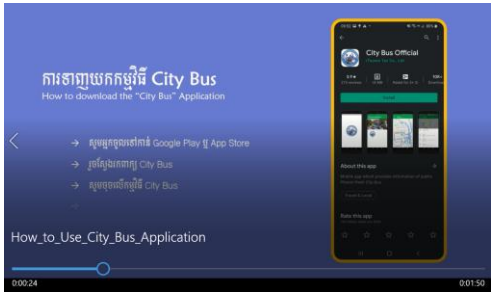
Bus Map
Repair Workshop

BOMS
Bus Location App

Source: JICA PiBO (2022)

3.1 JICA Project for Improvement of Public Bus Operation in Phnom Penh (PiBO)

Bus PR video
Bus PR event



Rapid Bus (ADB TA)
Bus Priority Signal
e-bike Share

3.2 Pilot Bus Priority Signal under JICA PiBO

Purpose of the Pilot Bus Priority Signal:

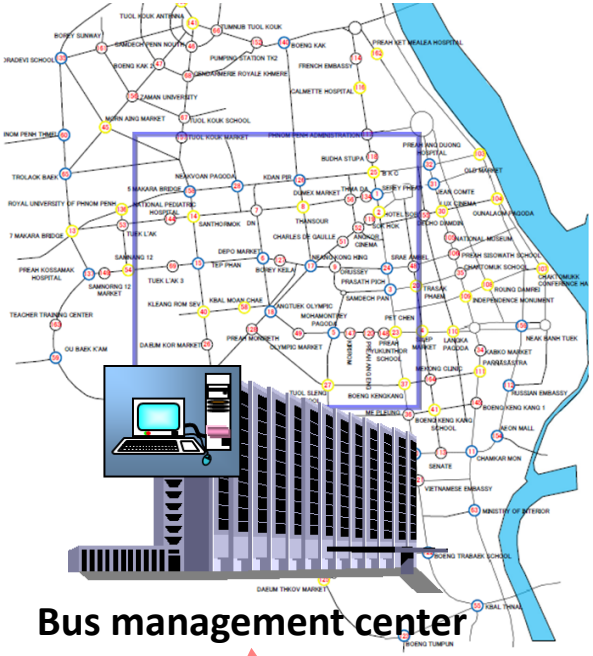
(1) To verify application of bus probe data for justifying and improving signal control parameter and for monitoring improved parameters performance

(2) To confirm reduction in travel time of bus route subject to priority control along Line 4A and 4B) and contribution to on-schedule operation of City Bus

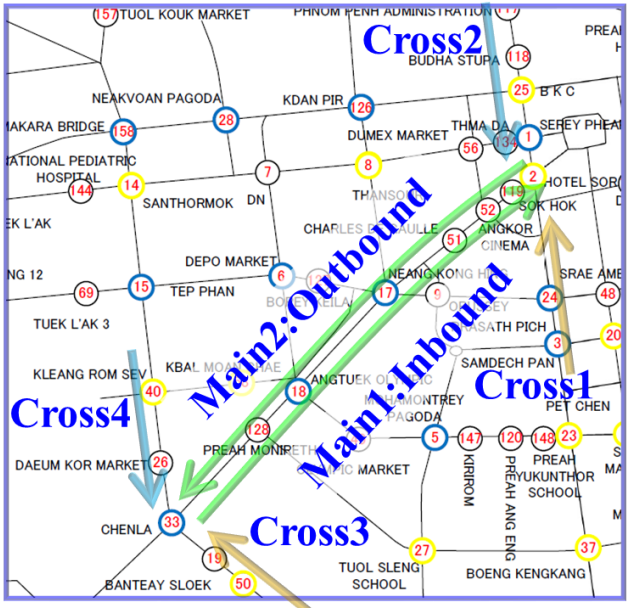
Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

Overview of Pilot Bus Priority Signal

Traffic control center



Bus management center



Target Routes:
Main1 and Main2
(Charles-de-Gaulle Blvd.)

