Improvement of High Accuracy People Flow Simulation by Comparing Real Data and Synthetic Data

Haruki ADACHI, Yanbo PANG, Yoshihide SEKIMOTO

Background

Although human flow simulation in smart city policies and urban planning plays an important role in congestion mitigation and traffic optimization, the use of real data is subject to privacy protection and cost constraints. Therefore, CSIS has been constructing pseudo human flow data that can be used nationwide, based on geospatial information such as census and building data, and historical traffic statistics data. However, although the pseudo-data is statistically consistent with the real data, it has not been sufficiently verified to what extent the behavior at the individual level is consistent with the real data.

Purpose

The purpose of this study is to evaluate how well the simulated data reproduces actual human flow behavior using the National Pseudo Human Flow Data ver. 2.0 and real human flow data, and to identify areas for improvement to improve the accuracy of the pseudo data generation model. By doing so, we aim to increase the usefulness of the pseudo-data in smart city policies and urban planning.

Method

In this study, we first collect national pseudo human flow data ver. 2.0 and real human flow data, and extract behavioral indicators such as action radius, trip distance, and length of stay. Next, we use statistical methods to analyze the differences between the indicators in both data sets and evaluate the reproducibility of individual-level behavior patterns. Based on the results of these analyses, we propose improvements to the pseudo-data generation model.
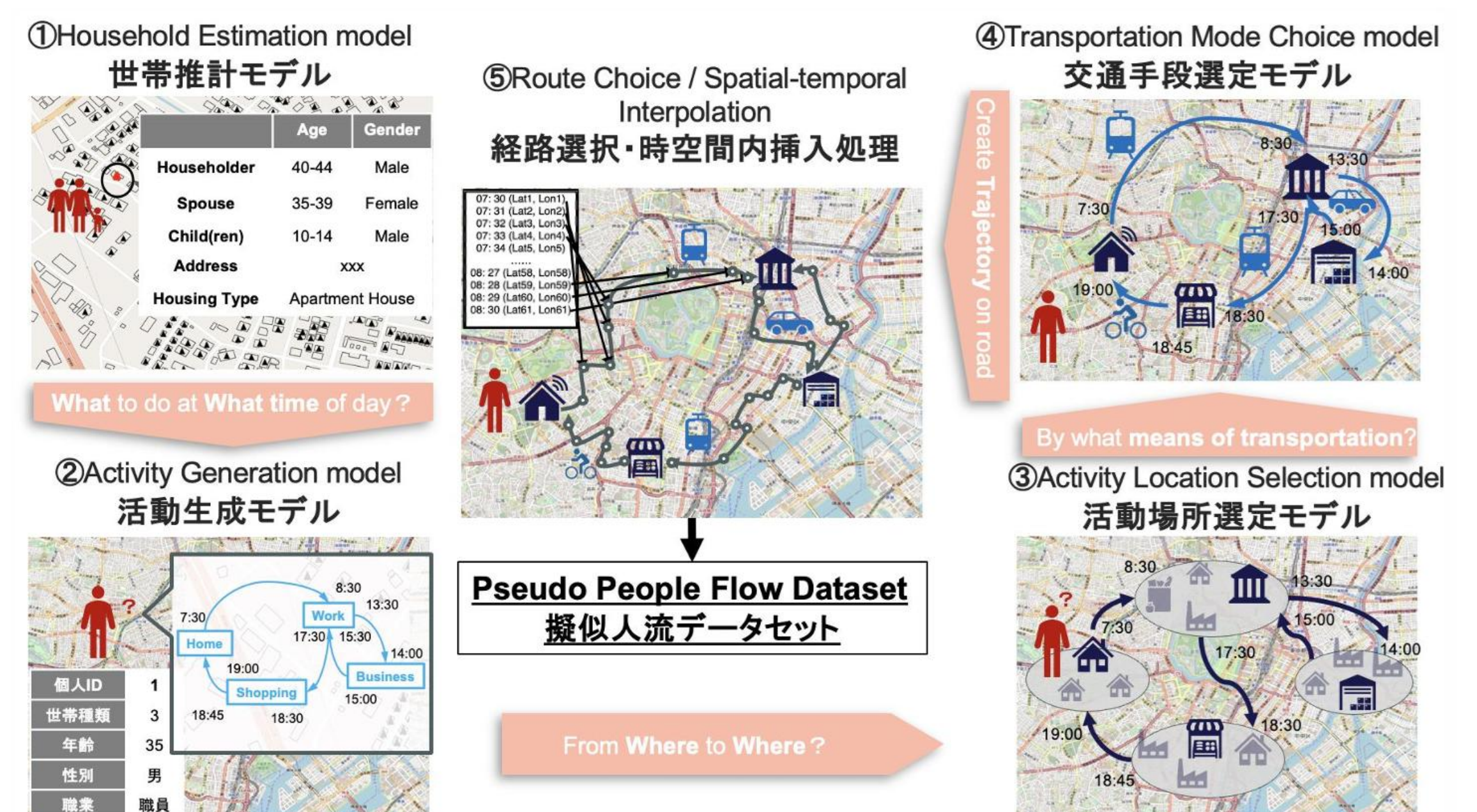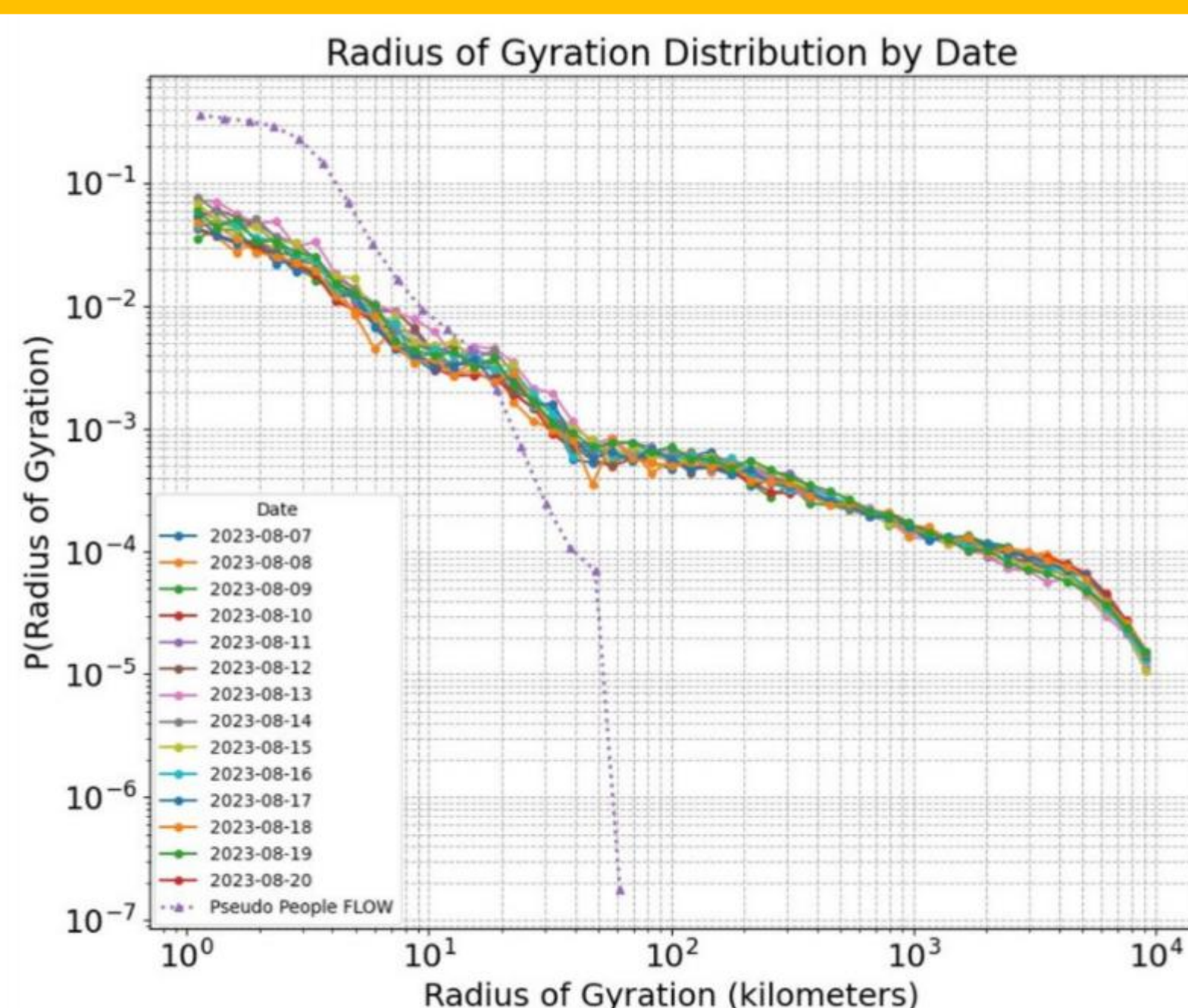
Real data：
Blogwatcher disaggregated point data
Psuedo people flow data：
Nationwide People Flow Data 2.0

