

# Uncertainty in Traffic Congestion Estimation Using Nationwide Pseudo Trip Data and Agent Based Simulation

## Background

Aayush, Yanbo Pang, Yoshihide Sekimoto

Availability of good quality people mass movement data helps in better urban planning and optimum infrastructure resource utilization. Although modern human tracking devices has made people mass movement data available for research purposes, but they have inherent constraints. Such data sets are expensive and bind by severe privacy policy and are aggregated for anonymization. Thus, can not be resourcefully used by researchers.

Sekimoto Lab is conducting a research on generation of nation-wide pseudo pflow (PPF) data. This data is anonymized and represents synthetic agents across entire Japan. The development in computing prowess in recent years has also led to agent-based traffic simulators which can simulate millions of agents over a large network. The variation in travel demand across geography is also better represented through a large network simulation rather than a small network e.g within a city. But such large network modelling require more computational resource and time. This study is to find a balance between computational time and improving accuracy of an agent's trajectory through a nationwide simulation using pseudo Pflow data.

## Contribution

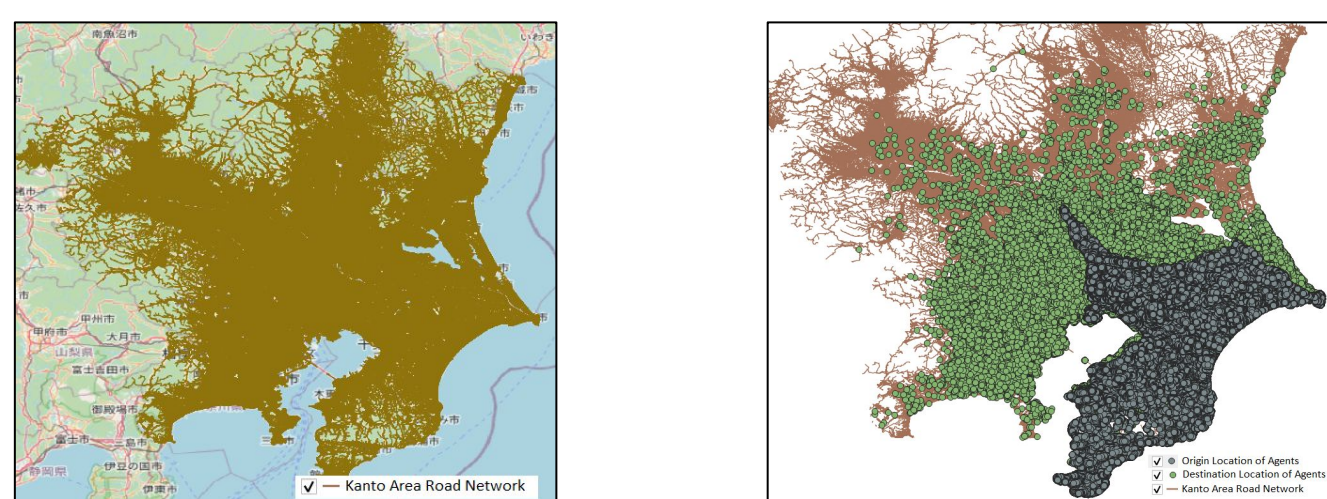
The pseudo pflow trajectory data considers shortest path between origin and destination of an agent. In real world people do not necessarily commute through a shortest path. Therefore, this research intends to improve the trajectory data by incorporating traffic congestion and agent to agent interaction during daily commute by each agent to include element of complex adaptive system in pseudo pflow data.

Also, an attempt to simulate nationwide movement of all agent over entire road network of Japan is made as it will represent a better variation in trajectory data by considering inter-region connectivity.

## Dataset

**(i) Open Psuedo Pflow Data set developed by Sekimoto Lab:** The dataset has trip data ,trajectory data and hourly link(road) volume for synthetic agents. The trip data has origin, destination coordinates along with time of arrival and departure. The link volume data has hourly volume of traffic in each link.

**(ii) Digital Road Network data(DRM) :** Digitized road network data of Japan is used as base network data for simulation.



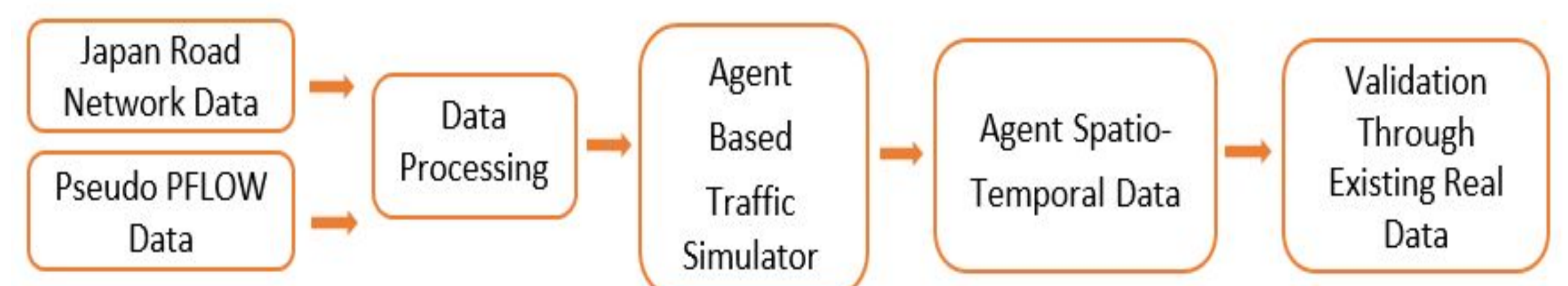
(a) DRM of Kanto region with Origin and destination coordinates of Agents of Chiba prefecture