Historical signal control data

Historical Bus probe data

Signal control data

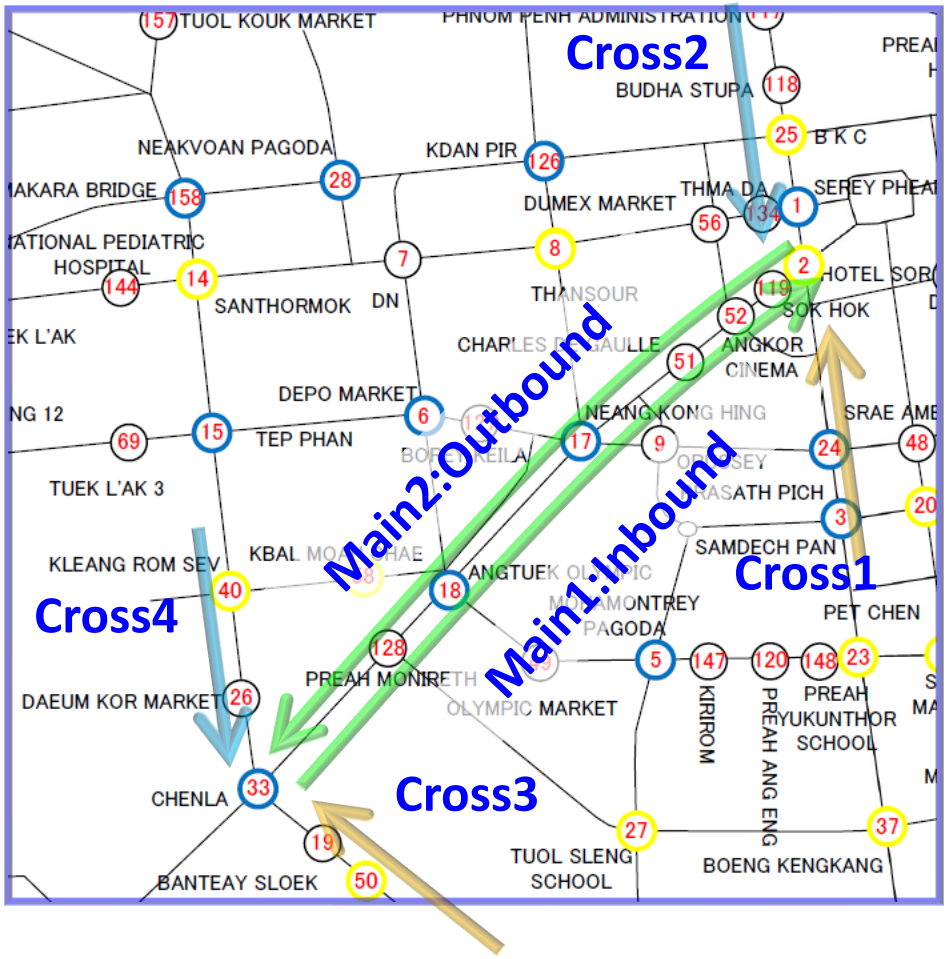
Bus probe data

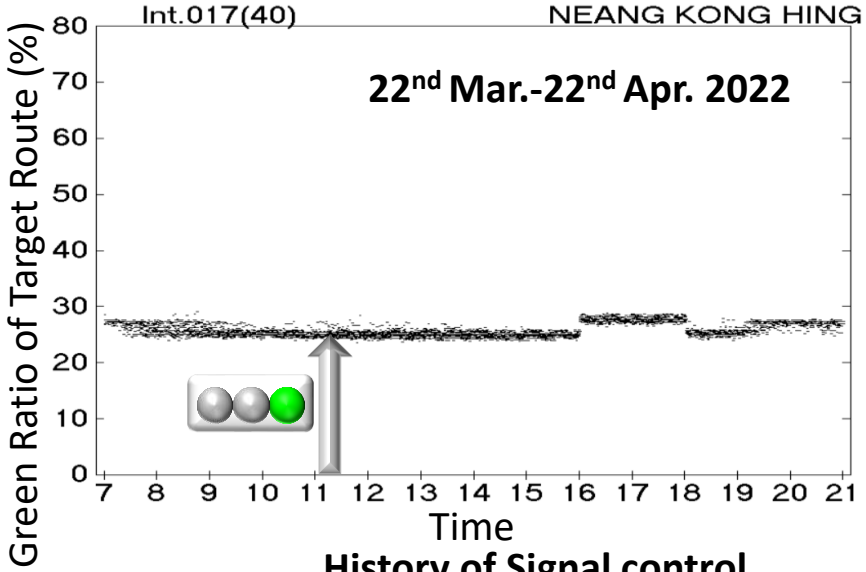
Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