	実データ	擬似人流データ
期間	2023/08/07～2023/08/19	2020/10/1
地域	東京23区	東京23区
ユーザ数	1,707,422	9,856,992
ログ数	1,827,825,300	(時空間内挿方法より)

Comparison result

The action radius reflects the range of an individual's daily activities. The pseudo-data show low probability in the area of long-distance movement, suggesting that there is a problem in reproducing long-distance movement.
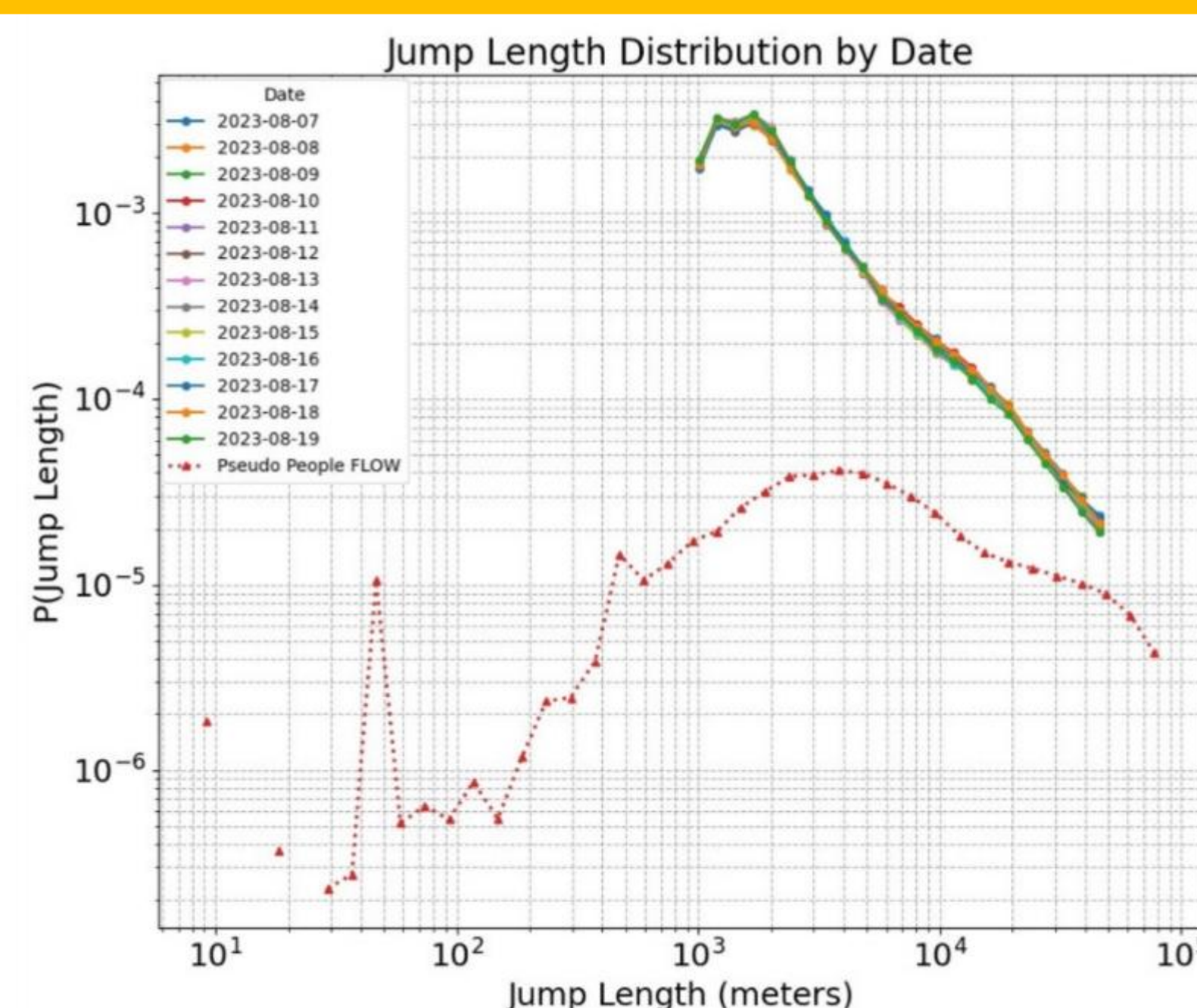
Trip distance indicates the distance between consecutive moves. The real data shows a smooth distribution from short to long distances, while the pseudo-human flow data shows an unnatural peak in the middle distance.