## Results

1)The selected Kanto area for simulation has 6677086 links spread across all prefectures of Kanto region. The number of agents simulated over this network were selected only from Chiba prefecture. Agents whose destination was beyond the network area were removed from simulation. Total number of agents simulated is 651396 and number of trips is 1302792. The total time for 1 simulation over Amazon Elastic computing cloud (EC2) instance of type m4.10x with 35 (virtual) CPUs and 150 GB of RAM was around 12 hours.

2) The link volume after simulation was compared with link volume from PPF data for two hours – 8-9 AM and 17-18 PM. Number of links considered for validation of result in South Chiba area is 170860.

## Methodology



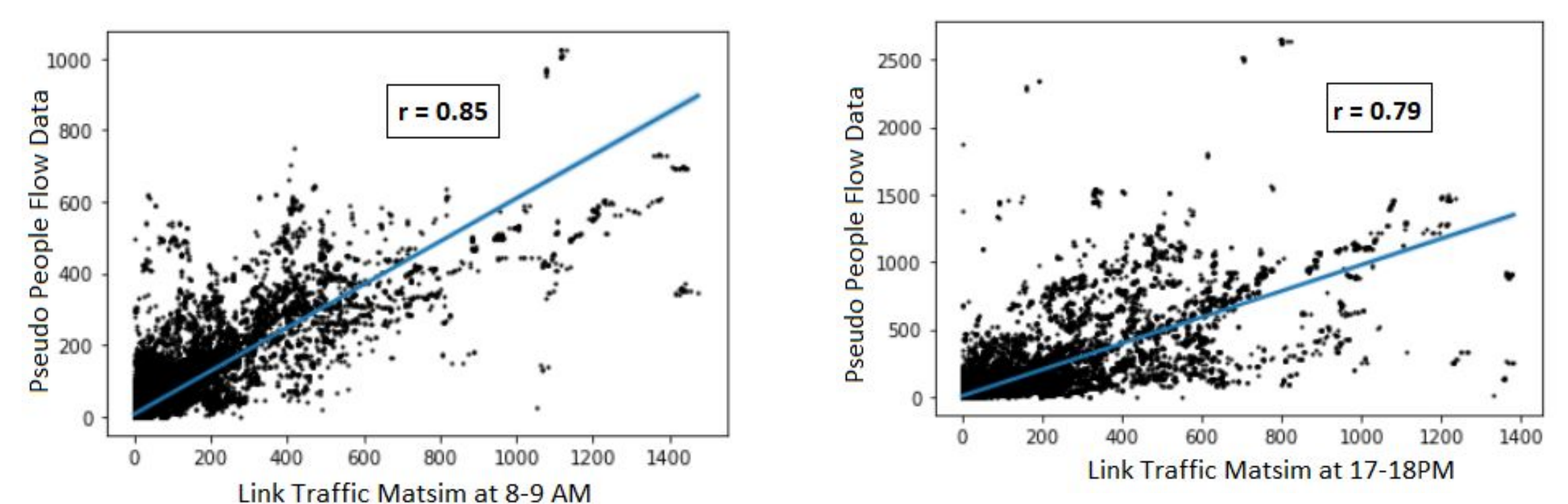
(b) Basic Flow Chart of Research

MATSim (Multi Agent Traffic Simulator) is an open-source framework for implementing large-scale agent-based transport simulations. It is proven to be computationally fast for large network simulations and is used as the traffic simulator in this research.

Pseudo Pflow trip data and digital road map data is used as input data. The raw data is converted to MATSim compatible format. To study traffic congestion trip data from pseudo pflow data is processed to extract agents whose mode of trip in one single day is only car. Activity of agents are limited to home to work and work to home.

High density Digital Road Map data of Japan is processed to create base network for simulator consisting of links and nodes. Links or road networks are considered bidirectional with no regulation.

The simulation parameters are optimized based on the available computing environment. The output is in the form of hourly link volume and agent's trajectory. The output is compared with the link volume data from PPF data for validation. Only links within South of Chiba prefecture were used to validate the result and analyze the correlation with the link volume of PPF data, as this region has fewer train routes and mostly traffic is via road.



The correlation coefficient at 8-9 AM is 0.85 and Correlation coefficient at 17-18 PM is 0.79. It shows a positive correlation. The link volume from PPF data considers agents from all prefectures and thus are more than when compared with simulation output, which considers only Agents from Chiba prefecture. The correlation is expected to improve when agents from adjoining prefectures are taken into consideration and with more iterations.