Bus Route Map

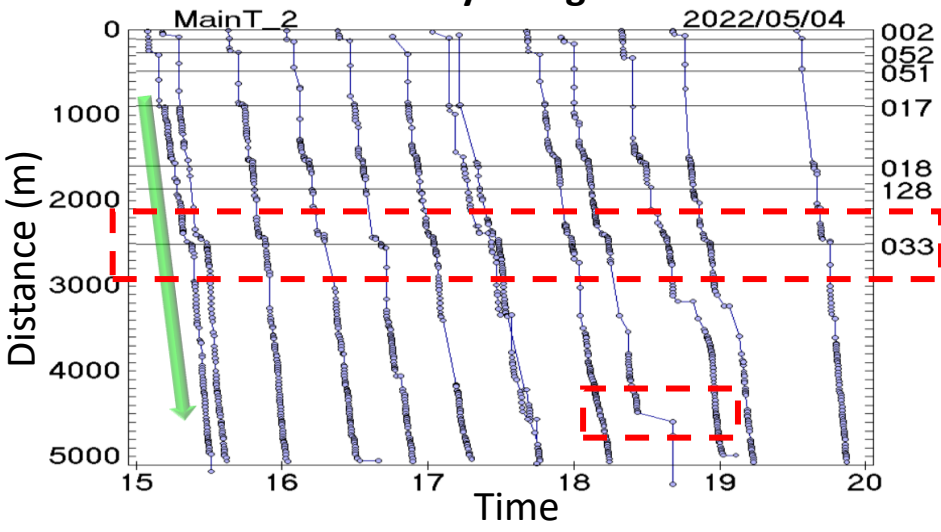
Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