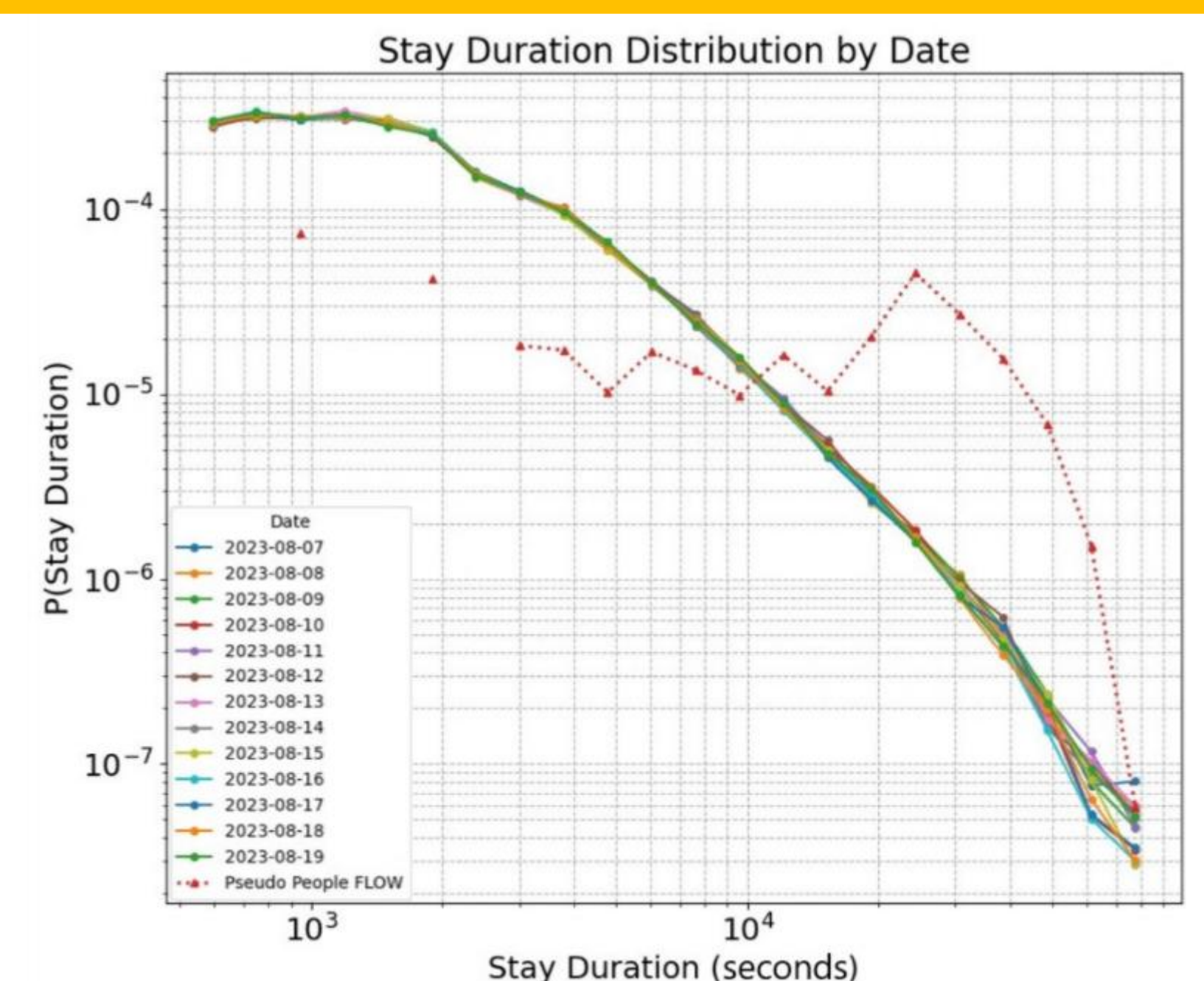
While the real data and the pseudo human flow data were similar for short stays, the pseudo human flow data decreased rapidly for the distribution of long stays, indicating that there is a limit to the reproduction of behavior over long periods of time.

◆Future plans

・Relax the 100 km radius action distance limit when creating pseudo-human flow data
・Compensate for short distances traveled that are lacking in person trip surveys.