Visualization of Bus probe data





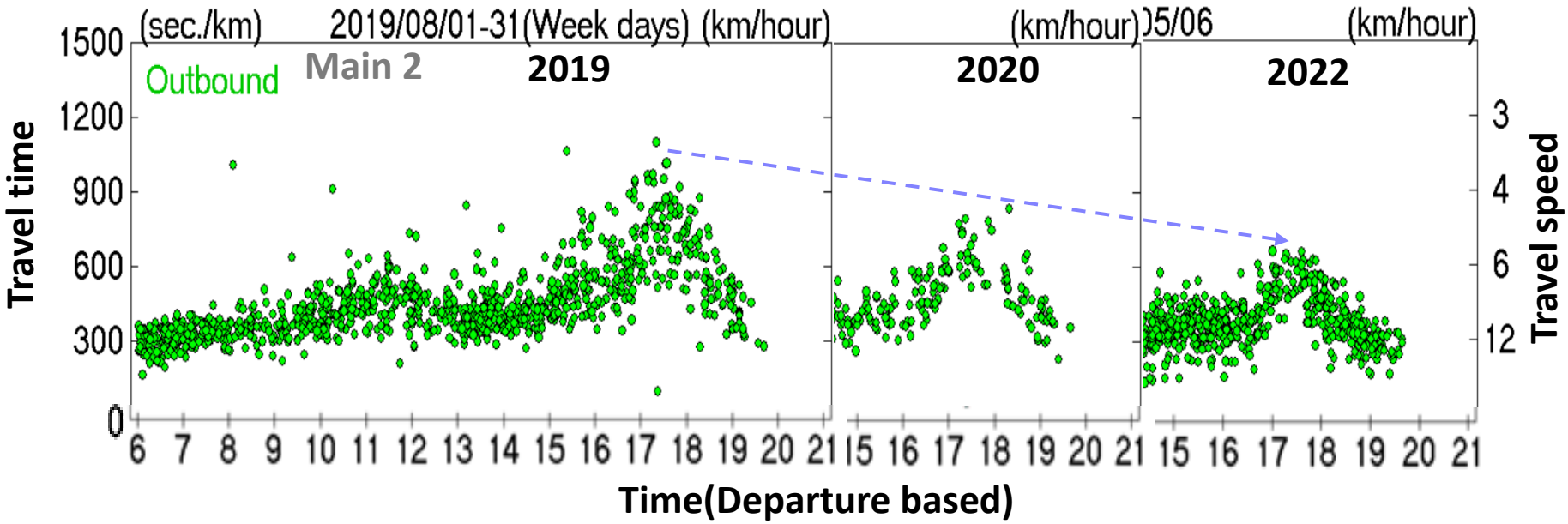
History of Signal control



History of Bus travel tracking

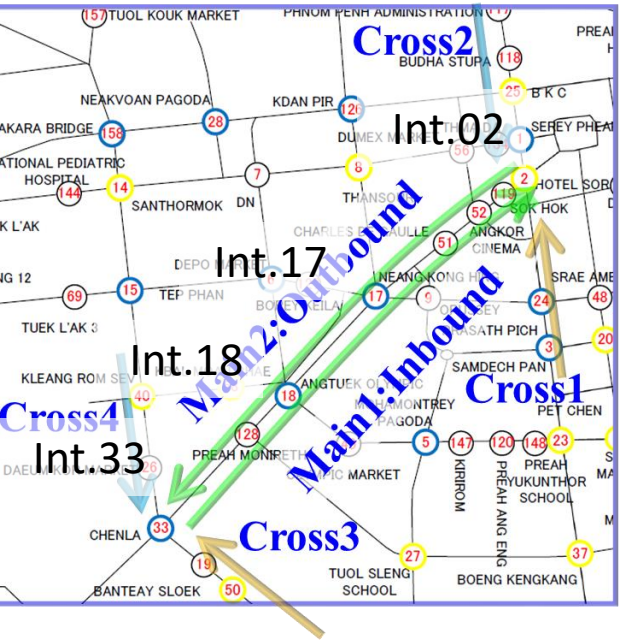
Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

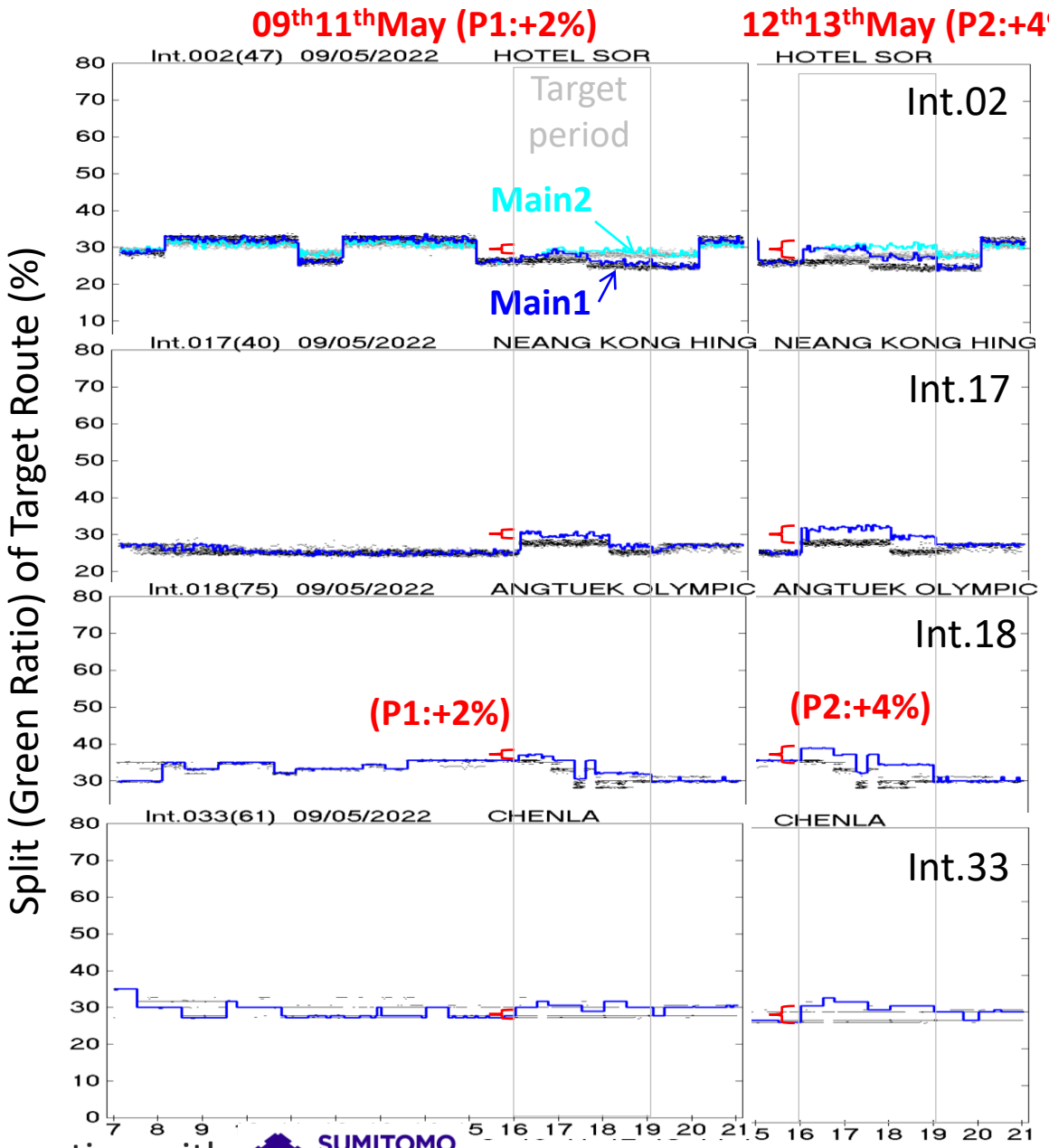
Visualization of Bus probe data



Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

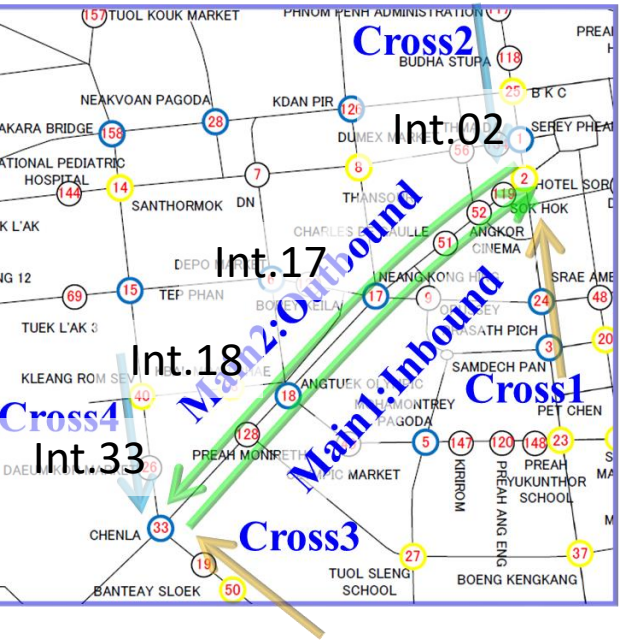
Pre-test: Green time modification for priority route

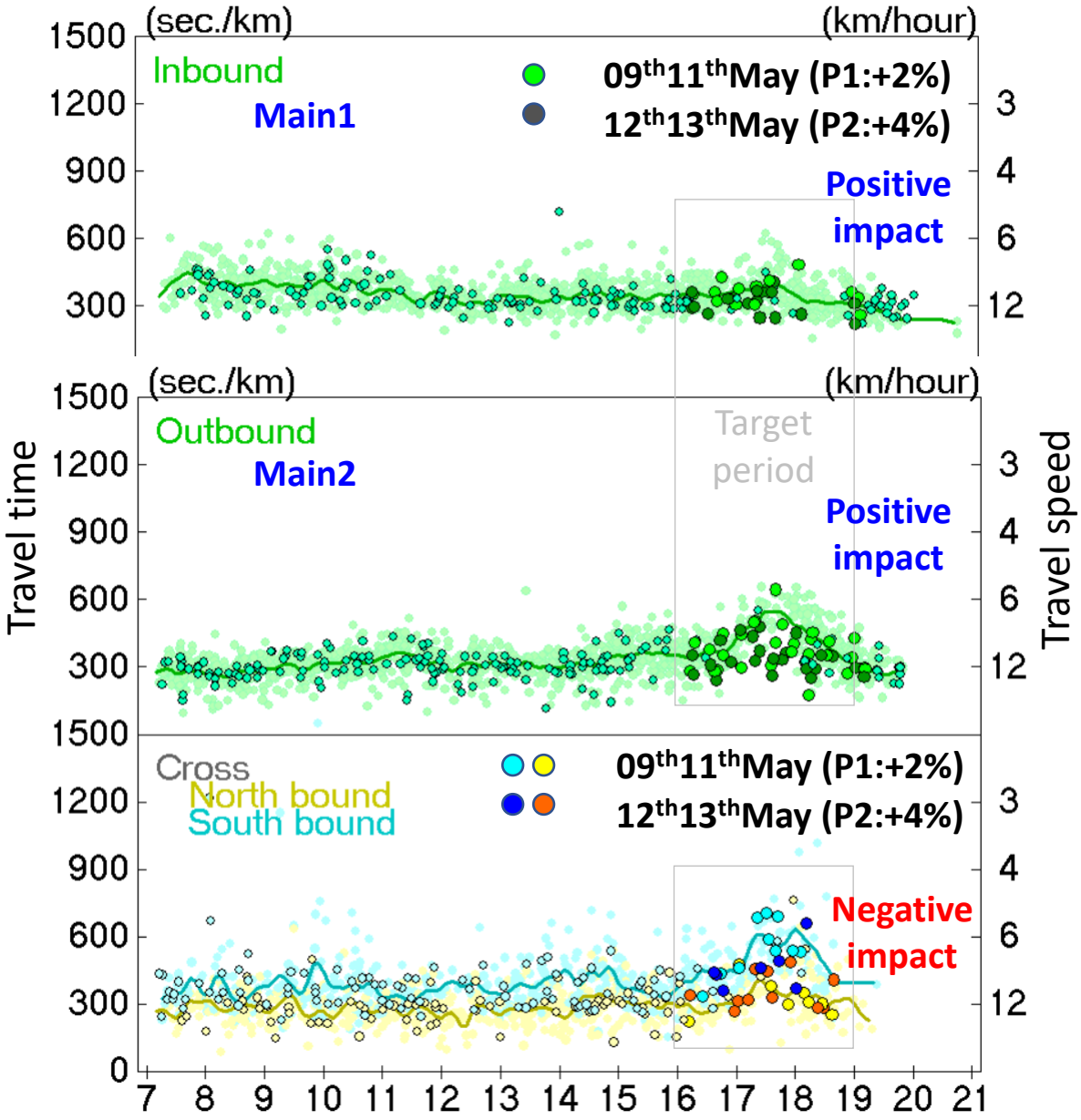




Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

Pre-test: Green time modification for priority route





Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

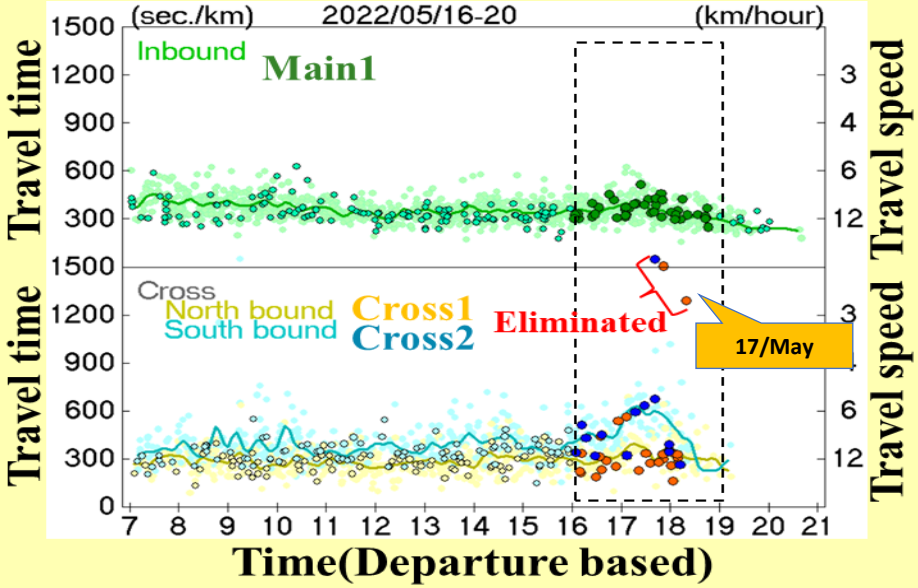
Pilot Bus Priority Signal and Overall Result

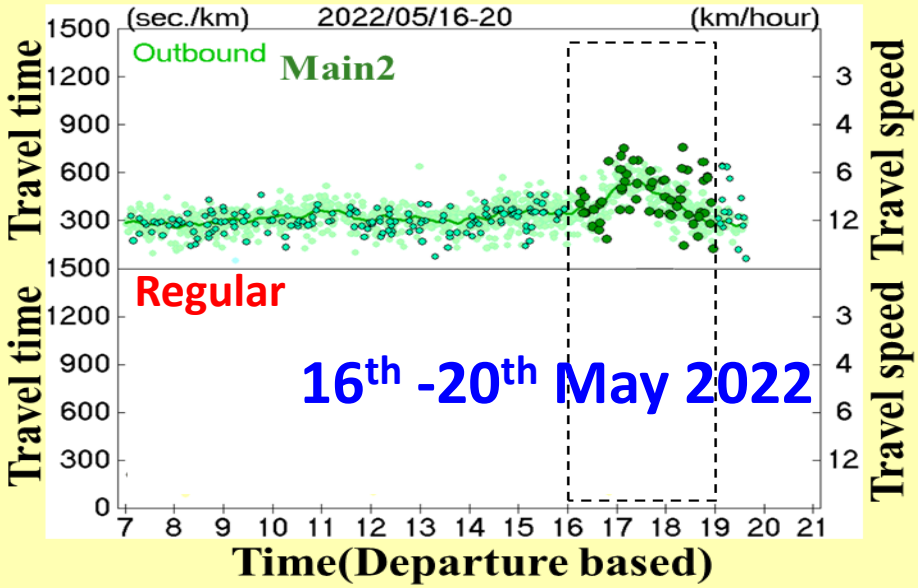
Term	16/May – 20/May, 2022	26/May – 1/June, 2022		
Parameter Plan	Regular (No priority control)	P1: #2, #17, #18 P2: #33		
	Average travel time (sec./km) {B}	Average travel time (sec./km) {A}	Difference (sec./km)	(A-B) / B
Main1	356sec.(42 probe data)	328sec.(37 probe data)	-28	-7.9%
Main2	449sec.(58 probe data)	370sec.(36 probe data)	-79	-17.6%
Cross1	304sec.(23 probe data)	460sec.(24 probe data)	+156	+51.3%
Cross2	433sec.(13 probe data)	419sec.(13 probe data)	-14	-3.0%
Cross3	No Data	No Data	--------	-------
Cross4	Ni Data	Ni Data	--------	-------

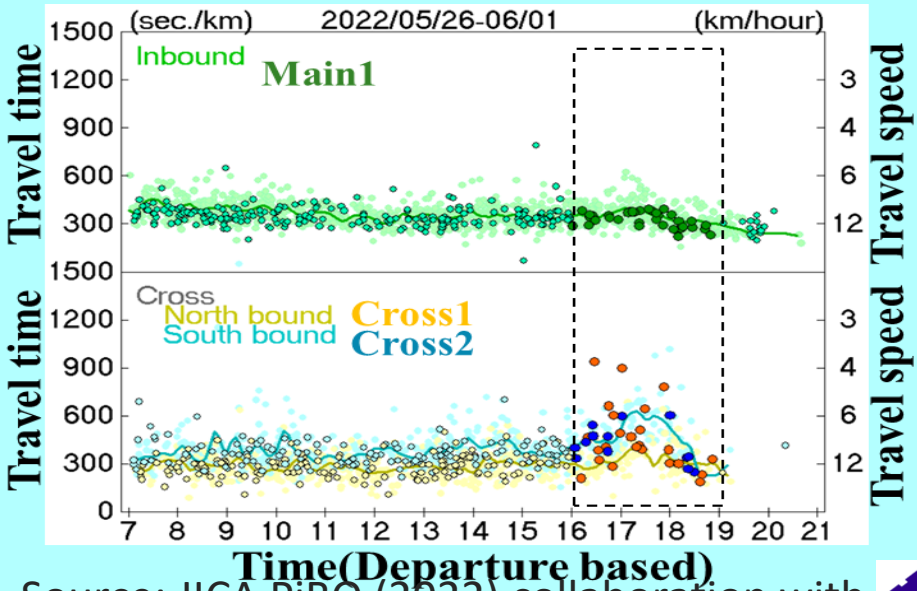
Main1: West to East, Charles-de-Gaulle Boulevard
Main2: East to West, Charles-de-Gaulle Boulevard
Cross1: South to North, Monibong Boulevard
Cross2: North to South, Monibong Boulevard
Cross3: South to North, Mao Tse-Tung Boulevard
Cross4: North to South, Mao Tse-Tung Boulevard

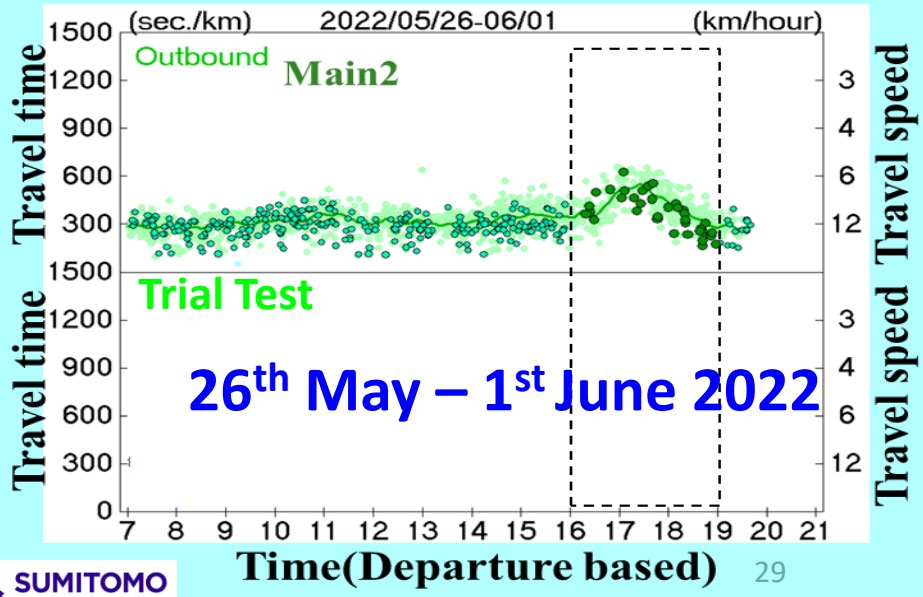
Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

Comparison with/without Pilot Bus Priority Signal









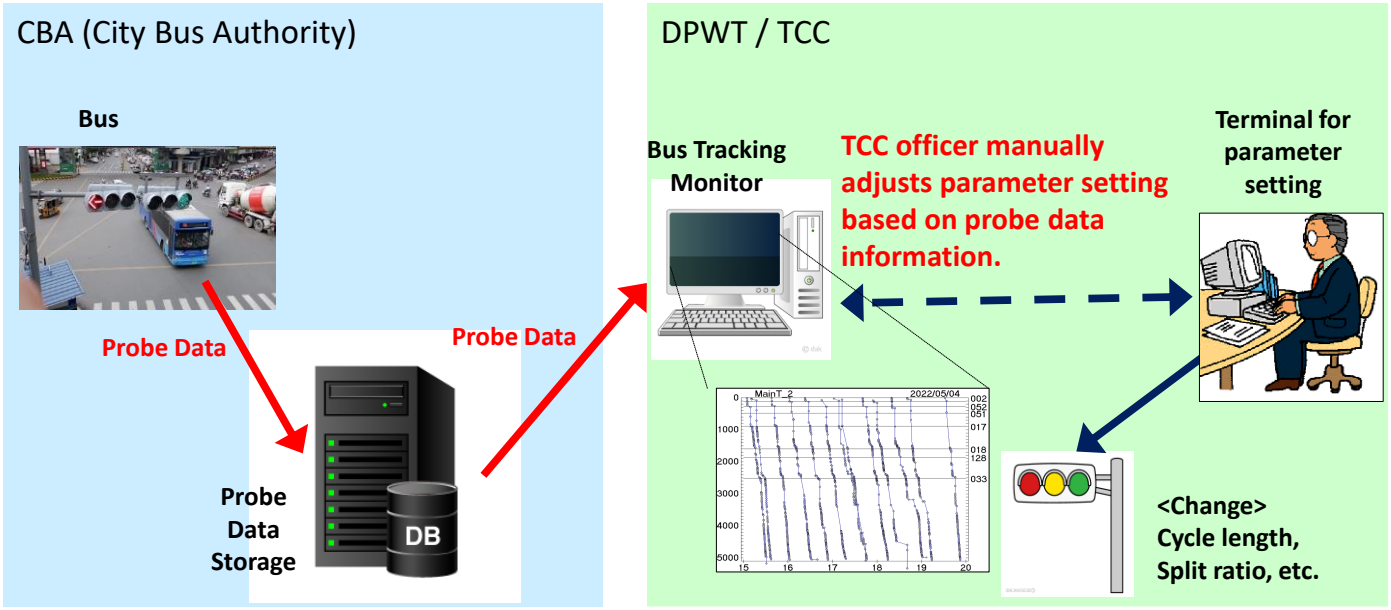
Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

Pilot Bus Priority Signal Result and Recommendation

Result and Recommendation:

(1) Bus probe data essential for improving signal control parameter and DPWT/TCC officers can grasp real-time information for traffic condition (under constraints of insufficient detector data)

(2) City Bus operation improved under slim control parameter changes which significantly contributes to on-schedule operation. Impacts on crossing traffic be also monitored by Bus probe data and adjusted to minimize the adverse impact



Source: JICA PiBO (2022) collaboration with SUMITOMO ELECTRIC

4. Wrap up

Wrap up

- Two case studies reveal potential and effective usage of ‘Big Data’ in transport planning and practices in Cambodia
- In case of mobile GPS log, algorithms for data cleaning, analysis, data fusion and sample expansion are vital and which may require further studies.
- Note that privacy policy of mobile OS tends to affect availability of mobile GPS log
- Taking Bus probe data as real-time ‘Big Data’ as an example, the bus priority signal tested as pilot
- It proved potential use of Bus probe data for optimization of traffic control signal, which requires organizational coordination